
Studi Perbandingan Analisis: Evaluasi Kinerja Algoritma Klasifikasi pada *Dataset* Terbatas

^{*1}Agus Wantoro, ²Aviv Fitria Yuliana, ³Dita Septasari, ⁴Ikna Awaliyani, ⁵Dwi Yana Ayu Andini

^{1,2,5} Fakultas Teknologi dan Informatika, Universitas Aisyah Pringsewu, Lampung

^{3,4} Fakultas Keguruan dan Ilmu Pendidikan, Universitas Aisyah Pringsewu, Lampung

email: *aguswantoro@aisyahuniversity.ac.id

Abstract

This research aims to evaluate and compare the performance of various classification algorithms under conditions of limited data quantity. Six algorithms were tested, including Naïve Bayes, Support Vector Machine (SVM), Decision Tree, Random Forest, AdaBoost, and k-NN against a small-scale public dataset from different domains. Experiments were conducted using a cross-validation technique (k-fold=5), and evaluation was based on confusion matrix to measure performance in terms of accuracy, precision, recall, F1-score, AUC, and model computation time. The results of the comparison of accuracy, precision, and recall performance show that the Naive Bayes algorithm consistently exhibits optimal performance with a value of 0.896. Timing tests show that the Naive Bayes algorithm demonstrates the fastest time in building the model, while the Random Forest algorithm shows the worst time. The AUC test results indicate that the Naive Bayes algorithm excels, followed by k-NN. Meanwhile, SVM shows the worst AUC performance. Based on the f1-score test, the Random Forest and Naive Bayes algorithms demonstrate the best performance, while the Tree algorithm shows the worst performance. This is because the Naive Bayes algorithm has ease of implementation, speed in calculations, and its ability to work well with large, medium, and limited data, as well as with many features. Each user should choose the algorithm tailored to the data used. In addition, the use of Cross-validation has proven to provide a more reliable performance estimate. These findings offer practical recommendations for researchers and practitioners in selecting effective classification algorithms for small-scale datasets, as well as highlighting the importance of validation techniques and data processing in enhancing model generalization under data limitation conditions.

Keywords : Classification; Multy-algorithm; Small data;

Abstrak

Penelitian ini bertujuan untuk mengevaluasi dan membandingkan performa berbagai algoritma klasifikasi pada kondisi jumlah data yang terbatas. Enam algoritma diuji seperti Naïve Bayes, Support Vector Machine (SVM), Decision Tree, Random Forest, AdaBoost, dan k-NN terhadap dataset publik berskala kecil dari domain yang berbeda. Eksperimen dilakukan menggunakan teknik validasi silang (k-fold=5), serta evaluasi dilakukan berdasarkan *confusion matrix* untuk mengukur kinerja akurasi, presisi, recall, F1-score, AUC, dan waktu komputasi model. Hasil perbandingan kinerja akurasi, presisi dan recall menunjukkan algoritma Naive Bayes secara konsisten menunjukkan kinerja optimal dengan nilai 0.896. Pengujian waktu menunjukkan algoritma Naive Bayes menunjukkan waktu tercepat dalam membangun model dan algoritma Random Forest menunjukkan waktu terburuk. Hasil pengujian AUC menunjukkan algoritma Naive Bayes unggul disusul k-NN. Sedangkan SVM menunjukkan kinerja AUC terburuk. Berdasarkan pengujian f1-score, algoritma Random Forest, dan Naive Bayes menunjukkan kinerja terbaik, sedangkan algoritma Tree menunjukkan kinerja terburuk. Hal ini dikarenakan algoritma Naive Bayes memiliki kemudahan implementasi, kecepatan dalam perhitungan, dan kemampuannya untuk bekerja dengan baik pada data yang besar, sedang, dan terbatas serta banyak fitur. Setiap user hendaknya memilih algoritma disesuaikan dengan data yang digunakan. Selain itu, penggunaan *cross-validation* terbukti mampu memberikan estimasi performa yang lebih andal. Temuan ini memberikan rekomendasi praktis bagi peneliti dan praktisi dalam memilih algoritma klasifikasi yang efektif untuk dataset berskala kecil, serta menyoroti pentingnya teknik validasi dan pemrosesan data dalam meningkatkan generalisasi model pada kondisi keterbatasan data.

Keywords: Dataset kecil; Multi algoritma; Klasifikasi;

1 PENDAHULUAN

Dalam era big data, sebagian besar model klasifikasi seperti *Machine Learning* (ML) mengandalkan jumlah data yang besar agar dapat menghasilkan performa optimal (Singh, Indu, and Majumdar 2025). Namun, pada banyak domain seperti kesehatan, bioinformatika, dan ilmu sosial, seringkali hanya tersedia dataset terbatas (Crossa et al. 2025). Tantangan ini menuntut penggunaan algoritma untuk tetap efektif dalam kondisi *low-data regime*.

Terdapat fenomena umum dimana kinerja klasifikasi menurun drastis saat data berkurang, termasuk rentan terhadap *overfitting* dan variasi tinggi (Williams and Cremaschi 2020). Meskipun berbagai algoritma telah digunakan, masih belum ada studi komprehensif yang membandingkan performa lintas algoritma khususnya dalam skenario data yang sangat terbatas (El Ghaly 2025)

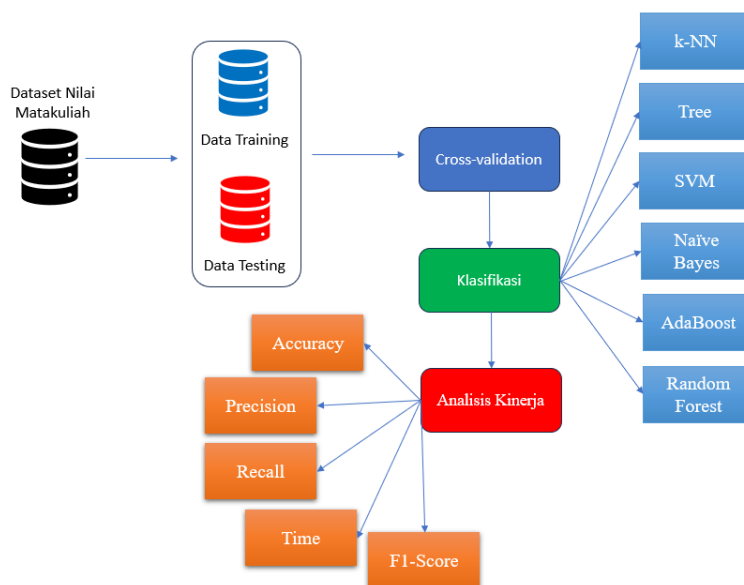
Algoritma seperti SVM, Naïve Bayes, dan Decision Tree sering direkomendasikan untuk data kecil (Deris et al. 2025). Naïve Bayes terbukti cepat dalam estimasi parameter (Vu 2022). ElasticNet dan model linear unggul pada dataset RNA-seq kecil. Transfer learning dan fine-tuning seperti ResNet50 terbukti meningkatkan performa dalam jumlah data terbatas, terutama pada citra medis atau hyperspectral (Ramdan et al. 2020). Penelitian empiris pada berbagai dataset tabular menunjukkan bahwa algoritma ensemble seperti LightGBM, AutoGluon, dan Random Forest sering mendominasi, bahkan di *low-data* (Suenaga et al. 2023)

Studi sebelumnya cenderung fokus hanya pada satu atau dua kategori algoritma (misalnya, linear vs. ensemble). Kurang membandingkan secara menyeluruh dalam kondisi data sangat kecil (<100 sampel) dengan fitur yang beragam. Selain itu, belum menguji dampak teknik validasi seperti *cross-validation* dalam skenario *low-data*. Serta, belum mengeksplorasi kombinasi antara algoritma konvensional dan metode *transfer learning* pada dataset tabular kecil. Oleh karena itu, dibutuhkan *framework* yang sistematis untuk memilih algoritma terbaik menjadi masalah krusial seperti ini. Algoritma yang salah pilih dapat memakan waktu, biaya, dan menghasilkan kesimpulan keliru yang berdampak pada kebijakan atau aplikasi praktis

Penelitian ini bertujuan untuk mengevaluasi dan membandingkan performa algoritma klasifikasi Tree, SVM, k-NN, AdaBoost, Random Forest, dan Naïve Bayes pada dataset kecil yaitu 48 sampel, serta menganalisis estimasi performa dalam kondisi *low-data* serta menghasilkan rekomendasi praktis dalam pemilihan algoritma yang sesuai untuk *dataset* terbatas

2 METODE

Penelitian ini merupakan studi eksperimental komparatif yang bertujuan mengevaluasi kinerja beberapa algoritma klasifikasi pada kondisi jumlah data yang terbatas. Penelitian dilakukan secara kuantitatif menggunakan pendekatan berbasis eksperimen berulang dan *cross-validation*. Desain penelitian diilustrasikan Gambar 1.



Gambar 1. Desain kerangka penelitian

2.1 Dataset

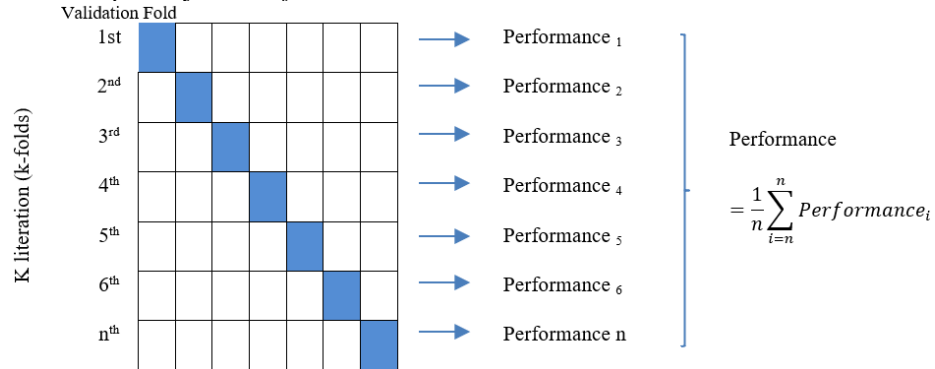
Dataset ini merupakan nilai matakuliah mahasiswa yang berisi 48 sampel, memiliki empat kelas, dan enam fitur. Jumlah kelas pada dataset ini seimbang (12 sampel) untuk masing-masing kelas, sehingga tidak perlu dilakukan penyeimbangan kelas. Tipe data pada fitur beragam yaitu angka dan teks. Dataset diambil pada tahun 2024 melalui link: <https://pelangi.aisyahuniversity.ac.id>. Fitur pada dataset ditampilkan pada Tabel 1.

Tabel 1. Fitur dataset

No.	Fitur	Tipe	Keterangan
1	Nama Mahasiswa	Teks	String
2	Nilai Quiz	Angka	Desimal
3	Nilai Tugas	Angka	Desimal
4	Ujian Tengah Semester	Angka	Desimal
5	Ujian Akhir Semester	Angka	Desimal
6	Indeks	Teks	Char

2.2 Cross-validation (k-folds)

Klasifikasi adalah bentuk teknik penambangan data yang saat ini populer. *Cross-validation* untuk mengevaluasi performa model secara lebih akurat dan menghindari overfitting. Dengan membagi data menjadi beberapa subset dan melakukan pelatihan serta pengujian berulang kali, *cross-validation* memberikan perkiraan kinerja model yang lebih andal, terutama ketika data terbatas (Sulistiani et al. 2024). Model klasifikasi akan divalidasi menggunakan *k-fold cross-validation*. Metode *cross-validation* umumnya digunakan untuk set pelatihan (Yan et al. 2022). Gambar 2 menampilkan jumlah *k-fold*



Gambar 2. Prosedur k-fold validation

2.3 Analisis Kinerja

Studi ini mengkaji bagaimana confusion matrix digunakan untuk mengukur akurasi dan tingkat kesalahan. Confusion matrix dengan ukuran $n \times n$ yang dikaitkan dengan pengklasifikasi, dimana n adalah jumlah total kelas, menampilkan ketegori yang diharapkan (*prediction*) dan aktual (Ohsaki et al. 2017). Confusion matrix untuk $n=2$ ditampilkan dalam Tabel 2.

Tabel 2. Confusion Matrix

Kelas	Prediksi Positiv	Prediksi Negativ
Jumlah positive sebenarnya	Jumlah True Positives (TP)	Jumlah False Negatives (FN)
Jumlah negative sebenarnya	Jumlah False Positives (FP)	Jumlah True Negatives (TN)

Akurasi, prediksi, presisi, recall, dan f1-score adalah metode untuk mengevaluasi dan membandingkan kinerja algoritma klasifikasi. Keempat nilai ini dapat diperoleh dari Confusion matrix pada Tabel 1 dan dihitung menggunakan persamaan (1) sampai (5)

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$f1 - score = 2x \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

$$AUC = (Percent Concordant + 0.5 * Percent Tied)/100 \quad (5)$$

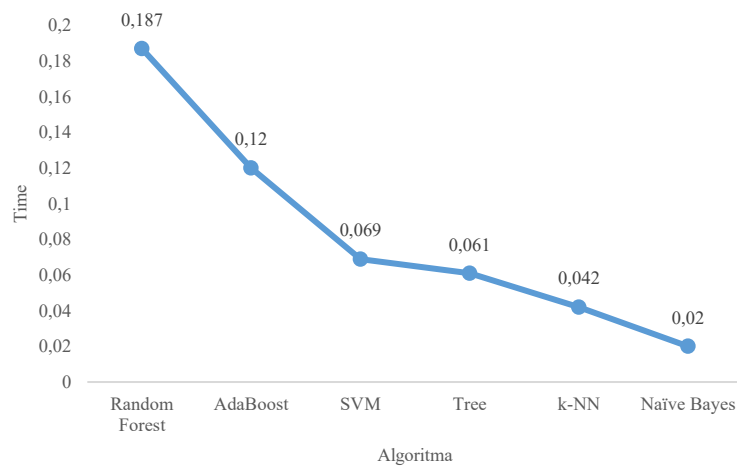
3 HASIL DAN PEMBAHASAN

Kami menggunakan *software* Orange (versi 3.3.8) untuk melakukan analisis perbandingan kinerja. Platform ini mampu menyederhanakan konstruksi beberapa data teknik analisis. Orange memiliki kemampuan mengategorikan, melakukan regresi, klasifikasi, menghilangkan fitur, membuat aturan asosiasi, dan penyesuaian kelas (Popchev and Orozova 2023). Kami menggunakan *cross-validation* (*k-fold=5*) karena temuan kami pilihan ini terbaik terhadap klasifikasi algoritma ML. Hasil kinerja algoritma ditampilkan dalam Tabel 3.

Tabel 3. Perbandingan kinerja akurasi, presisi, dan recall

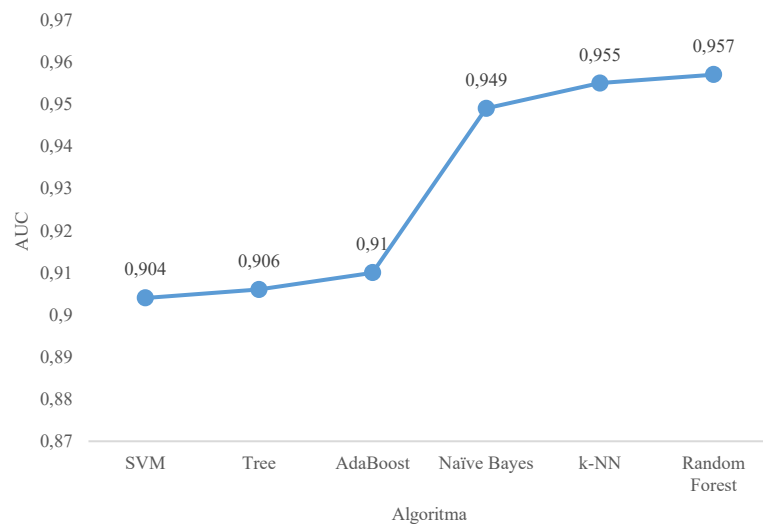
Algoritma	Akurasi	Presisi	Recall	Rata-rata
Tree	0.812	0.808	0.812	0.811
SVM	0.812	0.806	0.812	0.810
k-NN	0.833	0.84	0.833	0.835
AdaBoost	0.875	0.876	0.876	0.876
Random Forest	0.875	0.871	0.875	0.874
Naïve Bayes	0.896	0.895	0.896	0.896

Hasil pengujian akurasi, presisi, dan recall pada Table 4 menunjukkan bahwa algoritma Naive Bayes secara konsisten menjadi yang terbaik disusul algoritma AdaBoost. Hal ini dikarenakan algoritma Naive Bayes mudah diterapkan dan dipahami, efisien dalam komputasi. Selain itu dapat bekerja dengan baik pada data berdimensi tinggi, dan cocok untuk berbagai jenis data kuantitatif maupun kualitatif. Naive Bayes tangguh terhadap data yang hilang dan dapat digunakan untuk klasifikasi biner maupun multi-kelas. Selanjutnya kami melakukan perbandingan kinerja waktu yang dibutuhkan algoritma dalam membangun model yang ditampilkan pada Gambar 3.



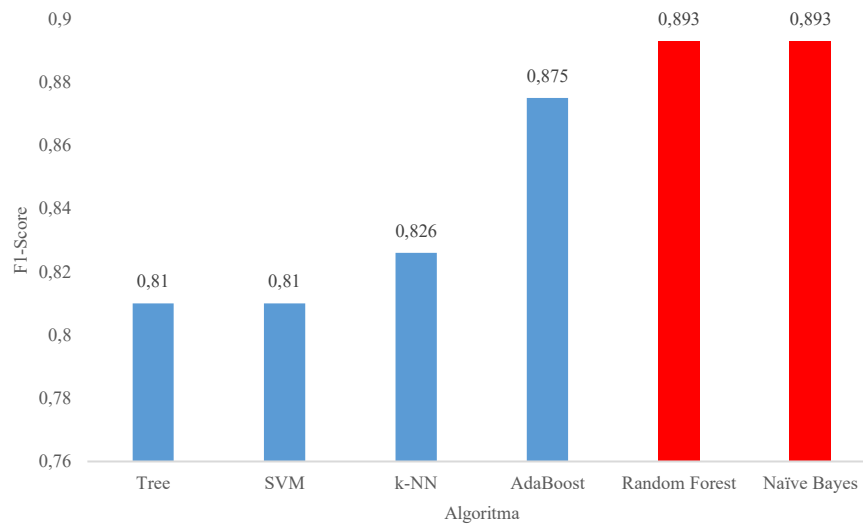
Gambar 3. Waktu komputasi membangun model klasifikasi

Waktu komputasi setiap algoritma menunjukkan waktu berbeda-beda. Algoritma Naive Bayes unggul dalam waktu eksekusi model klasifikasi disusul k-NN, sedangkan algoritma Random Forest menunjukkan waktu terburuk disusul AdaBoost. Selanjutnya kami melakukan perbandingan AUC ditampilkan pada Figure 4.



Gambar 4. Perbandingan kinerja berdasarkan AUC

Hasil analisis perbandingan AUC menunjukkan kinerja algoritma yang berbeda-beda. Algoritma Naive Bayes unggul dalam AUC disusul k-NN. Sedangkan SVM menunjukkan kinerja AUC terburuk. Selanjutnya kami melakukan perbandingan f1-score pada Gambar 5.



Gambar 5. Perbandingan kinerja berdasarkan F1-Score

Berdasarkan pengujian f1-score Gambar 5, algoritma Random Forest dan Naive Bayes menunjukkan kinerja terbaik, sedangkan algoritma Tree menunjukkan kinerja terburuk. Hal ini dikarenakan algoritma Naive Bayes memiliki kemudahan implementasi, kecepatan dalam perhitungan, dan kemampuannya untuk bekerja dengan baik pada data yang besar, sedang dan terbatas serta data dengan banyak fitur (dimensi tinggi). Selain itu, Naive Bayes juga bisa menangani data kuantitatif maupun kualitatif, serta mampu mengatasi data yang hilang. Algoritma Random Forest memiliki beberapa keunggulan utama, di antaranya adalah akurasi tinggi, kemampuan mengatasi *overfitting*, fleksibilitas dalam menangani berbagai jenis data dan tugas (klasifikasi dan regresi), serta kemudahan dalam menangani data berukuran besar dan fitur yang beragam

4 KESIMPULAN

Penelitian ini bertujuan untuk mengevaluasi performa berbagai algoritma klasifikasi pada dataset dengan jumlah data yang terbatas (<100). Dataset dipisah dua bagian yaitu data latih (*training*) dan data uji (*testing*). Persentase data *training* (80%) dan data *testing* (20%). Hasil pengujian menunjukkan persentase tersebut memiliki hasil optimal dari algoritma ML. *Cross-validation* (k-fold=5) terbukti tepat untuk memperoleh estimasi performa yang lebih akurat dan menghindari bias pada saat seleksi model

Berdasarkan hasil eksperimen berupa akurasi, presisi dan recall terhadap dataset berskala kecil dengan tipe fitur yang berbeda, kami menemukan bahwa Naive Bayes secara konsisten menampilkan kinerja terbaik disusul algoritma AdaBoost, sedangkan untuk Algoritma SVM menampilkan kinerja terburuk. Pengujian terhadap waktu komputasi membangun model, algoritma Naive Bayes menunjukkan performa waktu terbaik disusul kNN. Hasil analisis perbandingan AUC menunjukkan kinerja algoritma yang berbeda-beda. Algoritma Naive Bayes unggul dalam AUC disusul k-NN. Sedangkan SVM menunjukkan kinerja AUC terburuk. Berdasarkan pengujian f1-score, algoritma Random Forest dan Naive Bayes menunjukkan kinerja terbaik, sedangkan algoritma Tree menunjukkan kinerja terburuk. Hal ini dikarenakan algoritma Naive Bayes memiliki kemudahan implementasi, kecepatan dalam perhitungan, dan kemampuannya untuk bekerja dengan baik pada data yang besar, sedang dan terbatas serta data dengan banyak fitur (dimensi tinggi). Selain itu, Naive Bayes juga bisa menangani data kuantitatif maupun kualitatif, serta mampu mengatasi data yang hilang. Algoritma Random Forest memiliki beberapa keunggulan utama, di antaranya adalah akurasi tinggi, kemampuan mengatasi *overfitting*, fleksibilitas dalam menangani berbagai jenis data dan tugas (klasifikasi dan regresi), serta kemudahan dalam menangani data berukuran besar dan fitur yang beragam

Studi ini terbatas pada dataset tabular, dan belum dilakukan uji coba untuk data citra, video, dan audio. Belum dieksplorasi metode semi-supervised learning yang bisa berguna di kondisi data kecil. Berdasarkan keterbatasan tersebut, beberapa saran untuk penelitian selanjutnya yaitu (a) Ekspansi ke domain non-tabular, seperti citra dan teks, dengan mengevaluasi perbandingan algoritma klasifikasi di bawah pendekatan few-shot learning atau transfer learning. Eksplorasi model yang didesain khusus untuk data kecil, seperti (b) Prototypical Networks atau Siamese Networks pada konteks NLP atau citra, CatBoost atau TabNet pretrained (c) Integrasi teknik semi-supervised learning untuk memanfaatkan data tak berlabel secara optimal, yang umum terjadi pada studi berbasis data nyata di lapangan (d) Pengembangan framework otomatisasi pemilihan model dan parameter low-data agar pengguna non-teknis dapat

langsung memperoleh solusi klasifikasi optimal (e) Evaluasi lebih lanjut terhadap interpretabilitas model dalam kondisi data kecil, khususnya untuk aplikasi di bidang medis atau keuangan, yang membutuhkan transparansi keputusan.

UCAPAN TERIMA KASIH

Para penulis ingin menyampaikan rasa terima kasih yang mendalam kepada Fakultas Teknologi Pendidikan Universitas Aisyah Pringsewu (UAP), atas akses data penilaian matakuliah mahasiswa yang digunakan dalam studi ini. Kami juga mengucapkan terima kasih kepada Grup Riset Machine Learning (ML) di Fakultas Teknologi dan Informatika UAP atas masukan berharga mereka selama fase pengembangan dan evaluasi model.

DAFTAR PUSTAKA

- Crossa, José et al. 2025. "Machine Learning Algorithms Translate Big Data into Predictive Breeding Accuracy." *Trends in Plant Science* 30(2): 167–84. <https://doi.org/10.1016/j.tplants.2024.09.011>.
- Deris, M Mat et al. 2025. "High Accuracy Data Classification and Feature Selection for Incomplete Information Systems Using Extended Limited Tolerance Relation and Conditional Entropy Approach." *IEEE Access* 13: 27657–69.
- El Ghaly, Abdallah. 2025. "Hybrid ML Algorithm for Fault Classification in Transmission Lines Using Multi-Target Ensemble Classifier with Limited Data." *MDPI* 6(1).
- Ohsaki, Miho et al. 2017. "Confusion-Matrix-Based Kernel Logistic Regression for Imbalanced Data Classification." *IEEE Transactions on Knowledge and Data Engineering* 29(9): 1806–19.
- Popchev, Ivan, and Daniela Orozova. 2023. "Algorithms for Machine Learning with Orange System." *International journal of online and biomedical engineering* 19(4): 109–23.
- Ramdan, A et al. 2020. "Transfer Learning and Fine-Tuning for Deep Learning-Based Tea Diseases Detection on Small Datasets." In *2020 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET)*, , 206–11.
- Singh, Barkha, Sreedevi Indu, and Sudipta Majumdar. 2025. "Comparison of Machine Learning Algorithms for Classification of Big Data Sets." *Theoretical Computer Science* 1024: 114938. <https://www.sciencedirect.com/science/article/pii/S0304397524005553>.
- Suenaga, D et al. 2023. "Prediction Accuracy of Random Forest, XGBoost, LightGBM, and Artificial Neural Network for Shear Resistance of Post-Installed Anchors." *Structures* 50: 1252–63. <https://www.sciencedirect.com/science/article/pii/S2352012423002242>.
- Sulistiani, Heni, Admi Syarif, Kurnia Muludi, and Warsito. 2024. "Performance Evaluation of Feature Selections on Some ML Approaches for Diagnosing the Narcissistic Personality Disorder." *Bulletin of Electrical Engineering and Informatics* 13(2): 1383–91.
- Vu, Duy-Hien. 2022. "Privacy-Preserving Naive Bayes Classification in Semi-Fully Distributed Data Model." *Computers & Security* 115: 102630. <https://www.sciencedirect.com/science/article/pii/S0167404822000293>.
- Williams, Bianca, and Selen Cremaschi. 2020. "Classification Models." *Computer Aided Chemical Engineering*: 5.
- Yan, Tao, Shui-Long Shen, Annan Zhou, and Xiangsheng Chen. 2022. "Prediction of Geological Characteristics from Shield Operational Parameters by Integrating Grid Search and K-Fold Cross Validation into Stacking Classification Algorithm." *Journal of Rock Mechanics and Geotechnical Engineering* 14(4): 1292–1303. <https://www.sciencedirect.com/science/article/pii/S167477522000713>.