



Model Deteksi Mahasiswa Dropout Menggunakan Artificial Neural Network

Denia Igesti Nur Mellyati^{1*}, Kurniabudi², Jasmir³

¹⁻³ Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dinamika Bangsa, Indonesia

Email: igestidenia@gmail.com^{1*}, kurniabudi@unama.ac.id², Jasmir@unama.ac.id³

Alamat: Jl. Jenderal Sudirman, Thehok, Jambi

*Penulis Korespondensi: igestidenia@gmail.com

Abstract. Student dropout remains a significant challenge for higher education institutions as it impacts academic quality, educational management efficiency, and students' success in completing their studies. Therefore, an approach that can identify students at risk of dropping out is necessary so that timely academic interventions can be made. This study aims to develop a dropout detection model using an Artificial Neural Network (ANN). The data used come from a publicly available higher education dataset, ensuring research reproducibility. Data preprocessing steps were carried out to improve data quality before modeling, and the Synthetic Minority Over-Sampling Technique combined with Edited Nearest Neighbors (SMOTE-ENN) was applied to address class imbalance issues. The ANN model's performance was evaluated using accuracy, precision, recall, F1-score, and area under the ROC curve (ROC-AUC). The test results show that the ANN model can provide excellent predictive performance in detecting at-risk students. The application of SMOTE-ENN also proved to enhance the model's sensitivity toward the minority class, as indicated by improvements in recall and F1-score. These findings indicate that the developed ANN model has the potential to be used as a student dropout detection system to support data-driven decision-making and strategy development within higher education institutions.

Keywords: Artificial Neural Network; Dropout Prediction; Higher Education; Machine Learning.

Abstrak. Dropout mahasiswa merupakan permasalahan yang masih banyak dihadapi oleh perguruan tinggi karena berdampak pada kualitas akademik, efisiensi pengelolaan pendidikan, serta keberhasilan mahasiswa dalam menyelesaikan studi. Oleh karena itu, diperlukan suatu pendekatan yang mampu mendeteksi mahasiswa yang berisiko dropout agar intervensi akademik dapat dilakukan secara tepat waktu. Penelitian ini bertujuan untuk membangun model deteksi mahasiswa dropout menggunakan Artificial Neural Network (ANN). Data yang digunakan berasal dari dataset publik pendidikan tinggi yang memungkinkan reproduktibilitas penelitian. Tahapan pra-pemrosesan data dilakukan untuk meningkatkan kualitas data sebelum pemodelan, serta diterapkan teknik Synthetic Minority Over-Sampling Technique yang dikombinasikan dengan Edited Nearest Neighbors (SMOTE-ENN) untuk menangani permasalahan ketidakseimbangan kelas. Kinerja model ANN dievaluasi menggunakan metrik accuracy, precision, recall, F1-score, dan area under the curve (ROC-AUC). Hasil pengujian menunjukkan bahwa model ANN mampu memberikan performa prediksi yang sangat baik dalam mendeteksi mahasiswa berisiko dropout. Penerapan SMOTE-ENN juga terbukti meningkatkan sensitivitas model terhadap kelas minoritas, yang ditunjukkan oleh peningkatan nilai recall dan F1-score. Temuan ini mengindikasikan bahwa model ANN yang dibangun berpotensi digunakan sebagai sistem deteksi dropout mahasiswa guna mendukung pengambilan keputusan berbasis data dan strategi di dalam perguruan tinggi.

Kata kunci: Artificial Neural Network; Dropout Prediction; Higher Education; Machine Learning.

1. LATAR BELAKANG

Fenomena mahasiswa dropout masih menjadi permasalahan yang dihadapi oleh banyak perguruan tinggi dan berdampak langsung pada kualitas pendidikan, efisiensi pengelolaan akademik, serta reputasi institusi pendidikan tinggi. Dropout tidak hanya merugikan mahasiswa secara akademik dan psikologis, tetapi juga menyebabkan pemborosan sumber daya pendidikan yang telah dialokasikan oleh institusi (Kocsis, 2025). Laporan UNESCO

menunjukkan bahwa tingkat dropout pendidikan tinggi secara global masih berada pada kisaran yang signifikan, menjadikan isu ini sebagai tantangan strategis lintas negara (UNESCO, 2025).

Seiring dengan meningkatnya ketersediaan data akademik mahasiswa, pendekatan berbasis data mulai banyak diterapkan untuk mendukung pengambilan keputusan di bidang pendidikan. Machine learning menjadi salah satu pendekatan yang banyak digunakan karena kemampuannya dalam mengidentifikasi pola kompleks pada data pendidikan, yang sulit di tangkap oleh metode analisis konvensional (Yağcı, 2022). Berbagai penelitian menunjukkan bahwa algoritma machine learning mampu memberikan performa yang baik dalam memprediksi performa akademik dan risiko kegagalan studi mahasiswa (Holicza, 2023), (Cho et al., 2023).

Artificial Neural Network (ANN) merupakan salah satu metode machine learning yang banyak digunakan dalam prediksi performa akademik mahasiswa karena kemampuannya dalam memodelkan hubungan nonlinier antar variabel input. ANN dinilai efektif dalam menangkap kompleksitas data pendidikan yang bersifat multidimensional dan heterogen. Namun demikian, implementasi model ANN sebagai sistem deteksi mahasiswa dropout masih memerlukan kajian empiris yang lebih sederhana dan terfokus, terutama pada penerapannya sebagai alat prediksi yang dapat direplikasi dan diadaptasi oleh institusi pendidikan (Rodríguez et al., 2023), (Reznychenko et al., 2024).

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan model Artificial Neural Network dalam mendeteksi risiko mahasiswa dropout menggunakan data akademik mahasiswa. Selain itu, penelitian ini juga membandingkan performa model ANN tanpa penanganan ketidakseimbangan data dan ANN yang dikombinasikan dengan teknik SMOTE-ENN, guna mengevaluasi pengaruh penyeimbangan data terhadap kinerja prediksi. Hasil penelitian diharapkan dapat memberikan gambaran mengenai efektivitas ANN sebagai model deteksi dini dropout mahasiswa serta menjadi dasar bagi pengambilan keputusan akademik berbasis data (Thorat et al., 2022).

2. KAJIAN TEORITIS

Dropout mahasiswa didefinisikan sebagai kondisi ketika mahasiswa tidak menyelesaikan studi sesuai dengan ketentuan kurikulum yang berlaku. Dalam penelitian pendidikan, dropout sering dianalisis sebagai variabel biner yang menunjukkan status keberlanjutan studi mahasiswa. Beberapa penelitian menyebutkan bahwa dropout dipengaruhi oleh kombinasi faktor akademik, sosial, dan ekonomi, yang dapat dipresentasikan melalui data akademik mahasiswa (Kocsis, 2025), (Mnyawami, 2022).

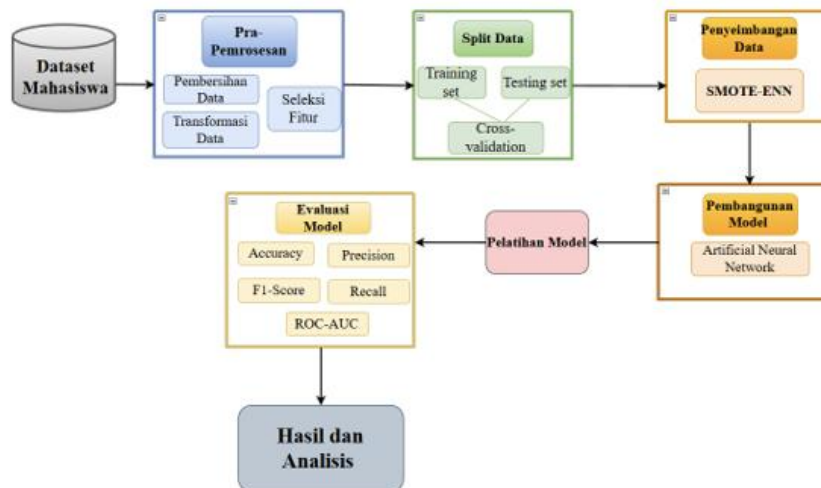
Machine learning merupakan pendekatan komputasional yang digunakan untuk membangun model prediktif berdasarkan pola dalam data. Dalam konteks pendidikan, machine learning telah diterapkan untuk memprediksi performa akademik, waktu kelulusan, serta risiko kegagalan studi mahasiswa (Marbouti, 2021), (Ng, 2018), (Primartha, 2018;). (Yağcı, 2022) dan (Holicza, 2023) menunjukkan bahwa algoritma machine learning mampu mengolah data akademik mahasiswa secara efektif dan menghasilkan model prediksi dengan tingkat akurasi yang baik dibandingkan pendekatan statistik tradisional.

Artificial Neural Network (ANN) adalah salah satu algoritma learning yang terdiri dari lapisan input, lapisan tersembunyi, dan lapisan output yang saling terhubung melalui bobot dan fungsi aktivasi (Samek & Müller, 2019). ANN memiliki kemampuan untuk memodelkan hubungan nonlinier antar variabel dan telah banyak digunakan dalam berbagai studi prediksi akademik (Thorat et al., 2022). (Alturki et al., 2022; Rodríguez et al., 2023) melaporkan bahwa ANN menunjukkan performa yang kompetitif dalam memprediksi capaian akademik mahasiswa.

Pada permasalahan klasifikasi dropout, distribusi data umumnya tidak seimbang karena jumlah mahasiswa yang mengalami dropout lebih sedikit dibandingkan mahasiswa yang tidak dropout. Ketidakseimbangan data dapat mempengaruhi kinerja model klasifikasi khususnya dalam memprediksi kelas minoritas. (Ghorbani & Ghouzi, 2020) mengemukakan bahwa teknik Synthetic Minority Over-Sampling Technique (SMOTE) yang dikombinasikan dengan Edited Nearest Neighbor (ENN) dapat meningkatkan kualitas data latih dan performa model prediksi pada data tidak seimbang. Berdasarkan temuan penelitian terdahulu tersebut, pendekatan ANN dengan dan tanpa penerapan SMOTE-ENN digunakan sebagai landasan konseptual dalam penelitian ini.

3. METODE PENELITIAN

Penelitian ini menggunakan desain penelitian kuantitatif dengan pendekatan eksperimen berbasis machine learning untuk membangun model deteksi risiko mahasiswa dropout. Fokus penelitian adalah penerapan model Artificial Neural Network (ANN) dengan dan tanpa penyeimbangan data menggunakan metode SMOTE-ENN. Alur eksperimen penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur Eksperimen

Alur penelitian pada gambar 1 dijelaskan secara detail sebagai berikut :

Dataset Penelitian

Penelitian ini menggunakan dataset publik dari Kaggle berjudul “Student Habits and Academic Performance Dataset” yang terdiri dari 80.000 catatan dengan 31 atribut numerik dan kategorikal. Dataset ini dipilih karena memungkinkan replikasi, memiliki karakteristik yang relevan untuk analisis risiko dropout (dengan variabel target Dropout Risk), berskala besar dan heterogen, serta memiliki ketidakseimbangan kelas yang signifikan antara mahasiswa dropout dan non-dropout, sehingga sesuai untuk menguji efektivitas metode penyeimbangan SMOTE-ENN.

Pra-Pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk memastikan kualitas data sebelum digunakan dalam proses pemodelan. Proses ini meliputi pembersihan data untuk menangani data hilang dan inkonsistensi, transformasi data agar sesuai dengan kebutuhan algoritma pembelajaran mesin, serta seleksi fitur untuk memilih variabel yang paling relevan dalam memprediksi risiko dropout mahasiswa. Tahapan ini bertujuan untuk meningkatkan kinerja dan stabilitas model yang dibangun.

Pembagian Data

Dataset yang telah dilakukan Pra-pemrosesan selanjutnya dibagi menjadi data latih dan data uji. Data latih digunakan untuk membangun model ANN, sedangkan data uji digunakan untuk mengevaluasi performa model. Untuk meningkatkan kestabilan pembelajaran model, diterapkan skema validasi internal pada data latih tanpa mengubah proporsi data uji.

Penyeimbangan Data

Untuk mengatasi permasalahan ketidakseimbangan kelas, penelitian ini menerapkan metode SMOTE-ENN pada data latih. SMOTE digunakan untuk menghasilkan data sintesis pada kelas minoritas, sedangkan Edited Nearest Neighbor (ENN) digunakan untuk menghapus data yang berpotensi sebagai noise. Penerapan SMOTE-ENN bertujuan untuk meningkatkan kemampuan model dalam mengenali mahasiswa berisiko dropout tanpa menyebabkan bias terhadap kelas mayoritas.

Pembangunan Model

Model klasifikasi dibangun menggunakan Artificial Neural Network (ANN) sebagai algoritma utama penelitian. ANN dipilih karena kemampuannya dalam memodelkan hubungan nonlinier antar variabel serta kinerjanya yang baik dalam berbagai studi prediksi akademik.

Pelatihan Model

Model dilatih menggunakan data latih, baik pada kondisi tanpa penyeimbangan data maupun dengan penerapan SMOTE-ENN, sehingga memungkinkan analisis perbandingan kinerja model.

Evaluasi Model

Evaluasi kinerja model dilakukan menggunakan beberapa metrik evaluasi, yaitu accuracy, precision, recall, F1-Score, dan ROC-AUC. Metrik tersebut digunakan untuk mengukur kemampuan model dalam mengklasifikasikan mahasiswa dropout dan non-dropout secara akurat, khususnya dalam kondisi data tidak seimbang.

4. HASIL DAN PEMBAHASAN

Pada bagian ini dijelaskan hasil kinerja model berdasarkan hasil eksperimen. Kinerja model disajikan menggunakan *confusion matrix*, nilai *Accuracy*, *Precision*, *Recall*, *F1-Score* dan *ROC-AUC*. Bagian ini juga menjelaskan hasil analisis kinerja model.

Kinerja Model Artificial Neural Network

Tabel 1 menyajikan evaluasi matriks pada kinerja ANN baik yang menggunakan teknik penyeimbangan data (SMOTE-ENN) maupun yang tidak.

Tabel 1. Hasil Evaluasi Model ANN Baseline dan Model ANN + SMOTE-ENN

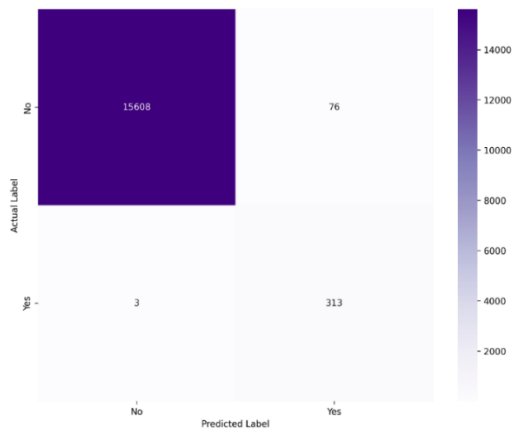
Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
ANN (Baseline)	0.9951	0.8046	0.9905	0.8879	0.9996
ANN + SMOTE-ENN	0.9982	0.9159	1.0000	0.9561	1.0000

Berdasarkan Tabel 1, model Artificial Neural Network (ANN) menunjukkan kinerja prediksi dropout yang sangat baik. Model tanpa penyeimbangan sudah mencapai akurasi 0,9951 dengan *recall* tinggi. Penerapan SMOTE-ENN meningkatkan kinerja model, terutama pada *precision* (naik dari 0,8046 ke 0,9159) dan F1-Score (naik dari 0,8879 ke 0,9561). Nilai ROC-AUC mendekati 1 pada kedua skenario, menunjukkan kemampuan membedakan kelas yang sangat baik.

Analisis Confusion Matrix

ANN (*Baseline*)

Gambar 2, model ANN menyajikan confusion matrix model Artificial Neural Network (ANN) tanpa penerapan penyeimbangan data yang menggambarkan distribusi hasil klasifikasi antara kelas dropout dan non-dropout berdasarkan label aktual dan prediksi model.

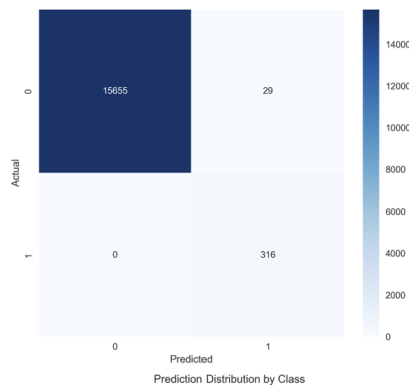


Gambar 2. Confusion Matrix Model ANN (*Baseline*)

Berdasarkan Gambar 2, terlihat bahwa model sangat baik dalam memprediksi kelas mayoritas No-Dropout (True Negative = 15.608), namun kurang sensitif dalam mendeteksi mahasiswa dropout. Hal ini ditunjukkan dengan tingginya False Negative (31), yang berarti banyak kasus dropout yang tidak terdeteksi. Sementara itu, False Positive-nya relatif rendah (76).

ANN + *SMOTE-ENN*

Gambar 3, menunjukkan confusion matrix model Artificial Neural Network (ANN) dengan penerapan metode SMOTE-ENN, yang merepresentasikan perubahan distribusi hasil klasifikasi setelah dilakukan penyeimbangan data pada kelas minoritas.



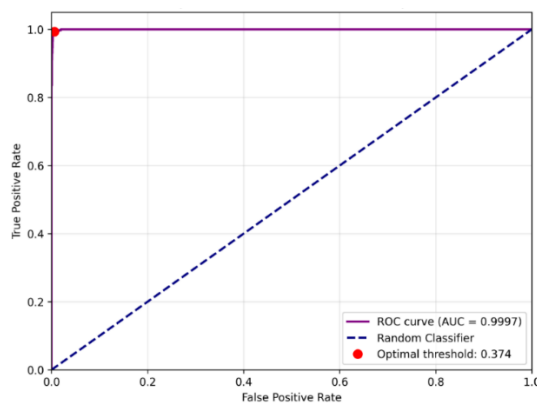
Gambar 3. Confusion Matirx Model ANN + SMOTE-ENN

Gambar 3 menunjukkan, peningkatan sensitivitas ini diikuti kenaikan False Positive (FP) dari 76 menjadi 316, menunjukkan lebih banyak mahasiswa non-dropout yang salah diklasifikasikan sebagai berisiko dropout. SMOTE-ENN berhasil mengatasi masalah deteksi dropout (mengurangi FN), tetapi meningkatkan prediksi alarm palsu (FP). Threshold dapat dioptimalkan untuk menyeimbangkan precision dan recall sesuai kebutuhan penerapan.

Analisis Kurva ROC

ANN (*Baseline*)

Gambar 4, menampilkan kurva Receiver Operating Characteristic (ROC) dari model Artificial Neural Network (ANN) tanpa penerapan penyeimbangan data yang menggambarkan hubungan antara True Positive Rate dan False Positive Rate pada berbagai nilai ambang klasifikasi.

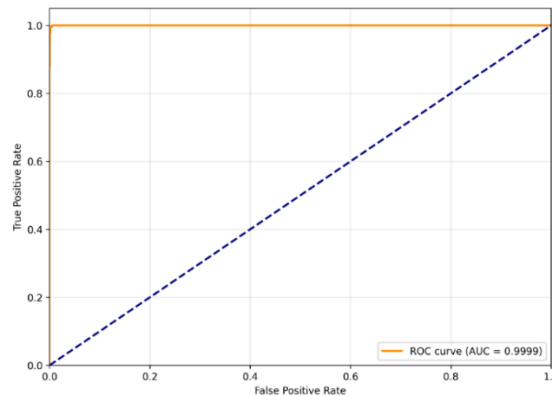


Gambar 4. Kurva ROC Model ANN + SMOTE-ENN

Berdasarkan Gambar 4, model menunjukkan performa luar biasa dengan AUC 0,9996, sangat dekat dengan kurva sempurna di sudut kiri atas. Kurva menunjukkan stabilitas model di berbagai threshold klasifikasi. Titik optimal threshold terletak pada probabilitas 0,413 yang menyeimbangkan kemampuan deteksi dan minimalisasi alarm palsu.

ANN + SMOTE-ENN

Gambar 5, menampilkan kurva Receiver Operating Characteristic (ROC) dari model Artificial Neural Network (ANN) dengan penerapan metode SMOTE-ENN, yang menunjukkan performa klasifikasi model setelah dilakukan penyeimbangan data pada kelas minoritas.



Gambar 5. Confusion Matrix Model ANN + SMOTE-ENN

Berdasarkan Gambar 4, Kurva ROC menunjukkan performa klasifikasi hampir sempurna dengan nilai AUC 0,9999. Kurva yang sangat dekat dengan sudut kira atas plot mengindikasikan kemampuan diskriminasi yang sangat baik dalam membedakan mahasiswa berisiko dropout. Model mencapai *True Positive Rate* tinggi dengan *False Positive Rate* mendekati nol, menandakan sensitivitas dan spesifisitas optimal.

5. KESIMPULAN DAN SARAN

Penelitian ini membangun model deteksi risiko dropout menggunakan ANN dan mengevaluasi efektivitas SMOTE-ENN dalam meningkatkan kinerja model. Hasil menunjukkan performa klasifikasi sangat baik. Pada model baseline, akurasi mencapai 0,9951 dan recall 0,9905. Setelah penerapan SMOTE-ENN, terjadi peningkatan pada precision (dari 0,8046 menjadi 0,9159) dan F1-score (dari 0,8879 menjadi 0,9561), membuktikan efektivitas metode ini dalam menangani data tidak seimbang tanpa mengorbankan kinerja keseluruhan. Penelitian memiliki keterbatasan berupa penggunaan dataset tunggal dari satu sumber publik, sehingga hasil belum dapat digeneralisasi secara luas, serta fokus pada satu algoritma (ANN) tanpa perbandingan dengan metode lain. Untuk penelitian selanjutnya, disarankan menggunakan dataset yang lebih beragam dari berbagai institusi dan membandingkan performa dengan algoritma machine learning lainnya guna mendapatkan model yang lebih kokoh dan komprehensif. Model ini berpotensi untuk dikembangkan sebagai sistem pendukung keputusan deteksi dropout.

UCAPAN TERIMA KASIH

Penulis menyampaikan ucapan terima kasih kepada semua pihak yang telah memberikan dukungan dalam pelaksanaan penelitian ini. Ucapan terima kasih disampaikan kepada dosen pembimbing atas arahan, masukan, dan bimbingan yang diberikan selama proses penelitian dan penyusunan naskah. Penulis juga mengucapkan terima kasih kepada pengelola repositori Kaggle yang telah menyediakan dataset Student Habits and Academic Performance Dataset yang digunakan dalam penelitian ini. Penelitian ini merupakan bagian dari tesis yang disusun oleh penulis sebagai salah satu syarat untuk memperoleh gelar akademik pada program pascasarjana.

DAFTAR REFERENSI

- Alturki, S., Cohausz, L., & Stuckenschmidt, H. (2022). Predicting Master's students' academic performance: An empirical study in Germany. *Smart Learning Environments*, 9(1). <https://doi.org/10.1186/s40561-022-00220-y>
- Cho, C. H., Yu, Y. W., & Kim, H. G. (2023). A Study on Dropout Prediction for University Students Using Machine Learning. *Applied Sciences*, 13(21), 12004. <https://doi.org/10.3390/app132112004>
- Ghorbani, R., & Ghousi, R. (2020). Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques. *IEEE Access*, 8, 67899–67911. <https://doi.org/10.1109/access.2020.2986809>
- Holicza, B. (2023). Predicting and Comparing Students' Online and Offline Academic Performance Using Machine Learning Algorithms. *Behavioral Sciences*, 13(4). <https://doi.org/10.3390/bs13040289>
- Kocsis, Á. (2025). Factors influencing academic performance and dropout rates in higher education. *Oxford Review of Education*, 51(3), 414–432. <https://doi.org/10.1080/03054985.2024.2316616>
- Kotsiantis, S. B. (n.d.). *Supervised Machine Learning: A Review of Classification Techniques*.
- Marbouti, F. (2021). Academic and Demographic Cluster Analysis of Engineering Student Success. *IEEE Transactions on Education*, 64(3), 261–266. <https://doi.org/10.1109/TE.2020.3036824>
- Mnyawami, Y. N. (2022). Comparative Study of AutoML Approach, Conventional Ensemble Learning Method, and KNearest Oracle-AutoML Model for Predicting Student Dropouts in Sub-Saharan African Countries. *Applied Artificial Intelligence*, 36(1). <https://doi.org/10.1080/08839514.2022.2145632>
- Ng, A. (2018). Machine Learning Yearning. *Deeplearning.Ai*.
- Primartha, R. (2018). *Belajar Machine Learning: Teori dan Praktik*. Informatika Bandung.
- Reznichenko, T., Uglickich, E., & Nagy, I. (2024). Accuracy Comparison of Logistic Regression, Random Forest, and Neural Networks Applied to Real MaaS Data. *2024 Smart City Symposium Prague (SCSP)*, 1–5. <https://doi.org/10.1109/SCSP61506.2024.10552715>

- Rodríguez, P., Villanueva, A., Dombrowskaia, L., & Valenzuela, J. P. (2023). A methodology to design, develop, and evaluate machine learning models for predicting dropout in school systems: The case of Chile. *Education and Information Technologies*, 28(8), 10103–10149. <https://doi.org/10.1007/s10639-022-11515-5>
- Samek, W., & Müller, K.-R. (2019). Towards Explainable Artificial Intelligence. In W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, & K.-R. Müller (Eds.), *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* (pp. 5–22). Springer International Publishing. https://doi.org/10.1007/978-3-030-28954-6_1
- Thorat, M., Pandit, S., & Balote, S. (2022). Artificial Neural Network: A brief study. *Asian Journal of Convergence in Technology*, 8(3). <https://doi.org/10.33130/AJCT.2022v08i03.003>
- UNESCO. (2025). *Dropout Rates in Universities Worldwide: Trends and Reasons*. <https://www.educations.com>. <https://www.educations.com/higher-education-news/rising-dropout-rates-in-universities-worldwide-reasons-and-solutions>
- Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(1), 11. <https://doi.org/10.1186/s40561-022-00192-z>