

DAFTAR PUSTAKA

- Arifudin, M. A., Setiana, E., & Nugraha, A. B. (2026). Ensemble Learning for Early Warning Systems in Higher Education: A Comparative Study of Student Attrition. *Bulletin of Intelligent Machines and Algorithms*, 1(3), 101–110.
- Arthana, I. K. R., Maysanjaya, I. M. D., Pradnyana, G. A., & Dantes, G. R. (2024). Optimizing Dropout Prediction in University Using Oversampling Techniques for Imbalanced Datasets. *International Journal of Information and Education Technology*, 14(8), 1052–1060.
<https://doi.org/10.18178/ijiet.2024.14.8.2133>
- Baker, R. (2010). Data mining for education. *International encyclopedia of education*, 7(3), 112–118.
- Bowers, A. J., Sprott, R., & Taff, S. A. (2013). Do We Know Who Will Drop Out?: A Review of the Predictors of Dropping out of High School: Precision, Sensitivity, and Specificity. *The High School Journal*, 96(2), 77–100.
<https://doi.org/10.1353/hsj.2013.0000>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.
<https://doi.org/10.1023/A:1010933404324>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321–357.
- Chen, T. (2016). XGBoost: A Scalable Tree Boosting System. *Cornell University*.
- Chicco, D., Tötsch, N., & Jurman, G. (2021). The matthews correlation coefficient (Mcc) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining*, 14.
<https://doi.org/10.1186/s13040-021-00244-z>
- Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine learning research*, 7(Jan), 1–30.
- Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation*, 10(7), 1895–1923.
- Fakih, A., Hamzami, M. A., Hadiano, M. R., & Alifah, N. I. S. (2025). Perbandingan Akurasi Algoritma C4. 5 dan K-NN Untuk Prediksi Kelulusan Mahasiswa Penerima Beasiswa. *Jurnal Komputer Antartika*, 3(1), 18–25.

- Goren, O., Cohen, L., & Rubinstein, A. (2024). Early Prediction of Student Dropout in Higher Education Using Machine Learning Models. *International Educational Data Mining Society*.
- Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? Dalam S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh (Ed.), *Advances in Neural Information Processing Systems* (Vol. 35, hlm. 507–520). Curran Associates, Inc.
https://proceedings.neurips.cc/paper_files/paper/2022/file/0378c7692da36807bdec87ab043cdadc-Paper-Datasets_and_Benchmarks.pdf
- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284.
<https://doi.org/10.1109/TKDE.2008.239>
- Hermawati, M. (2023). Penerapan Algoritma C4.5 dengan Optimasi Particle Swarm Optimization untuk Prediksi Kelulusan Mahasiswa. *Faktor Exacta*, 16(3). <https://doi.org/10.30998/faktorexacta.v16i3.17296>
- Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. *Scientific Reports*, 12(1), 5979.
<https://doi.org/10.1038/s41598-022-09954-8>
- Isphording, I., & Qendrai, P. (2019). Gender differences in student dropout in STEM. *IZA Research Reports*, 87, 1–16.
- J. Ross Quinlan. (1993). C4.5: Programs for Machine Learning. Dalam *Machine Learning Kluwer Academic Publishers, Boston, Manufactured in The Netherlands*.
- Jain, A., Dubey, A. K., Khan, S., Panwar, A., Alkhatib, M., & Alshahrani, A. M. (2025). A PSO weighted ensemble framework with SMOTE balancing for student dropout prediction in smart education systems. *Scientific Reports*, 15(1), 17463. <https://doi.org/10.1038/s41598-025-97506-1>
- Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1). <https://doi.org/10.1186/s40537-019-0192-5>
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Ijcai*, 14(2), 1137–1145.
- Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. Dalam *Progress in Artificial Intelligence* (Vol. 5, Nomor 4).
<https://doi.org/10.1007/s13748-016-0094-0>
- López-Meneses, E., Mellado-Moreno, P. C., Gallardo Herrerías, C., & Pelicano-Piris, N. (2025). Educational Data Mining and Predictive Modeling in the

- Age of Artificial Intelligence: An In-Depth Analysis of Research Dynamics. *Computers*, 14(2). <https://doi.org/10.3390/computers14020068>
- Mduma, N. (2023). Data Balancing Techniques for Predicting Student Dropout Using Machine Learning. *Data*, 8(3). <https://doi.org/10.3390/data8030049>
- Namoun, A., & Alshantqiti, A. (2021). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. Dalam *Applied Sciences (Switzerland)* (Vol. 11, Nomor 1, hlm. 1–28). MDPI AG. <https://doi.org/10.3390/app11010237>
- Pawitra, M. A. S., Hung, H.-C., & Jati, H. (2024). A Machine Learning Approach to Predicting On-Time Graduation in Indonesian Higher Education. *Elinvo (Electronics, Informatics, and Vocational Education)*, 9(2), 294–308. <https://doi.org/10.21831/elinvo.v9i2.77052>
- Permendikti Saintek No. 39 Tahun 2025, Legislation 39, Indonesia, Kementerian Pendidikan Tinggi, Sains, dan Teknologi (2025). <https://peraturan.bpk.go.id/Details/333967/permendikti-saintek-no-39-tahun-2025>
- Putra, H., Nasution, K., & Rilyani, E. (2025). Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Decision Tree: Studi Perbandingan Algoritma Id3 Dan C4. 5. *Jurnal Media Akademik (JMA)*, 3(7).
- Putra, L. G. R., Prasetya, D. D., & Mayadi, M. (2025). Student Dropout Prediction Using Random Forest and XGBoost Method. *INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi*, 9(1), 147–157. <https://doi.org/10.29407/intensif.v9i1.21191>
- Putrama, I. M., & Martinek, P. (2025). SPEAR: Sparse PCA-Enhanced Adaptive Resampling for Multi-class Imbalanced Data. *2025 International Conference on Emerging Systems and Intelligent Computing (ESIC)*, 520–525. <https://doi.org/10.1109/ESIC64052.2025.10962701>
- Putrama, I., & Martinek, P. (2025). Neighbor displacement-based enhanced synthetic oversampling for multiclass imbalanced data. *arXiv preprint arXiv:2501.04099*.
- Putri, D. P. H., Dewi, N. P. N. P., Purnamawan, I. K., & Marti, N. W. (2023). Perbandingan Performansi Support Vector Machine (Svm) dan Backpropagation untuk Klasifikasi Studi Mahasiswa Undiksha. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 9(3), 492–501.
- Putri, S. A., & Rachmatika, R. (2025). Penerapan Algoritma Random Forest dan SMOTE untuk Prediksi Risiko Putus Sekolah Siswa Sekolah Menengah Kejuruan. *Decode: Jurnal Pendidikan Teknologi Informasi*, 5(3), 903–910.
- Santhi, T., Sari, A. M., Putra, D. K. A. M., Mahendra, G. S., & Ariasih, M. P. (2023). Implementasi Business Intelligence Menggunakan Tableau untuk

Visualisasi Prediksi Kelulusan Mahasiswa. *Journal of Software Engineering and Information System (SEIS)*, 51–58.

Satvika, G. A. J., Sukajaya, I. N., & Gunadi, I. G. A. (2024). Improving k-nearest neighbor performance using permutation feature importance to predict student success in study. *Indonesian Journal of Electrical Engineering and Computer Science*, 35(3), 1835–1844.
<https://doi.org/10.11591/ijeecs.v35.i3.pp1835-1844>

Sidik, P., Sunarya, I. M. G., & Gunadi, I. G. A. (2025). Comparison of Random Forest and Support Vector Machine Methods in Sentiment Analysis of Student Satisfaction Questionnaire Comments at ITB STIKOM Bali. *Journal of Applied Informatics and Computing*, 9(3), 794–802.

Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PLoS ONE*, 14(11).
<https://doi.org/10.1371/journal.pone.0224365>

Wang, L., Shi, E., Meyers, B., Vlachos, P., Tchong, J., & Denardo, S. (2026). Machine learning performance for a small dataset: random oversampling improves data imbalances and fairness. *BMC Medical Research Methodology*, 26(1), 67. <https://doi.org/10.1186/s12874-026-02779-3>

Wardani, N. W., Dantes, G. R., & Indrawan, G. (2018). Prediksi customer churn dengan algoritma decision tree C4. 5 berdasarkan segmentasi pelanggan untuk mempertahankan pelanggan pada perusahaan retail. *Jurnal RESISTOR (Rekayasa Sistem Komputer)*, 1(1), 16–24.

Widyanto, A. (2023). Pengaruh Keseimbangan Data terhadap Akurasi Model Support Vector Machine pada Data Set Donor Darah. *Jurnal Teknologi Terpadu*, 9(2), 79–88.

Wolpert, D. H., & Macready, W. G. (1997). No Free Lunch Theorems for Optimization. Dalam *IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION* (Vol. 1, Nomor 1).