

KLASIFIKASI KUALITAS UDARA MENGGUNAKAN METODE EXTREME LEARNING MACHINE (ELM)

Rachma Raudhatul Jannah¹, Muhammad Zulfikar Sholahuddin², Dina Zatusiva Haq³, Dian C. R. Novitasari⁴

^{1,2,3,4}Program Studi Matematika, Fakultas Sains dan Teknologi,
UIN Sunan Ampel Surabaya
Gunung Anyar, Surabaya, East Java 60294

¹rachmaraudhatul18@gmail.com, ²muhammadzulfikars746@gmail.com, ⁴diancrininov@gmail.com

ABSTRACT

Air quality is a critical factor affecting both ecological and human well-being. Air pollution is a global epidemic that poses a threat to human health and the environment. High population density resulting from industrial expansion and the increased number of motor vehicles are two primary causes of declining air quality in metropolitan areas. Air pollutants include surface ozone (O₃), dust particles (PM₁₀), sulfur dioxide (SO₂), nitrogen dioxide (NO₂), and carbon monoxide (CO). Researchers have begun exploring the use of Extreme Learning Machine (ELM) to classify air quality. The ELM method assesses air quality as either very good or poor. In this study, we compare datasets to evaluate the effectiveness of hidden node parameters using the split method. Our tests indicate that the split method impacts accuracy, sensitivity, and specificity. The ideal model with a 70:30 split ratio and 15 hidden nodes achieved a 90% success rate.

Keywords- ELM, Dampak Polusi udara, Kualitas udara, Klasifikasi

PENDAHULUAN

Pencemaran udara merupakan terkontaminasinya udara dimana aktivitas manusia melepaskan zat, energi, serta komponen lain ke udara menyebabkan udara melebihi baku mutu yang ditentukan [1]. Adanya zat pencemar di udara ditambah dengan dampak lainnya, cenderung meningkatkan jumlah orang yang mengidap atau terjangkit penyakit kanker paru-paru, emphysema, tuberkulosis, pneumonia, bronchitis, asma, dan bahkan influenza [2]. Kesehatan manusia menjadi perhatian utama ketika mempertimbangkan dampak jangka panjang dari polusi udara. Khususnya di pusat-pusat perkotaan [3].

Pasalnya, hal itu berdampak langsung pada kenyamanan kota dan kesehatan masyarakat. Udara yang kita hirup dipengaruhi oleh kombinasi variabel alami dan buatan manusia. Harus ada upaya peningkatan untuk menghindari dan memantau polusi udara secara ketat karena kualitas udara memburuk setiap hari. Polutan udara juga berasal dari ozon permukaan (O₃), partikel debu (PM₁₀), sulfur dioksida (SO₂), nitrogen dioksida (NO₂), dan karbon monoksida (CO). Kami mengikuti peraturan Menteri Negara Lingkungan Hidup nomor: KEP 45/MENLH/1997 tentang Indeks Standar Pencemar Udara (ISPU), sekaligus menghimpun data dari negara ini.[4].

Dengan pesatnya perkembangan ilmu pengetahuan dan teknologi, banyak perubahan yang terjadi dalam kehidupan. Karena jumlah penduduk di kota lebih besar dibanding dengan jumlah penduduk di desa, maka permasalahan lingkungan di kota dapat berbeda dengan di desa, seperti kualitas udara [5]. Kemajuan pesat dalam teknologi komunikasi merupakan ciri khas era digital

modern. Karena berbagai keuntungan yang ditawarkannya, teknologi komunikasi menjadi lebih penting bagi semakin banyak individu. Penggunaan jaringan wireless sensor network (WSN) merupakan salah satu cara di mana teknologi informasi dan komunikasi telah meningkatkan pemantauan kualitas udara [6]. Dalam jaringan sensor nirkabel, beberapa sensor node beroperasi bersama sebagai jaringan. Hal ini memungkinkan pengambilan bacaan secara real time.

Studi sebelumnya telah menyelidiki algoritma ELM karena berkaitan dengan prediksi laser cutting. Dibandingkan dengan algoritma Pemrograman Genetika, metode Extreme Learning Machine memberikan prediksi yang lebih baik dan lebih akurat, menurut studinya [7]. Serta terdapat penelitian tentang penerapan metode Extreme Learning Machine terhadap penyakit Chronic Kidney Disease, dari hasil penelitian tersebut didapat hasil nilai akurasi yang sangat baik yakni sebesar 96,7% [8]. Hal tersebut membuktikan bahwa algoritma ELM mampu menyelesaikan kasus klasifikasi dengan hasil yang baik dan menghasilkan nilai akurasi yang tinggi.

Di sini kita melihat temuan pengukuran kualitas udara yang dilakukan dengan menggunakan pendekatan confusion matrix. Confusion Matrix adalah alat untuk menentukan seberapa baik ide penambahan data bekerja. Data pelatihan dapat ditambang untuk informasi tersembunyi menggunakan metode data mining termasuk clustering, rule mining, dan klasifikasi [7].

Cara terbaik untuk mengatasi masalah yang disebutkan di atas adalah dengan membuat prediksi tentang kualitas udara. Salah satu pendekatan Artificial Neural Networks (ANN) yang mungkin berhasil adalah

algoritma Extreme Learning Machine (ELM); ini memiliki tingkat kesalahan yang lebih rendah, dan diproyeksikan untuk memberikan hasil prediksi yang lebih baik.

METODE PENELITIAN

Sejumlah langkah terorganisir membentuk penyelidikan ini. Mengumpulkan data untuk melatih dan menguji model adalah langkah pertama. Mengembangkan kerangka kerja untuk analisis kualitas udara mengikuti pengumpulan data. Penelitian kemudian beralih ke fase berikutnya, yang melibatkan penerapan arsitektur sistem yang dikembangkan. Pengujian, langkah terakhir dari penelitian, menentukan seberapa baik kinerja sistem yang dimaksud. Kemanjuran dan ketepatan model yang dibangun dapat dipahami dengan lebih baik dengan bantuan temuan dari tahap ini.

A. Pengumpulan Data

Pertimbangkan variabel yang mempengaruhi kualitas udara dalam penelitian kuantitatif ini. Adapun parameter udaranya adalah PM2.5, PM10, SO2, CO, O3, NO2, dan total tujuh parameter dengan masing-masing seratus lembar data, Max Critical. Data ini diambil dari kaggle pada tanggal 1 Mei 2024 [9].

Tabel 1 Data Kualitas Udara

No	PM 2.5	PM 10	SO ₂	CO ₂	O ₃	NO ₂	Max Critical	Category
1	45	30	2	69	19	0	69	0
2	44	16	2	58	33	0	58	0
3	43	12	2	46	15	0	46	1
4	40	8	2	34	29	0	34	0
...
100	33	19	1	32	0	0	32	1

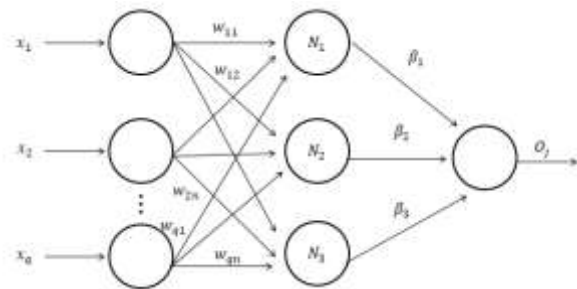
B. Alur Penelitian

Memasukkan data kualitas udara dari situs Kaggle merupakan tahap pertama dalam penelitian ini. Data tersebut terbagi dalam dua kategori: kualitas udara baik (56 catatan) dan kualitas udara buruk (44 catatan). Selain itu, teknik split digunakan untuk membagi data menjadi set pelatihan (*Training*) dan pengujian (*Testing*). Split dalam pemrograman adalah operasi yang digunakan untuk membagi atau memisahkan suatu string atau objek menjadi bagian-bagian yang lebih kecil, biasa disebut token atau elemen, berdasarkan kriteria tertentu yang disebut pembatas. Pembatas ini dapat berupa karakter tunggal, serangkaian karakter, atau pola yang menentukan breakpoint. Proses split umumnya digunakan dalam pemrograman, khususnya dalam pengolahan data [10]. Dalam metode tersebut dataset menjadi dua bagian secara acak: satu bagian sebagai data training dan bagian lainnya sebagai data testing, lalu dataset dibagi sesuai dengan rasio pembagian yang telah ditentukan sebelumnya. Data pelatihan digunakan untuk melatih model, dan fitur data diuji untuk menguji keakuratan dan validitas hasil pelatihan [11]. Setelah itu, masukkan angka acak untuk menentukan berat lapisan yang terkubur. Data tampaknya tidak memiliki rentan jumlah yang berbeda jauh, tetapi tetap penting untuk menormalkannya untuk berjaga-jaga. Setelah itu, gunakan metode Extreme Learning Machine untuk

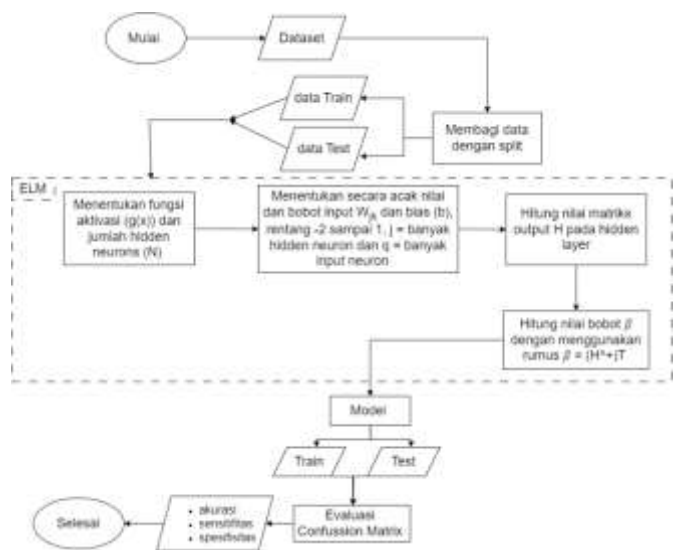
melanjutkan kategorisasi.

C. Extreme Learning Machine

Salah satu jenis jaringan saraf tiruan dengan hanya satu lapisan tersembunyi adalah Extreme Learning Machine (ELM), yang muncul sebagai pendekatan baru untuk mempelajari jaringan saraf tiruan feedforward. [12]. Jaringan feedforward menggunakan parameter-parameter yang ditentukan secara manual seperti input weight dan bias [13]. Ketika kecepatan pembelajaran jaringan saraf feedforward buruk, ELM dapat membantu meningkatkan kinerjanya. Jika dibandingkan dengan teknik pembelajaran Propagasi mundur, kecepatan ELM dan kinerja generalisasi ribuan kali lebih besar [14]. Sementara hidden neuron sangat penting untuk model jaringan saraf tiruan, menurut teori algoritma ELM, nilainya sepenuhnya tidak bergantung pada aplikasi dan dapat diputuskan secara acak. Meskipun begitu, ELM mampu memberikan pendekatan universal untuk fungsi non-konstanta kontinu sebagian [15]. Node hidden adalah neuron yang lapisan antara input dan output dari sebuah jaringan saraf tiruan. Node ini bertanggung jawab untuk memproses input dan mengirimkan hasilnya ke lapisan berikutnya. Mereka disebut "tersembunyi" karena mereka tidak langsung terlihat dari input atau output jaringan. Node tersembunyi memainkan peran penting dalam mengidentifikasi pola dan fitur dalam data input.



Gambar 1 Arsitektur ELM



Gambar 2 Flowchart ELM

Berikut ini adalah perkembangan algoritma ELM:
 a. Menyiapkan bias dan bobot awal. Bobot (W_{jk}) dihasilkan secara acak dari -1 hingga 1, dimana j adalah jumlah neuron tersembunyi dan q adalah jumlah neuron masukan. Di sisi lain bias (b) dihasilkan secara acak dari 0 hingga 1, hingga $1 \times j$.
 b. Hitung matriks keluaran lapisan tersembunyi menggunakan persamaan 1:

$$H_{init} = X(i)W(i)T + b(i) \tag{1}$$

Informasi:

- $X(i)$ = data input ke- i
- $W(i)T$ = bobot input yang di transpose ke- i
- $b(i)$ = nilai bias ke- i

Matriks H_{init} kemudian harus dihitung dengan menggunakan fungsi aktivitas sigmoid biner. Prosedur ini menghasilkan Pemetaan Matriks dari matriks H_{init} yang dibuat menggunakan fungsi aktivitas sigmoid biner non-linier, yang nilainya berkisar dari 0 hingga 1. Dengan menggunakan Persamaan 2, kita dapat menentukan proses aktivasi sigmoid biner:

$$H = \frac{1}{1 + \exp(-H_{init})} \tag{2}$$

Dimana:

c. Hitung invers Moore-Penrose menggunakan persamaan 3:

$$H^+ = (H^T \cdot H)^{-1} \cdot H^T \tag{3}$$

Keterangan:

- H^+ = Matriks Moore-Penrose
- H = Matriks output hidden layer
- H^T = Matriks output hidden layer yang ditranspose

d. Hitung bobot akhir persamaan 4:

$$\hat{B} = H^+ T \tag{4}$$

Keterangan:

- \hat{B} = Matriks output bobot
- T = Matriks kelas data latih

D. Confusion Matrix

Confusion Matrix digunakan untuk memastikan bahwa temuan tersebut akurat dalam penelitian ini. Kinerja model algoritma dapat dinilai dengan menggunakan Confusion Matrix, yang merupakan semacam tabel Matriks. Menurut data, setiap baris pada matriks mewakili kelas tertentu. Selain itu, data digunakan untuk membuat prakiraan, yang ditampilkan di setiap kolom [16]. Confusion Matrix bertujuan untuk mengevaluasi kinerja metode klasifikasi biner atau multikelas, dengan membandingkan hasil klasifikasi pada data asli [17]. Model yang dikembangkan dievaluasi dengan menghitung nilai precision, recall, F1-Score dan Accuracy menggunakan confusion matrix [18].

Tabel 2 Confusion Matrix

		Aktual	
		Kualitas Udara Buruk	Kualitas Udara Baik
Prediksi	Kualitas Udara Buruk	True Positive (TP)	False Positive (FP)
	Kualitas Udara Baik	False Negative (FN)	True Negatif (TN)

Keterangan: [19]

TP (True Postive)= jumlah dokumen kualitas udara baik diklasifikasi dengan benar sebagai kualitas udara baik
 TN (True Negative)= jumlah dokumen kualitas udara buruk diklasifikasi dengan benar sebagai kualitas udara buruk

FP (False Positive)= jumlah dokumen kualitas udara buruk yang salah diklasifikasi sebagai kelas kualitas udara buruk

FN (False Negative)= jumlah dokumen dari kualitas udara baik yang salah diklasifikasi sebagai kualitas udara buruk

Rumus confusion matrix untuk menghitung accuracy, sencivity, specificity adalah sebagai berikut: [20]

1. Accuracy merupakan hasil prediksi yang benar dari seluruh data pengujian [21]

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

2. Precision merupakan rasio prediksi TP (*True Positive*) dibandingkan dengan jumlah total prediksi positif

$$Precision = \frac{TP}{TP+FP}$$

3. Recall merupakan rasio prediksi TP (*True Postive*) dibandingkan dengan jumlah total data yang benar

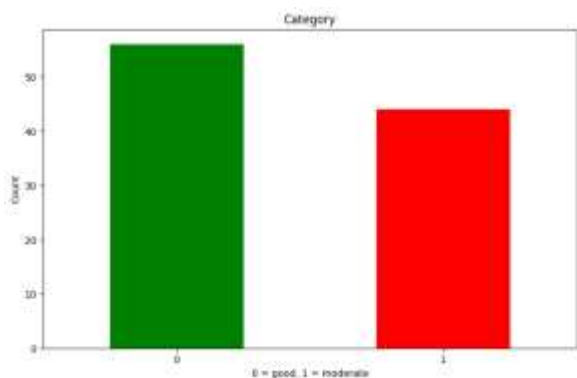
$$Recall = \frac{TP}{TP+FN}$$

4. F1- Score merupakan perbandingan rata-rata *precision* dengan *recall*.

$$F1- Score = 2 \times \frac{(Precision \times Recall)}{(Precision+Recall)}$$

HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data dari kaggle. Penelitian ini menggunakan 7 variabel untuk pengolahan data. Di sini, kami mengklasifikasikan kualitas udara menjadi dua kelas yakni kualitas udara baik dan kualitas udara tidak baik. Untuk mengetahui banyaknya jumlah dari dua kelas tersebut, maka dilakukan dengan membuat count yang dapat dilihat sebagai berikut:



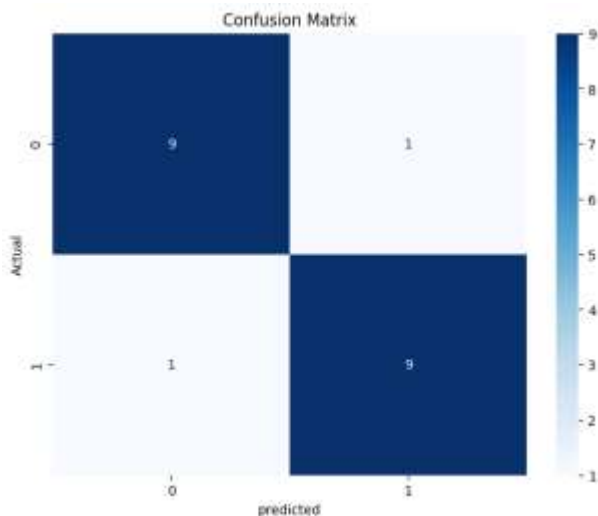
Gambar 3 Hasil Count

Extreme Learning Machine (ELM) adalah teknik yang digunakan untuk klasifikasi ini. Dengan menggunakan pendekatan split, bagilah data pelatihan dan eksperimen menjadi dua set dengan perbandingan 70:30. Ini adalah fase pertama. Dengan menggunakan nilai note hidden 10, 15, 20, dan 25, penelitian ini menjalankan banyak percobaan. Tabel 3 menunjukkan hasil pengujian.

Tabel 3 Hasil Uji Coba Note Hidden

Split		Node Hidden	Akurasi	Sensivitas	Spesifitas
Training	Testing				
70	30	10	0,9	0,9	0,83
		15	0,9	0,9	0,9
		20	0,9	0,9	0,81
		25	0,85	0,85	0,81
80	20	10	0,83	0,83	0,81
		15	0,86	0,86	0,92
		20	0,86	0,86	0,78
		25	0,86	0,86	0,71

Temuan dari Tabel 3 menunjukkan bahwa pembagian 70:30, dengan nilai note hidden 15, menghasilkan sensitivitas 90% dan spesifisitas 90%, yang merupakan kombinasi optimal untuk akurasi. Hasil pengujian dapat dipengaruhi oleh proses pemisahan, menurut hal ini. Gambar 4 menunjukkan matriks confusion untuk hasil pengujian terbaik dengan menggunakan split 70:30 dan note hidden 15.



Gambar 4 Confusion Matrix

Dari hasil confusion matrix yang diberikan untuk klasifikasi kualitas udara didapat:

True Positive (TP): 9 (kasus sebenarnya buruk, diprediksi buruk)

True Negative (TN): 9 (kasus sebenarnya baik, diprediksi baik)

False Positive (FP): 1 (kasus sebenarnya baik, diprediksi buruk)

False Negative (FN): 1 (kasus sebenarnya buruk, diprediksi baik)

Setelah itu dimasukkan pada rumus confusion matrix sebagai berikut:

$$\text{Accuracy: } \frac{9+9}{9+9+1+1} = \frac{18}{20} = 0,9$$

$$\text{Precision: } \frac{9}{9+1} = \frac{9}{10} = 0,9$$

$$\text{Recall: } \frac{9}{9+1} = \frac{9}{20} = 0,9$$

$$\text{F-1 Score: } 2 \times \frac{(0,9 \times 0,9)}{(0,9+0,9)} = 0,9$$

KESIMPULAN

Hasil diatas dapat disimpulkan penerapan split dapat memberikan hasil yang baik. Nilai evaluasi kinerja data split memberikan hasil yang terbaik bila menggunakan rasio split 70:30 dan nilai note hidden 15 yang menghasilkan akurasi 90%, sensitifitas 90%, dan spesifisitas 90%, yang berarti menggunakan split dapat meningkatkan performa kinerja klasifikasi. Ini juga menunjukkan bahwa model mampu memprediksi dengan baik untuk kasus udara baik maupun buruk, dengan kesalahan prediksi yang relatif rendah (10%). Perencanaan strategis untuk menjaga kualitas udara diharapkan mendapat manfaat dari hasil kategorisasi ini.

REFERENSI

- [1] J. Abidin dan F. A. Hasibuan, "Pengaruh dampak pencemaran udara terhadap kesehatan untuk menambah pemahaman masyarakat umum tentang bahaya dari polusi udara." *Pros. Semin. Nas. Fis. Univ. Riau IV*, vol. 5, no. 4, hal. 1–7, 2019.
- [2] M. Raharjo, "Dampak Pencemaran Udara Pada Lingkungan Dan Kesehatan Manusia," hal. 1–13, 2018
- [3] R. A. Hapsari dan A. Purwinarko, "Implementation of Convolutional Neural Network Algorithm Using Vgg-16 Architecture for Image Classification in Facial Images," *Recursive J. Informatics*, vol. 1, no. 2, hal. 83–92, 2023, doi: 10.15294/rji.v1i2.68059.
- [4] Menteri Negara Lingkungan Hidup, "Keputusan Menteri Negara Lingkungan Hidup No. 45 Tahun 1997 Tentang : Indeks Standar Pencemar Udara," no. 45, 1997.
- [5] M. Yasir, "Pencemaran Udara Di Perkotaan Berdampak Bahaya Bagi Manusia, Hewan, Tumbuhan dan Bangunan," *J. OSF.Oi*, hal. 1–10, 2021, [Daring]. Tersedia pada: <https://doi.org/10.31219/osf.io/nc5rg>
- [6] D. Culler, D. Estrin, dan M. Srivastava, "Overview of sensor networks," *Computer (Long. Beach. Calif.)*, vol. 37, no. 8, hal. 41–49, 2004, doi: 10.1109/MC.2004.93.
- [7] P. Mayadewi dan E. Rosely, "Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma

- Klasifikasi Data Mining,” *Semin. Nas. Sist. Inf. Indones.*, vol. 2015, no. November, hal. 329–334, 2015.
- [8] I. Fadilla, P. P. Adikara, dan R. Setya Perdana, “Klasifikasi Penyakit Chronic Kidney Disease (CKD) Dengan Menggunakan Metode Extreme Learning Machine (ELM),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 10, hal. 3397–3405, 2018
- [9] Iq Air, “Air quality in South Tangerang,” *IQ Air*, 2024. <https://www.kaggle.com/datasets/ourwit/air-quality-in-south-tangerang-indonesia-20-23?resource=download>
- [10] R. Adinugroho, *Perbandingan Rasio Split Data Training Dan Data Testing Menggunakan Metode Lstm Dalam Memprediksi Harga Indeks Saham Asia*. 2022.
- [11] Suherman dan I. Muzaky, “Analisis Penjualan Barang Laris Dan Kurang Laris Terhadap Percetakan Awfa Digiti Printing Menggunakan Metode Decision Tree Dengan Optimasi Algoritma Genetika,” *J. Teknol. Pelita Bangsa*, vol. 10, no. 1, hal. 118–130, 2019.
- [12] A. P. Ariyanti, M. I. Mazdadi, A.- Farmadi, M. Muliadi, dan R. Herteno, “Application of Extreme Learning Machine Method With Particle Swarm Optimization to Classify of Heart Disease,” *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 17, no. 3, hal. 281, 2023, doi: 10.22146/ijccs.86291.
- [13] A. Giusti, A. W. Widodo, dan S. Adinugroho, “Prediksi Penjualan Mi Menggunakan Metode Extreme Learning Machine (ELM) di Kober Mie Setan Cabang Soekarno Hatta,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 8, hal. 2972–2978, 2018.
- [14] Y. V. Sari, Z. Muallifah, dan A. Fanani, “Klasifikasi Kualitas Air Menggunakan Metode Extreme Learning Machine (ELM),” *J. JUPITER*, vol. 15, no. 2, hal. 983–994, 2023
- [15] I. Riandri, D. Prawira, R. M. #2, N. #3, dan A. Winarko, “Implementasi Algoritma Extreme Learning Machine pada Prediksi Aktivitas Badai Geomagnetik Extreme Learning Machine Algorithm Implementation on Predicting Geomagnetic Storm Activity,” vol. 1, no. 1, hal. 7–13, 2018
- [16] I. W. Saputro dan B. W. Sari, “Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa,” *Creat. Inf. Technol. J.*, vol. 6, no. 1, hal. 1, 2020, doi: 10.24076/citec.2019v6i1.178.
- [17] V. V. Nurdiansyah, I. Cholissodin, dan P. P. Adikara, “Klasifikasi Penyakit Tuberkulosis (TB) menggunakan Metode Extreme Learning Machine (ELM),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 5, hal. 1387–1393, 2020.
- [18] R. J. Rumandan, R. Nuraini, N. Sadikin, dan Y. Rahmanto, “Klasifikasi Citra Jenis Daun Berkhasiat Obat Menggunakan Algoritma Jaringan Syaraf Tiruan Extreme Learning Machine,” *J. Comput. Syst. Informatics*, vol. 4, no. 1, hal. 145–154, 2022, doi: 10.47065/josyc.v4i1.2586.
- [19] D. Normawati dan S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, hal. 697–711, 2021.
- [20] M. N. Hidayat dan R. Pramudita, “Analisis Sentimen Terhadap Pembelajaran Secara Daring Pasca Pandemi Covid-19 Menggunakan Metode IndoBERT,” *Inf. Manag. Educ. Prof. J. Inf. Manag.*, vol. 8, no. 2, hal. 161, 2024, doi: 10.51211/imbi.v8i2.2719.
- [21] R. R. Wahid, F. T. Anggraeny, dan B. Nugroho, “Implementasi Metode Extreme Learning Machine untuk Klasifikasi Tumor Otak pada Citra Magnetic Resonance Imaging,” *Pros. Semin. Nas. Inform. Bela Negara*, vol. 1, no. November 2020, hal. 16–20, 2020, doi: 10.33005/santika.v1i0.45.