Jurnal Pendidikan Vokasi Raflesia
Volume 6 – Nomor 2, Oktober 2026, 328-339
https://doi.org/10.53494/jpvr.v6i2.1521

ISSN 2776-3897 (print)
ISSN 2776-3978 (online)

Available online at:
https://ejournal.polraf.ac.id/index.php/JPVR/article/view/1521

Analisis Perbandingan Kinerja Algoritma Machine Learning pada Bidang Engineering Berdasarkan Karakteristik Dataset Melalui Kajian Sistematis

Jaka Perdana[1*], Hasanah Nur Aini[2], Firda Nabilah[3], Muhammad Yuhda Umama[4], Khalifhan Sadri[5], Azri Anggia Putri[6], Elsa Sabrina[7]
[1,2,3,4,5,6,7] Pendidikan Teknologi Informatika dan Komputer, Universitas Negeri Medan
**Corresponding Author:* jaka.5243151043@mhs.unimed.ac.id

Info Artikel

Sejarah Artikel:
Diterima:
05 Oktober 2026
Direvisi:
13 Oktober 2026
Disetujui:
30 Oktober 2026

Kata Kunci:
Machine Learning, Karakteristik *Dataset*, Kinerja Algoritma, *Engineering*, Kajian Literatur Sistematis.

Abstrak. Kinerja algoritma *machine learning* dalam permasalahan *engineering* dapat bervariasi bergantung pada karakteristik *dataset* dan jenis permasalahan yang dihadapi. Penelitian ini bertujuan untuk membandingkan kinerja berbagai algoritma *machine learning* melalui kajian literatur sistematis. Dari 37 artikel yang diidentifikasi dan disaring, sebanyak 27 artikel relevan dianalisis berdasarkan jenis dan jumlah data, jumlah fitur, *preprocessing*, algoritma, metrik evaluasi, serta kinerja model. Hasil kajian menunjukkan bahwa *dataset* tabular merupakan jenis yang paling dominan, yaitu sebanyak 21 artikel, diikuti oleh dataset citra sebanyak 4 artikel dan *time-series* atau sinyal sebanyak 2 artikel. Algoritma *random forest* menjadi algoritma yang paling banyak digunakan, yaitu dalam 16 artikel. Meskipun demikian, tidak terdapat algoritma yang selalu menjadi yang terbaik karena kinerjanya dipengaruhi oleh struktur *dataset*, target prediksi, distribusi kelas, *preprocessing*, dan jenis permasalahan. Teknik *preprocessing* seperti *feature selection*, *data balancing*, *encoding*, serta transformasi *dataset* juga berperan dalam memengaruhi kinerja model. Secara keseluruhan, tidak ada satu algoritma yang secara konsisten unggul pada seluruh permasalahan *engineering*, sehingga pemilihan algoritma perlu disesuaikan dengan karakteristik *dataset* dan tujuan penelitian.

***How to Cite*:**
Perdana, J., Aini, H. N., Nabilah, F., Umama, M. Y., Sadri, K., Putri, A. A., & Sabrina, E. (2026). Analisis Perbandingan Kinerja Algoritma Machine Learning Pada Bidang Engineering Berdasarkan Karakteristik Dataset Melalui Kajian Sistematis. *Jurnal Pendidikan Vokasi Raflesia*, 6(2), 328-339. https://doi.org/10.53494/jpvr.v6i2.1521

Penerbit:
Politeknik Raflesia
jpvraflesia@gmail.com

PENDAHULUAN

Perkembangan *machine learning* (ML) telah mendorong pemanfaatan pendekatan berbasis data untuk menyelesaikan berbagai permasalahan *engineering* yang membutuhkan prediksi, klasifikasi, pemantauan, dan pengambilan keputusan otomatis. Penerapan ML telah mencakup *geotechnical engineering* untuk klasifikasi tanah dan kestabilan lereng (Yaghoubi et al., 2024), *structural health monitoring* (SHM) untuk klasifikasi kerusakan struktur berdasarkan data sensor (Ahmadian et al., 2024), serta prediksi sifat mekanis beton (Hassan et al., 2024; Inqiad et al., 2024). Selain itu, ML telah digunakan untuk diagnosis kerusakan kendaraan berbasis sinyal voiceprint (Gong et al., 2022), alokasi sumber daya pada jaringan 6G melalui quantum-inspired *machine learning* (Duong et al., 2022), dan prediksi cacat

Copyright (c) 2026 Perdana et al.
This work is licensed under a Creative Commons Attribution 4.0 International License.

aplikasi mobile dalam rekayasa perangkat lunak (Jorayeva et al., 2022). Keragaman penerapan tersebut menunjukkan bahwa ML digunakan pada berbagai domain *engineering* dengan karakteristik permasalahan dan jenis data yang berbeda. Perbedaan karakteristik tersebut kemudian menimbulkan kebutuhan untuk memilih algoritma yang sesuai dengan kondisi data yang digunakan. Dengan demikian, keberhasilan penerapan ML tidak hanya berkaitan dengan algoritma yang digunakan, tetapi juga dengan kesesuaiannya terhadap karakteristik data.

Kesesuaian antara algoritma dan karakteristik data terlihat pada berbagai penelitian yang menunjukkan bahwa algoritma yang berbeda dapat memberikan kinerja yang berbeda pada jenis data tertentu. Ahmadian et al., (2024) menunjukkan bahwa XGBoost efektif untuk seleksi fitur data sensor jembatan, sedangkan metode *stacking* menghasilkan akurasi klasifikasi kerusakan terbaik sebesar 93,1% dibandingkan algoritma konvensional. Pada prediksi kuat tekan *self-compacting concrete*, Inqiad et al., (2024) menemukan bahwa XGBoost lebih unggul dengan nilai R^2 sebesar 0,998 dibandingkan *random forest* sebesar 0,986 dan *multi expression programming* sebesar 0,923. Hassan et al., (2024) melalui kajian terhadap 42 studi prediksi properti beton, melaporkan bahwa tidak terdapat algoritma yang secara konsisten unggul pada seluruh studi. Pada domain yang berbeda, Gong et al., (2022) menemukan bahwa *convolutional neural network* (CNN) memberikan kinerja terbaik pada data *voiceprint* berbasis *linear predictive coding* (LPC), sedangkan *deep neural network* (DNN) lebih unggul pada data berbasis *wavelet transform*. Temuan tersebut menunjukkan bahwa karakteristik data berperan dalam menentukan algoritma yang memberikan kinerja optimal.

Hubungan antara karakteristik *dataset* dan kinerja algoritma juga terlihat dari pengaruh fitur statistik dan *meta-level* terhadap hasil prediksi. Uddin & Lu, (2024) menganalisis 200 *dataset* tabular berdasarkan tiga fitur *meta-level,* yaitu ukuran *dataset*, jumlah atribut, dan rasio kelas, serta empat fitur statistik, yaitu *mean*, standar deviasi, *skewness*, dan *kurtosis*. Hasil penelitian menunjukkan bahwa kurtosis berpengaruh negatif secara signifikan terhadap akurasi algoritma *non-tree-based*, seperti *support vector machine* (SVM), *logistic regression*, dan *k-nearest neighbors* (KNN), sedangkan pengaruh tersebut terbatas pada algoritma *tree-based*. Temuan tersebut menunjukkan bahwa pengaruh karakteristik *dataset* tidak selalu sama pada setiap algoritma. Oleh karena itu, analisis hubungan antara karakteristik *dataset* dan algoritma menjadi penting untuk memahami perbedaan kinerja yang dihasilkan.

Tidak terdapat satu pendekatan ML yang paling sesuai untuk seluruh kondisi data karena kinerja algoritma dapat berubah sesuai dengan karakteristik *dataset* dan konteks permasalahan. Yaghoubi et al., (2024) menemukan bahwa meskipun *artificial neural network* (ANN) merupakan pendekatan yang paling banyak digunakan dalam *geotechnical engineering, ensemble learning* justru menunjukkan kinerja yang lebih baik dalam memprediksi perilaku geoteknik. Jorayeva et al., (2022), melalui kajian terhadap 47 studi prediksi cacat aplikasi mobile, melaporkan bahwa algoritma dengan kinerja terbaik bervariasi antarstudi. Duong et al., (2022) juga menunjukkan bahwa keunggulan *quantum-inspired machine learning* pada jaringan 6G tetap bergantung pada karakteristik data yang digunakan. Dengan demikian, pemilihan algoritma perlu mempertimbangkan karakteristik data dan konteks permasalahan, bukan hanya jenis algoritma yang tersedia.

Meskipun berbagai penelitian telah membahas penerapan dan kinerja ML pada sejumlah domain *engineering*, masing-masing penelitian masih memiliki ruang lingkup yang berbeda. Ahmadian et al., (2024) berfokus pada SHM, Hassan et al., (2024) dan Inqiad et al., (2024) pada material konstruksi, Yaghoubi et al., (2024) pada geoteknik, Jorayeva et al., (2022) pada rekayasa perangkat lunak, dan Gong et al., (2022) pada diagnosis kendaraan. Sementara itu, Uddin & Lu, (2024) secara khusus menganalisis hubungan antara karakteristik *dataset* dan kinerja algoritma, tetapi menggunakan *dataset* tabular umum dan tidak berfokus pada domain *engineering* tertentu. Perbedaan ruang lingkup tersebut

Copyright (c) 2026 Perdana et al.
This work is licensed under a Creative Commons Attribution 4.0 International License.

menunjukkan bahwa temuan mengenai hubungan karakteristik *dataset* dan kinerja algoritma masih tersebar pada berbagai penelitian dan belum memberikan gambaran lintas domain *engineering*. Kondisi tersebut menunjukkan perlunya suatu pendekatan yang dapat mengintegrasikan temuan dari berbagai penelitian secara sistematis.

Kajian sistematis dapat digunakan untuk mengintegrasikan temuan penelitian yang tersebar dan mengidentifikasi pola yang muncul dari berbagai studi. Thompson et al., (2022) menunjukkan bahwa *systematic review* dan *meta-analysis* dapat digunakan untuk menggabungkan hasil dari banyak penelitian dan mengidentifikasi pola metodologis, sebagaimana ditunjukkan dalam kajian terhadap 125 studi deteksi *fake news*. Agar proses kajian menghasilkan temuan yang valid dan dapat direplikasi, tahapan penelitian perlu mengikuti kaidah metodologis yang jelas (Sauer & Seuring, 2023). Oleh karena itu, kajian sistematis dapat digunakan sebagai pendekatan untuk membandingkan kinerja algoritma berdasarkan karakteristik *dataset* pada berbagai domain *engineering*.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan membandingkan kinerja algoritma ML pada berbagai bidang *engineering* berdasarkan karakteristik *dataset* melalui kajian sistematis. Kajian dilakukan terhadap penelitian yang mencakup beberapa domain, yaitu SHM, material konstruksi, geoteknik, rekayasa perangkat lunak, diagnosis kendaraan, dan jaringan komunikasi. Berbeda dengan Uddin & Lu, (2024) yang berfokus pada *dataset* tabular umum serta Yaghoubi et al., (2024) yang berfokus pada satu domain, penelitian ini menempatkan karakteristik *dataset* sebagai dasar perbandingan lintas domain *engineering*. Kebaruan penelitian terletak pada sintesis temuan dari berbagai domain untuk mengidentifikasi keterkaitan antara karakteristik data dan kinerja algoritma sehingga dapat memberikan gambaran yang lebih terpadu mengenai pemilihan algoritma pada permasalahan *engineering* tertentu.

METODE PENELITIAN

Penelitian ini menggunakan metode *Systematic Literature Review* (SLR) untuk mengidentifikasi, mengevaluasi, dan mensintesis penelitian terdahulu mengenai penerapan algoritma *machine learning* di bidang *engineering*. SLR dilakukan melalui proses sistematis dalam pencarian, seleksi, evaluasi, dan sintesis literatur (Sauer & Seuring, 2023). Fokus penelitian adalah membandingkan kinerja berbagai algoritma *machine learning* berdasarkan karakteristik *dataset* yang digunakan.

Artikel ilmiah menggunakan kata kunci *engineering, education, science, machine learning, algorithm,* dan *dataset* dengan bantuan *Publish or Perish, Google Scholar,* dan dataset melalui *Publish or Perish, Google Scholar,* serta beberapa basis data jurnal lainnya.

Proses penyaringan dilakukan secara sistematis. Dari 37 artikel yang teridentifikasi, 27 artikel relevan dipertahankan untuk dianalisis, sedangkan 10 artikel dikeluarkan karena tidak sesuai dengan fokus kajian. Dengan demikian, 27 artikel dinyatakan memenuhi kriteria dan digunakan sebagai sumber informasi penelitian. Proses identifikasi dan seleksi yang terstruktur penting dalam SLR agar pemilihan studi dapat dilakukan secara jelas dan dapat dipertanggungjawabkan (Sauer & Seuring, 2023).

Artikel yang telah melewati tahap penyaringan kemudian dianalisis melalui proses ekstraksi data. Informasi yang dicatat mencakup karakteristik *dataset*, algoritma *machine learning*, lingkup permasalahan *engineering*, metode *preprocessing*, metode pengujian, standar evaluasi, dan hasil kinerja algoritma. Karakteristik *dataset* dianalisis berdasarkan jumlah data, jumlah fitur, jenis data, dan informasi lain yang dilaporkan. Proses ini dilakukan secara teratur untuk memperoleh informasi yang dapat dibandingkan antarpenelitian (Sauer & Seuring, 2023).

Unit sintesis dalam kajian ini adalah artikel dan skenario evaluasi yang dijelaskan di dalamnya. Informasi dikelompokkan berdasarkan domain, bentuk data, karakteristik data,

Copyright (c) 2026 Perdana et al.
This work is licensed under a Creative Commons Attribution 4.0 International License.

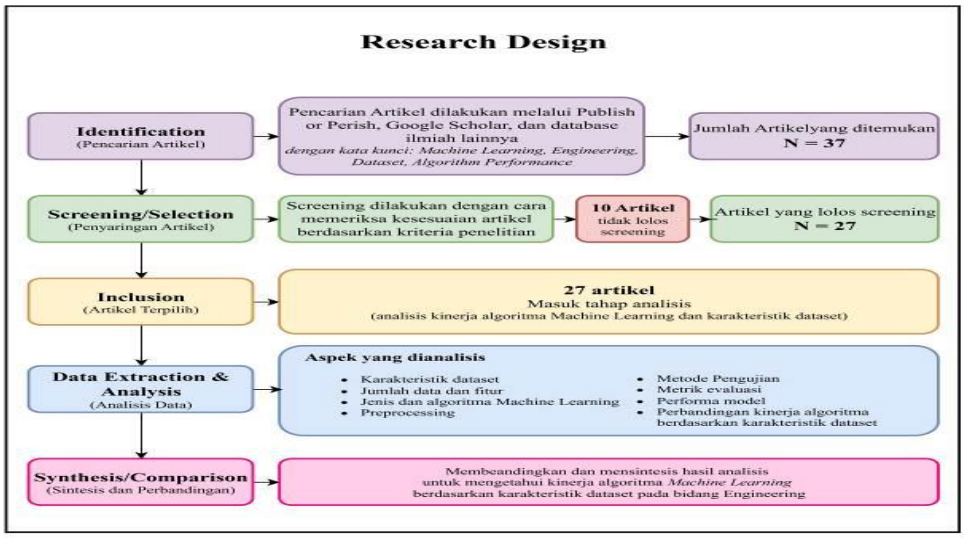
preprocessing, algorotma, metrik, dan model terbaik. Frekuensi dihitung pada tingkat artikel dan tidak dianggap sebagai bukti keunggulan algoritma.

Pada tahap sintesis, data dikategorikan menjadi tabular atau statistik, citra, serta *time-series* atau sinyal. Algoritma dikelompokkan berdasarkan pendekatan. sedangkan *preprocessing* dicatat berdasarkan fungsinya, seperti penanganan kelas tidak seimbang, pengkodean, pemilihan fitur, dan transformasi. Nilai klasifikasi, regresi, deteksi, dan peramalan tidak digabung menjadi satu skor.

Metode ini menghasilkan sintesis naratif dan pengelompokan deskriptif, bukan meta-analisis statistik. Perbedaan domain, target, satuan, dan protokol pengujian antarartikel . Tanggal pencarian, *search string*, rincian eksklusi, dan penilaian risiko bias belum terdokumentasi secara lengkap dan perlu dilengkapi sebelum naskah diajukan.

Tahap berikutnya adalah membandingkan dan menggabungkan hasil dari 27 artikel yang telah dikaji. Penggabungan dilakukan dengan memperhatikan hubungan antara karakteristik *dataset*, metode yang diterapkan, dan hasil algoritma *machine learning*. Analisis bertujuan menemukan pola kinerja algoritma pada karakteristik *dataset* dan permasalahan *engineering* yang berbeda. Dengan demikian, hasil sintesis tidak hanya mengumpulkan penelitian sebelumnya, tetapi juga membandingkan hasil berdasarkan karakteristik data yang digunakan (Sauer & Seuring, 2023).

Secara keseluruhan, tahapan penelitian meliputi pencarian literatur, seleksi artikel, ekstraksi data, analisis karakteristik *dataset* dan algoritma, evaluasi kinerja, serta sintesis hasil. Alur keseluruhan proses SLR disajikan secara kronologis pada Gambar 1.



Gambar 1: Alur desain penelitian

HASIL DAN PEMBAHASAN

Hasil

Sebanyak 27 artikel yang lolos tahap penyaringan dianalisis untuk membandingkan kinerja algoritma *machine learning* berdasarkan karakteristik *dataset* pada berbagai bidang *engineering*. Hasil digunakan untuk melihat pola lintas domain.

Karakteristik *dataset* menunjukkan variasi yang luas, mulai dari data tabular berukuran kecil hingga data jaringan dan perilaku pengguna dengan jumlh observasi lebih

Copyright (c) 2026 Perdana et al.
This work is licensed under a Creative Commons Attribution 4.0 International License.

besar. Variasi ini menjadi dasar untuk menafsirkan perbedaan algoritma dan hasil evaluasi secara kontekstual.

Karakteristik *Dataset* pada 27 Artikel

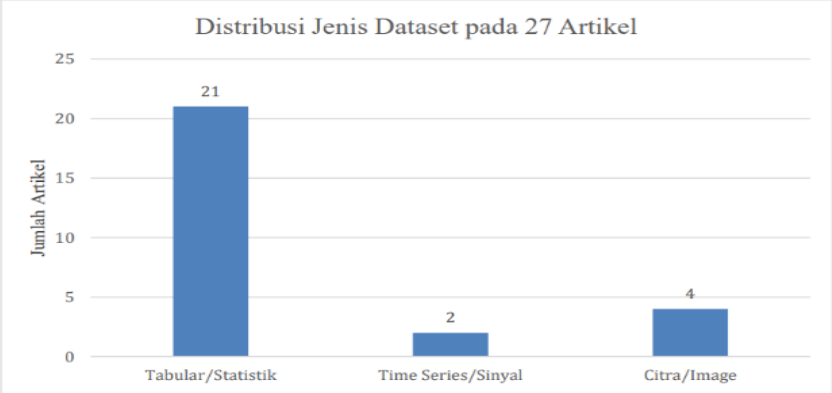
Berdasarkan 27 artikel, *dataset* dikelompokkan menjadi tiga bentuk utama, yaitu tabular atau statistik, time-series atau sinyal, dan citra. Data tabular mendominasi dengan 21 artikel (77,8%), sedangkan time-series atau sinyal ditemukan pada 2 artikel (7,4%) dan citra pada 4 artikel (14,8%). Sebagian besar penelitian engineering menggunakan data berbentuk baris dan kolom. Distribusi jenis dataset dirangkum pada Tabel 1 dan Gambar 2.

Perbedaan bentuk data memengaruhi cara hasil penelitian diinterpretasikan. Dominasi data tabular menunjukkan komposisi korpus kajian, bukan bukti bahwa data tabular selalu lebih unggul.

Tabel 1. Distribusi Jenis *Dataset* pada 27 Artikel

No.	Jenis *dataset*	Jumlah artikel	Proporsi
1	Tabular/statistik	21	77,8%
2	*Time-series*/sinyal	2	7,4%
3	Citra/*image*	4	14,8%
	Total	27	100%

Sumber: Analisis Data Artikel (2026)



Gambar 2: Distribusi jenis *dataset* pada 27 artikel
(Sumber: Hasil pengolahan data artikel, 2026)

Pada kelompok tabular, ukuran *dataset* sangat bervariasi, mulai dari 27 *record* pada penelitian MWCNT-PMMA hingga lebih dari satu juta interaksi pada penelitian *e-commerce*. Beberapa penelitian material menggunakan *dataset* kecil, seperti sekitar 60 sampel pada prediksi kontaminasi air, 120 data pada aluminium, 172 sampel pada *high-entropy alloy*, dan 174 observasi pada kesehatan mesin diesel. Sebaliknya, penelitian keamanan jaringan menggunakan data dalam jumlah besar, seperti 700.001 baris pada UNSW-NB15 dan 117.359 *record* pada Eth-PSD setelah *balancing*. Rangkuman skala data disdajikan pada Tabel 2.

Rentang ukuran dari puluhan hingga ratusan ribu baris menunjukkan bahwa kategori dataset kecil dan besar bersifat relatif terhadap tugas, jumlah fitur, dan desain eksperimen. Oleh karena itu, pada sampel kecil perlu diperhatikan pemisahan data, validasi, dan potensi ketergantungan antara data latih dan data uji. Heterogenitas ukuran dataset lebih tepat dipahami sebagai perbedaan konteks penelitian, bukan dasar untuk menyimpulkan hubungan ukuran sempel dan kinerja model.

Copyright (c) 2026 Perdana et al.
This work is licensed under a Creative Commons Attribution 4.0 International License.

Tabel 2. Ringkasan Skala Data dan Struktur *Dataset*

Kategori	Rentang data	Contoh	Karakteristik utama
Terkecil	27 *record*	MWCNT-PMMA	3 *input* + 2 target
Kecil	±60–174	*Water, aluminium,* HEA*, diesel*	3–10+ variabel
Sedang	916–11.054	Baja, ECC, *elastic properties, phishing URL*	20–221 fitur
Besar	84.664–700.001+	Ethereum*, network intrusion*	4–48 fitur akhir/utama
Citra	1.330–11.791*	*Pothole, cloud, masonry*	Jumlah gambar, anotasi, *slices*
Time-series	3.200–26.304	*Tactile sensor, electricity load*	Urutan waktu/sinyal

Sumber: Hasil sintesis 27 artikel (2026).

Pada kelompok tabular, jumlah fitur bervariasi dari tiga input hingga 221 fitur awal. Beberapa *dataset* berdimensi tinggi mengalami reduksi, seperti *elastic properties* dari 221 menjadi 150 fitur dan *grade prediction* dari 129 menjadi 35 fitur. Pada *software defect*, jumlah atribut berkisar 22–40, sedangkan *network traffic* sekitar 41–48 fitur. Perbedaan jumlah fitur menambah kompleksitas selain ukuran data. Pada data berdimensi tinggi, redundansi dan fitur yang tidak relevan perlu diperhatikan.

Class imbalance juga ditemukan pada beberapa penelitian, terutama *software defect*, *intrusion detection*, *phishing*, dan keamanan IoT. Teknik seperti *oversampling* dan SMOTE digunakan untuk meningkatkan representasi kelas minoritas; pada NSL-KDD, SMOTE meningkatkan *recall* kelas minoritas dari 71,3% menjadi 91,0%. Hal ini menunjukkan bahwa distribusi kelas, bukan hanya ukuran *dataset*, turut memengaruhi kinerja model. *Accuracy* dapat terlihat tinggi meskipun model kurang mampu mengenali kelas minoritas, sehingga *recall*, F1, AUC, dan matriks kebingungan lebih informatif.

Distribusi dan Penggunaan Algoritma

Dari 27 artikel, algoritma dikelompokkan menjadi *tree-based learning*, *ensemble and boosting*, *instance-based learning*, *margin-based learning*, *linear/statistical learning*, *neural network/deep learning*, *probabilistic learning*, dan *computer vision*. *Random forest* menjadi algoritma yang paling sering digunakan, diikuti SVM/SVC dan KNN. Frekuensi tersebut disajikan pada Tabel 3 dan Gambar 3.

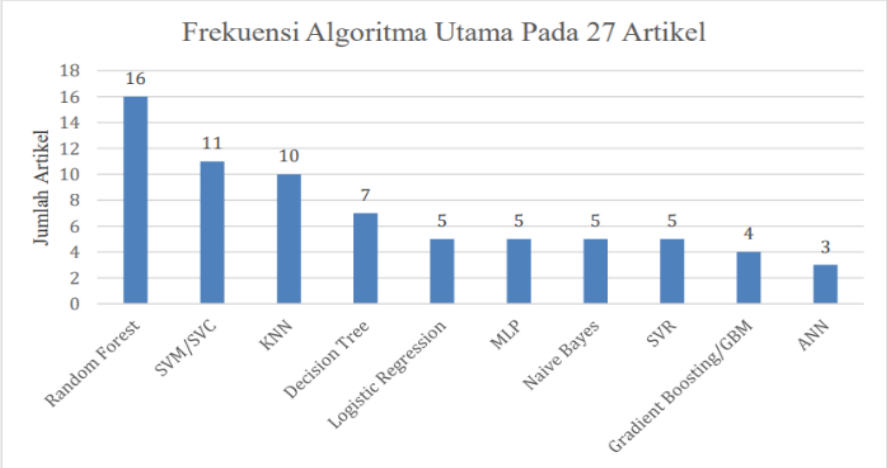
Tabel 3. Frekuensi Algoritma yang Digunakan pada 27 Artikel

No.	Algoritma	Jumlah artikel	Proporsi
1	*Random Forest* (RF)	16	59,3%
2	SVM/SVC	11	40,7%
3	KNN	10	37,0%
4	Decision Tree/J48/C4.5	7	25,9%
5	Logistic Regression	5	18,5%

Copyright (c) 2026 Perdana et al.
This work is licensed under a Creative Commons Attribution 4.0 International License.

No.	Algoritma	Jumlah artikel	Proporsi
6	MLP	5	18,5%
7	Naive Bayes	5	18,5%
8	SVR	5	18,5%
9	Gradient Boosting/GBM	4	14,8%
10	ANN	3	11,1%

Sumber: Hasil sintesis 27 artikel (2026)



Gambar 3: Frekuensi algoritma utama pada 27 artikel
(Sumber: Hasil pengolahan data artikel, 2026)

Random forest digunakan pada 16 dari 27 artikel (59,3%), SVM/SVC pada 11 artikel (40,7%), dan KNN pada 10 artikel (37,0%). Namun, tingginya frekuensi penggunaan tidak menunjukkan bahwa algoritma tersebut selalu memiliki performa terbaik karena hasil dapat berbeda berdasarkan karakteristik data dan target.

Dominasi *random forest* menunjukkan popularitasnya dalam studi yang dikaji, tetapi frekuensi penggunaan dapat dipengaruhi oleh kemudahan penerapan, ketersediaan perangkat lunak, kebiasaan penelitian, dan kesesuaian dengan tugas.

Pada klasifikasi, *random forest*, SVM, KNN, *Decision Tree*, dan *Naive Bayes* banyak digunakan pada *dataset* tabular seperti *software defect, intrusion detection,* dan *phishing*. Pada regresi, algoritma yang digunakan lebih beragam, seperti *random forest Regressor*, *Gradient Boosting*, SVR, *Linear Regression*, dan *neural network*. Data *time-series* banyak menggunakan LSTM, sedangkan data citra menggunakan YOLO, U-Net, dan DeepLabv3+ sesuai karakteristik visual dan spasialnya.

Kesesuaian algoritma bergantung pada jenis tugas dan tujuan evaluasi. Prediksi numerik, klasifikasi, deteksi objek, segmentasi, dan *forecasting* memiliki kebutuhan yang berbeda sehingga model terbaik tidak selalu sama.

Perbedaan performa juga ditemukan pada *dataset* yang sama dengan target berbeda. Pada penelitian aluminium, *random forest* terbaik untuk *Yield Strength*, sedangkan ANN lebih baik untuk *Tensile Strength*. Hal ini menunjukkan bahwa pemilihan algoritma bersifat *dataset-specific* dan dapat pula *target-specific*.

Preprocessing* Berdasarkan Karakteristik *Dataset

Teknik *preprocessing* meliputi *data cleaning*, penanganan *missing value*, normalisasi atau standardisasi, *encoding, feature selection, feature engineering, data balancing*, serta

Copyright (c) 2026 Perdana et al.
This work is licensed under a Creative Commons Attribution 4.0 International License.

transformasi khusus untuk citra dan *time-series*. Teknik tersebut disesuaikan dengan struktur data dan kebutuhan model, sehingga tidak terdapat satu *pipeline preprocessing* yang berlaku untuk seluruh artikel. Teknik *preprocessing* dirangkum pada Tabel 4.

Cleaning dan penanganan *missing value* meningkatkan kualitas data, *encoding* mengubah data kategorikal, *scaling* menyesuaikan skala, *feature selection* memilih atribut yang relevan, sedangkan *augmentation* dan transformasi menyiapkan citra atau sinyal. Dampaknya dapat berbeda antaralgoritma dan *dataset*. Oleh karena itu, peningkatan performa tidak dapat langsung dikaitkan dengan satu teknik, terutama ketika kontribusinya tidak diuji secara terpisah.

Tabel 4. Teknik *Preprocessing* Menurut Karakteristik *Dataset*

Jenis *dataset*	Teknik yang ditemukan	Tujuan utama
Tabular numerik	*Cleaning, missing value, scaling, feature selection*	Meningkatkan kualitas dan kesesuaian data
Tabular kategorikal/campuran	*Cleaning, encoding,* normalisasi	Mengubah atribut kategorikal menjadi representasi numerik
Fitur berdimensi tinggi	*Feature selection, feature engineering*	Mengurangi fitur relevan/redundan dan membentuk representasi yang lebih informatif
Imbalanced	*Oversampling*, SMOTE, *balancing*	Meningkatkan representasi kelas minoritas
Citra	*Annotation, resizing, slicing, augmentation, segmentation*	Menyiapkan citra untuk *computer vision*
Time-series/sinyal	*Cleaning*, standardisasi, *time encoding, sliding window*	Mempertahankan pola temporal

Sumber: Hasil sintesis 27 artikel (2026)

Feature selection terbukti meningkatkan akurasi rata-rata sekitar 5% pada *software defect prediction*, sementara pada IECGA tidak selalu meningkatkan akurasi, tetapi mengurangi waktu eksekusi sekitar 51,52% pada *training* dan 52,31% pada *testing* (Ali et al., 2024). Pada dataset dengan banyak atribut, teknik ini digunakan untuk mengurangi redundansi dan memilih fitur yang sesuai dengan model.

Pada *dataset imbalanced*, *preprocessing* berfokus pada keseimbangan kelas. Eth-PSD menggunakan *oversampling* yang meningkatkan data dari 84.664 menjadi 117.359 *record*, sedangkan NSL-KDD menggunakan SMOTE dan melaporkan peningkatan *recall* kelas minoritas. Data kategorikal menggunakan *One-Hot Encoding* atau *Label Encoding*, sementara data citra dan *time-series* menggunakan *resizing, augmentation, standardisasi,* dan *sliding window*.

Metrik Evaluasi dan Performa Model

Metrik evaluasi disesuaikan dengan jenis permasalahan. Klasifikasi umumnya menggunakan *Accuracy, Precision, Recall,* dan *F1-Score*, regresi menggunakan MAE, RMSE, dan R^2, *forecasting* menggunakan MAPE, sedangkan *object detection* dan *computer vision* menggunakan mAP, IoU, FPS, dan *inference time*. Ringkasan metrik disajikan pada Tabel 5.

Pemilihan metrik memengaruhi interpretasi performa model. *Accuracy* dapat menyembunyikan kesalahan pada kelas minoritas, sedangkan *precision, recall,* dan F1 memberikan informasi yang lebih spesifik. Pada regresi, MAE dan RMSE mengukur kesalahan dengan karakteristik berbeda, sementara R^2 menunjukkan variasi yang dapat dijelaskan model. Oleh karena itu, nilai performa antarjenis masalah tidak dapat disusun dalam satu peringkat.

Copyright (c) 2026 Perdana et al.
This work is licensed under a Creative Commons Attribution 4.0 International License.

Tabel 5. Metrik Evaluasi Menurut Jenis Permasalahan

Jenis masalah	Metrik utama	Fokus evaluasi
Klasifikasi	*Accuracy, Precision, Recall*, *F1-Score*	Penilaian ketepatan klasifikasi dan keseimbangan kesalahan antar-kelas
Regresi	MAE, RMSE, R^2, MAPE	Besarnya *error* dan kesesuaian prediksi terhadap nilai aktual
Object detection	mAP, IoU, *Precision, Recall, F1-Score*	Ketepatan deteksi dan lokasi objek
Semantic segmentation	*Accuracy, Precision, Recall, F1-Score*	Kualitas segmentasi piksel/*region* sesuai target
Real-time/edge	FPS, *inference time*	Kecepatan inferensi dan efisiensi implementasi
Sinyal/*tactile*	*Accuracy, separability, dispersion, confusion matrix*	Keterpisahan kelas dan keberhasilan klasifikasi sinyal

Sumber: Hasil sintesis 27 artikel (2026)

Hasil performa menunjukkan variasi yang kuat antarstudi. Pada data tabular, *random forest* dan *ensemble* sering memperoleh performa tinggi, seperti *accuracy* 93% pada konfigurasi SAP, 97,12%–98,76% pada *intrusion detection*, dan R^2 0,89 pada prediksi kuat tekan batu bata. Namun, pada penelitian kontaminasi air, *Multivariable Linear Regression* lebih konsisten untuk beberapa target, sedangkan SVR unggul pada pH dan *chlorides* (Banerjee et al., 2022). Hal ini menunjukkan bahwa model terbaik pada satu target belum tentu terbaik pada target lainnya.

Perbedaan skor antarartikel tidak selalu menunjukkan perbedaan kualitas algoritma karena dapat dipengaruhi ukuran dan komposisi data, kualitas label, fitur, penanganan kelas, *hyperparameter*, protokol validasi, dan cara pelaporan hasil. Perbandingan algoritma yang lebih langsung memerlukan dataset, pembagian data, *tuning*, dan metrik yang sebanding serta, jika tersedia, pengujian pada dataset eksternal untuk menilai generalisasi.

Pada data *non-tabular*, performa cenderung mengikuti struktur data. Linear SVM mencapai akurasi 99,58% pada sensor taktil, sedangkan 1D CNN mencapai 96,12% pada klasifikasi Braille (Lu et al., 2023). Pada citra, YOLOv5 mencapai mAP@0,5 sebesar 95% untuk *pothole*, DeepLabv3+ memperoleh akurasi 96,86% pada *block detection*, dan U-Net memperoleh F1 79,6% pada *crack detection*. Pada *forecasting* beban listrik, LSTM menghasilkan MAE 11.965,8, RMSE 18.848,5, dan MAPE 6,83%, yang menunjukkan pentingnya mempertahankan pola temporal.

Nilai performa yang tinggi perlu dilihat bersama ukuran *dataset* dan skenario pengujian. *Hybrid GLAR+ANN* memperoleh R=1 pada *grade prediction*, tetapi hasil tersebut terutama menunjukkan *fitting* dan belum membuktikan *generalisasi*. Pada MWCNT-PMMA yang hanya memiliki 27 *record*, R^2 0,99 juga perlu dipahami dalam konteks dataset dan eksperimen yang terbatas.

Hubungan Karakteristik *Dataset* dengan Kinerja Algoritma

Pada data tabular, *random forest* dan *ensemble* berbasis pohon sering digunakan dan pada beberapa studi memberikan performa tinggi. Data citra lebih banyak menggunakan *deep learning* untuk menangkap informasi spasial, sedangkan *time-series* atau sinyal menggunakan model seperti LSTM dan CNN untuk mempertahankan pola temporal.

Hubungan tersebut menunjukkan bahwa karakteristik data memengaruhi *preprocessing* dan pilihan algoritma, sedangkan jenis tugas menentukan metrik evaluasi. Namun, hubungan ini tidak bersifat mutlak karena data tabular tidak selalu menghasilkan *random forest* terbaik dan *time-series* tidak selalu menjadikan LSTM sebagai model terbaik.

Copyright (c) 2026 Perdana et al.
This work is licensed under a Creative Commons Attribution 4.0 International License.

Ukuran *dataset* juga dapat memengaruhi pemilihan model. Pada data kecil, algoritma terbaik dapat berbeda berdasarkan target. Seperti pada penelitian aluminium, hasil model dapat berbeda meskipun sumber datanya sama.

Class imbalance turut memengaruhi performa, terutama pada *cybersecurity*. Pada NSL-KDD, *recall* meningkat setelah SMOTE, sedangkan pada Eth-PSD, KNN mencapai AUC 98,11% setelah proses *balancing*.

Model *ensemble* dan *hybrid* seperti IECGA, FusionNet, dan hybrid LSD menunjukkan performa yang tinggi atau kompetitif dibandingkan *baseline*, tetapi membutuhkan kompleksitas dan sumber daya komputasi lebih besar. Hubungan faktor-faktor tersebut ditunjukkan pada Gambar 4 dan dirangkum pada Tabel 6.



Gambar 4: Kerangka hubungan karakteristik *dataset*, *preprocessing*, algoritma, dan kinerja
(Sumber: Hasil sintesis 27 artikel, 2026)

Tabel 6. Sintesis Hubungan Karakteristik *Dataset* dan Kinerja Algoritma

Kondisi data	**Pola model yang ditemukan**	**Jenis tugas**	**Implikasi sintesis**
Tabular, ukuran dan fitur moderat	RF, DT, SVM, KNN, *boosting*	Klasifikasi/regresi	RF/*ensemble* sering kompetitif
Tabular *imbalanced*	RF/SVM/KNN/*ensemble* + *balancing*	Klasifikasi keamanan	Performa sensitif terhadap distribusi kelas
Fitur banyak	*Feature selection/engineering* + model sesuai target	Regresi/klasifikasi	Reduksi fitur dapat meningkatkan efisiensi dan kadang akurasi
Time-series/sinyal	LSTM, CNN, model temporal	*Forecasting*/klasifikasi sinyal	Representasi urutan menjadi faktor utama
Citra	YOLO, U-Net, DeepLabv3+, CNN	*Object detection/segmentation*	Representasi visual dan spasial menentukan pendekatan
Target ganda	Model berbeda per target	*Multi-output/material properties*	Tidak ada satu model terbaik untuk semua target
Real-time/edge	Konfigurasi/*runtime* efisien	*Deployment*	FPS dan *inference time* dapat mengubah pilihan konfigurasi

Sumber: Hasil sintesis 27 artikel (2026)

Secara keseluruhan, tidak ada satu algoritma yang selalu menjadi yang terbaik pada seluruh kondisi. Perbandingan lintas domain lebih tepat dilakukan berdasarkan pola daripada *ranking* tunggal. Data tabular cenderung menggunakan algoritma *tree-based* dan *ensemble*, data citra menggunakan *computer vision*, sedangkan *time-series* membutuhkan model yang mempertahankan informasi temporal.

Pembahasan

Pola lintas domain menunjukkan bahwa keunggulan algoritma bergantung pada karakteristik data dan tujuan prediksi. Hassan et al., (2024) menemukan bahwa tidak ada algoritma yang selalu unggul pada prediksi properti beton. Yaghoubi et al., (2024)

Copyright (c) 2026 Perdana et al.
This work is licensed under a Creative Commons Attribution 4.0 International License.

menunjukkan bahwa *ensemble learning* dapat memberikan kinerja lebih baik pada geoteknik, sedangkan Jorayeva et al., (2022) menemukan variasi algoritma terbaik pada prediksi cacat aplikasi mobile. Temuan ini mendukung pemilihan model berdasarkan struktur data, target, dan konteks, bukan sekadar frekuensi penggunaan.

Kajian ini melengkapi Uddin & Lu, (2024) dengan mencakup berbagai permasalahan *engineering*, termasuk data tabular, citra, dan *time-series* atau sinyal. Namun, perbedaan cakupan, dataset, validasi, dan metrik membuat hasil tidak dapat dibandingkan secara langsung.

Bagi pendidikan vokasional, pembelajaran *machine learning* perlu dikaitkan dengan permasalahan teknik nyata. Peserta didik dapat dilatih mengenali jenis data, memeriksa kualitas dan distribusi data, memilih metrik, serta membandingkan model. Mereka juga perlu memahami bahwa model yang baik pada satu *dataset* belum tentu berlaku pada data dari alat, lokasi, atau periode berbeda.

Implikasi kurikulum dapat diterapkan melalui kegiatan yang terintegrasi dengan mata pelajaran teknik, mulai dari pengumpulan dan pembersihan data, visualisasi, pengenalan fitur, hingga perbandingan model dan evaluasi menggunakan metrik yang sesuai. Dengan demikian, keterampilan digital dapat mendukung kompetensi pengukuran, pemantauan, diagnosis, dan pemeliharaan.

Bagi praktisi, pemilihan model sebaiknya dimulai dari kebutuhan operasional, seperti jenis keputusan, biaya kesalahan, kualitas data, kebutuhan interpretasi, keterbatasan komputasi, dan waktu respons. Model kemudian diuji dengan prosedur yang sama dan dibandingkan dengan *baseline*.

Kajian ini memiliki keterbatasan berupa jumlah artikel yang hanya 27, sebaran domain yang tidak merata, serta perbedaan dataset, validasi, *tuning*, target, dan metrik. Informasi pencarian dan penilaian risiko bias juga belum terdokumentasi lengkap. Oleh karena itu, penelitian selanjutnya perlu melengkapi protokol pencarian, memperluas korpus, menggunakan *benchmark* yang konsisten, dan menguji generalisasi model pada dataset serta kondisi yang berbeda.

Penerapan model juga perlu memperhatikan tata kelola data dan pemeliharaan model. Perubahan sensor, prosedur, atau kondisi operasi dapat memengaruhi kinerja. Karena itu, hasil kajian sebaiknya menjadi dasar perancangan pengujian, bukan jaminan kinerja produksi.

Dalam mengevaluasi hasil, perlu dibedakan antara kemampuan model mempelajari data eksperimen, kemampuan melakukan generalisasi pada data baru, dan kegunaannya dalam pengambilan keputusan *engineering*. Oleh karena itu, hasil perlu dibaca bersama informasi pembagian data, validasi, *threshold*, kelas minoritas, dan keterbatasan penelitian.

SIMPULAN DAN SARAN

Penelitian ini membandingkan kinerja algoritma *machine learning* pada bidang *engineering* berdasarkan karakteristik *dataset* melalui kajian terhadap 27 artikel. *Dataset* didominasi data tabular sebanyak 21 artikel, diikuti citra sebanyak 4 artikel dan *time-series* atau sinyal sebanyak 2 artikel. *Random forest* paling sering digunakan, yaitu 16 dari 27 artikel (59,3%), tetapi tidak menjadi algoritma terbaik secara universal karena performanya bergantung pada jenis data, target, dan permasalahan. *Preprocessing* juga disesuaikan dengan karakteristik data, seperti *feature selection*, *balancing*, *encoding*, serta transformasi citra dan *time-series*. Secara keseluruhan, kinerja model dipengaruhi oleh keterkaitan karakteristik *dataset*, *preprocessing*, algoritma, dan metode evaluasi sehingga pemilihan model harus disesuaikan dengan konteks permasalahan.

Penelitian selanjutnya disarankan memperluas jumlah artikel dengan melaporkan strategi pencarian dan kriteria seleksi secara transparan serta melakukan validasi dan uji lintas-*dataset* yang sebanding. Praktisi perlu mempertimbangkan bentuk dan ukuran data,

Copyright (c) 2026 Perdana et al.
This work is licensed under a Creative Commons Attribution 4.0 International License.

jumlah fitur, distribusi kelas, target, serta waktu *inference* dalam memilih algoritma dan membandingkannya dengan *baseline* menggunakan metrik yang sesuai. Dalam pendidikan vokasi bidang TIK dan teknik, hasil kajian dapat digunakan untuk mengajarkan pemilihan algoritma, *preprocessing*, dan evaluasi berdasarkan karakteristik data.

DAFTAR PUSTAKA

Ahmadian, V., Beheshti, S. B., Noori, M., Wang, T., & Altabey, W. A. (2024). Engineering Applications of Artificial Intelligence Comparative study of a newly proposed machine learning classification to detect damage occurrence in structures. *Engineering Applications of Artificial Intelligence*, *127*(PA), 107226. https://doi.org/10.1016/j.engappai.2023.107226

Ali, M., Mazhar, T., Al-rasheed, A., & Shahzad, T. (2024). Enhancing software defect prediction : a framework with improved feature selection and ensemble machine learning. *PeerJ Computer Science*, 1–37. https://doi.org/10.7717/peerj-cs.1860

Duong, T. Q., Ansere, J. A., Narottama, B., Sharma, V., Dobre, O. A., & Shin, H. (2022). Quantum-Inspired Machine Learning for 6G : Fundamentals , Security , Resource Allocations , Challenges , and Future Research Directions. *IEEE Open Journal of Vehicular Technology*, *3*(8), 375–387. https://doi.org/10.1109/OJVT.2022.3202876

Gong, C.-S. A., Su, C.-H. S., Liu, Y.-E., Guu, D.-Y., & Chen, Y.-H. (2022). Deep Learning with LPC and Wavelet Algorithms for Driving Fault Diagnosis. *Sensors*, *22*(7072), 1–27. https://doi.org/https://doi.org/10.3390/s22187072

Hassan, S. I., Syed, S. A., Ali, S. W., Zahid, H., Tariq, S., ud, M. M. S., & Alam, M. M. (2024). Systematic literature review on the application of machine learning for the prediction of properties of different types of concrete. *PeerJ Computer Science*, *10*, 1–30. https://doi.org/10.7717/PEERJ-CS.1853

Inqiad, W. Bin, Siddiquea, M. S., Alarifib, S. S., Buttc, M. J., Najehd, T., & Gamile, Y. (2024). Heliyon Comparative analysis of various machine learning algorithms to predict 28-day compressive strength of Self-compacting concrete. *Heliyon*, *9*(11), e22036. https://doi.org/10.1016/j.heliyon.2023.e22036

Jorayeva, M., Akbulut, A., Catal, C., & Mishra, A. (2022). Machine Learning-Based Software Defect Prediction for Mobile Applications : A Systematic Literature Review. *Sensors*, *22*(2551), 1–17. https://doi.org/https://doi.org/ 10.3390/s22072551

Lu, Y., Kong, D., Yang, G., Wang, R., Pang, G., & Luo, H. (2023). Machine Learning-Enabled Tactile Sensor Design for Dynamic Touch Decoding. *Advance Science*, *10*(2303949), 1–9. https://doi.org/10.1002/advs.202303949

Sauer, P. C., & Seuring, S. (2023). How to conduct systematic literature reviews in management research: a guide in 6 steps and 14 decisions. In *Review of Managerial Science* (Vol. 17, Number 5). Springer Berlin Heidelberg. https://doi.org/10.1007/s11846-023-00668-3

Thompson, R. C., Joseph, S., & Adeliyi, T. T. (2022). A Systematic Literature Review and Meta-Analysis of Studies on Online Fake News Detection. *Information*, *13*(527), 1–20. https://doi.org/https://doi.org/10.3390/info13110527

Uddin, S., & Lu, H. (2024). Dataset meta - level and statistical features affect machine learning performance. *Scientific Reports*, *14*(1670), 1–11. https://doi.org/10.1038/s41598-024-51825-x

Yaghoubi, E., Yaghoubi, E., Khamees, A., & Hossein, A. (2024). A systematic review and meta-analysis of artificial neural network , machine learning , deep learning , and ensemble learning approaches in field of geotechnical engineering. In *Neural Computing and Applications* (Vol. 36, Number 21). Springer London. https://doi.org/10.1007/s00521-024-09893-7

Copyright (c) 2026 Perdana et al.
This work is licensed under a Creative Commons Attribution 4.0 International License.