

ANALISA TINGKAT KEHADIRAN SISWA MENGGUNAKAN ALGORITMA DECISION TREE UNTUK MENDETEKSI POLA DAN RISIKO DROP OUT DI SMK SWASTA TUNAS BANGSA

Habi Haikal¹, Hotler Manurung², Siswan Syahputra³
Teknik Informatika, STMIK Kaputama, Binjai
E-mail: *habihaiikal0107@gmail.com¹

ABSTRAK

Tingkat kehadiran siswa merupakan salah satu indikator penting dalam menentukan keberhasilan akademik dan risiko putus sekolah (drop out). Penelitian ini bertujuan untuk menganalisis pola kehadiran siswa serta mengidentifikasi siswa yang berisiko drop out dengan menggunakan algoritma Decision Tree C4.5. Data yang digunakan berasal dari 225 siswa SMK Swasta Tunas Bangsa dengan variabel prediktor berupa kehadiran, nilai akademik, dan nilai sikap, serta label target berupa status drop out (Ya/Tidak). Metode penelitian meliputi tahap pengumpulan data, praproses, pembangunan model pohon keputusan, validasi model dengan teknik k-fold cross validation, serta evaluasi kinerja model menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa atribut Kehadiran menjadi simpul akar (root node) utama pada pohon keputusan, yang kemudian bercabang pada variabel Nilai Akademik dan Nilai Sikap. Model yang dihasilkan mampu mencapai akurasi rata-rata sebesar 85,33%, dengan performa terbaik ditunjukkan pada kelas "Tidak Drop Out" (presisi 96,08% dan recall 90,52%). Sementara itu, pada kelas "Drop Out" recall mencapai 80%, namun presisi masih relatif rendah (60,87%) sehingga menghasilkan sejumlah *false positive*. Sebagai implementasi, penelitian ini juga menghasilkan aplikasi dalam bentuk *Graphical User Interface* (GUI) berbasis MATLAB yang mempermudah pihak sekolah dalam melakukan klasifikasi, validasi, serta menampilkan hasil evaluasi model secara langsung. Dengan demikian, algoritma C4.5 dapat digunakan sebagai sistem peringatan dini untuk membantu sekolah dalam mengidentifikasi siswa yang berisiko drop out, sehingga intervensi preventif dapat dilakukan secara lebih cepat dan tepat sasaran.

Kata kunci

Kehadiran, Drop Out, Decision Tree C4.5, Data Mining, GUI, SMK Tunas Bangsa

ABSTRACT

Student attendance rate is one of the key indicators in determining academic success and the risk of school dropout. This research aims to analyze student attendance patterns and identify students at risk of dropping out using the Decision Tree C4.5 algorithm. The dataset consists of 225 students from SMK Swasta Tunas Bangsa with predictor variables including attendance, academic achievement, and attitude scores, while the target label is dropout status (Yes/No). The research methodology involves several stages: data retrieval, preprocessing, decision tree model construction, model validation using k-fold cross validation, and performance evaluation using accuracy, precision, recall, and F1-score metrics. The results indicate that Attendance is the root node of the decision tree, branching further into Academic Achievement and Attitude variables. The model achieves an average accuracy of 85.33%, with the best performance shown in the "Not Drop Out" class (precision 96.08% and recall 90.52%). Meanwhile, for the "Drop Out" class, the recall reaches 80%, but the precision is relatively low (60.87%), resulting in several false positives. As part of the implementation, this research also develops a Graphical User Interface (GUI) based on MATLAB, which facilitates the classification process, validation, and direct visualization of model evaluation results. Therefore, the C4.5 algorithm can be effectively utilized as an early warning system to help schools identify students at risk of dropping out, enabling faster and more targeted preventive interventions.

Keywords

Attendance, Drop Out, Decision Tree C4.5, Data Mining, GUI, SMK Tunas Bangsa

1. PENDAHULUAN

Kehadiran siswa di sekolah merupakan salah satu indikator penting dalam menilai sejauh mana keterlibatan siswa dalam proses pembelajaran. Dalam wawancara bersama seorang guru di SMK Swasta Tunas Bangsa, tercatat bahwa setiap semester terdapat sekitar 12–15% siswa yang absen lebih dari 20 hari. Ketidakhadiran tersebut tidak selalu berturut-turut, melainkan terjadi sesekali. Sesuai peraturan sekolah, siswa yang memiliki catatan absen lebih dari 8 kali dalam satu semester akan diberikan Surat Peringatan (SP) ke-3, dan apabila tidak ada perbaikan, siswa tersebut dapat dikeluarkan dari sekolah. Kondisi ini menunjukkan adanya pola ketidakhadiran berulang yang signifikan, yang apabila tidak segera ditangani dapat menurunkan prestasi akademik dan menimbulkan risiko putus sekolah (drop out) (Rumberger & Lim, 2009; Sosu et al., 2021).

Tingkat ketidakhadiran tersebut disebabkan oleh berbagai faktor yang kompleks, seperti latar belakang sosial ekonomi (Sosu et al., 2021) atau rendahnya motivasi belajar (Budiningsih, 2025). Menurut Al-Dawood (2020), masalah kehadiran siswa seringkali diperburuk oleh sistem manual yang tidak efisien, yang menyebabkan keterlambatan dalam mengidentifikasi pola absensi. Pendekatan berbasis data, seperti data mining, menjadi solusi untuk memproses informasi kehadiran secara sistematis dan akurat (Han et al., 2022). Salah satu metode yang populer adalah algoritma Decision Tree, yang merupakan teknik klasifikasi efektif dalam data mining (Amna et al., 2023). Dalam penelitian ini, digunakan algoritma C4.5, yang merupakan pengembangan dari ID3, karena kemampuannya dalam menangani atribut kontinu, nilai yang hilang (missing value), dan pruning otomatis (Quinlan, 1993; Mienye & Jere, 2024).

Analisis adalah proses menguraikan suatu pokok permasalahan menjadi bagian-bagian yang lebih kecil untuk dipelajari secara mendalam (Suhendar, 2021). Menurut Wulandari & Sholihin (2024), proses ini mencakup pengelompokan, pembedaan, serta penelaahan hubungan antarbagian untuk memperoleh gambaran yang jelas dan menyeluruh. Siswa didefinisikan sebagai individu yang sedang berada dalam proses pengembangan potensi diri melalui kegiatan pembelajaran di jalur, jenjang, dan jenis pendidikan tertentu (Budiningsih, 2025). Andriani Safitri et al. (2022) memandang siswa dari perspektif karakteristik dan perannya dalam proses pendidikan, yang meliputi gaya belajar, motivasi, dan latar belakang sosial.

Beberapa penelitian terdahulu menunjukkan bahwa algoritma C4.5 menawarkan tingkat akurasi prediksi yang tinggi dalam konteks ketidakhadiran siswa dan kinerja akademik. Misalnya, penelitian oleh Purnama & Apsiswanto (2025) penggunaan algoritma C4.5 pada data siswa SMK menghasilkan akurasi prediksi hingga 98,02%. Selain itu, penelitian oleh Anwar (2025) yang menggabungkan algoritma K-Means dan C4.5 juga berhasil mengidentifikasi risiko drop out siswa dengan tingkat akurasi yang dipengaruhi kuat oleh faktor seperti performa akademik dan latar belakang ekonomi siswa (Anwar, 2025). Hal ini membuktikan bahwa pendekatan Decision Tree sangat relevan dalam konteks pendidikan. Studi lain oleh Surip et al. (2021) juga menekankan bahwa penggunaan C4.5 yang dioptimasi dengan metode lain mampu meningkatkan akurasi dalam mendeteksi siswa putus sekolah.

Data mining adalah proses mengekstraksi pengetahuan atau informasi tersembunyi dari sejumlah besar data. Proses ini mencakup tahapan seperti seleksi data, praproses, transformasi, dan penambangan pola (Han et al., 2022). Menurut Schröer et al. (2021), proses *data mining* mengikuti prosedur standar CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Algoritma C4.5 adalah salah satu algoritma klasifikasi yang banyak digunakan untuk membuat pohon keputusan (*decision tree*). Algoritma ini

merupakan pengembangan dari algoritma ID3 dan dikenal luas karena kemampuannya dalam menangani data numerik, data dengan nilai hilang, serta mekanisme *pruning* otomatis untuk mencegah *overfitting* (Quinlan, 1993; Barutu, 2024).

Penelitian yang dilakukan oleh Muhamad, Windarto, dan Suhada (2019) menunjukkan bahwa penerapan algoritma C4.5 efektif dalam mengklasifikasikan potensi siswa yang berisiko *drop out* dengan akurasi 96,15%. Menurut Surip et al. (2021), penggunaan algoritma C4.5 yang dioptimasi dengan *Particle Swarm Optimization* (PSO) mampu meningkatkan akurasi dalam menganalisis data siswa putus sekolah hingga 90,86%. Ermillian & Nugroho (2024) juga menyimpulkan bahwa metode *Decision Tree* memberikan visualisasi yang mudah dipahami dalam deteksi potensi siswa putus sekolah. Selanjutnya, Fatma & Rochmawati (2024) mendapatkan akurasi 92% dan presisi 84% saat memprediksi siswa putus sekolah menggunakan C4.5, sementara Lestari & Dina (2023) menggunakannya untuk mengklasifikasikan ketepatan kelulusan siswa.

Penelitian ini bertujuan untuk mengembangkan model klasifikasi dengan algoritma Decision Tree C4.5 untuk mendeteksi siswa yang berisiko mengalami drop out di SMK Swasta Tunas Bangsa. Hasilnya diharapkan dapat membantu sekolah dalam menyusun strategi intervensi yang lebih tepat sasaran, baik dari sisi bimbingan konseling, dukungan sosial, maupun bantuan finansial. Dengan demikian, penelitian ini berjudul “Analisa Tingkat Kehadiran Siswa Menggunakan Algoritma Decision Tree untuk Mendeteksi Pola dan Risiko Drop Out di SMK Swasta Tunas Bangsa” diharapkan dapat memberikan kontribusi nyata dalam bidang pendidikan, khususnya dalam pemanfaatan teknologi data mining untuk mendeteksi secara dini siswa yang berisiko putus sekolah. Melalui hasil analisis yang diperoleh, sekolah diharapkan mampu merancang intervensi yang lebih tepat sasaran guna meningkatkan kehadiran dan mengurangi angka drop out.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan algoritma *Decision Tree* C4.5. Alur penelitian mengacu pada tahapan dalam metode *Knowledge Discovery in Databases* (KDD), yaitu proses sistematis dalam menggali pengetahuan dari data (Han et al., 2022). Metodologi penelitian yang digunakan juga sejalan dengan kerangka kerja yang sistematis untuk menjamin validitas dan reliabilitas hasil (Ojoboh & Igben, 2024; Susilawati et al., 2025).

a. Pengumpulan Data

Data penelitian diperoleh dari 225 siswa SMK Swasta Tunas Bangsa. Variabel yang dikumpulkan adalah Kehadiran, Nilai Akademik, Nilai Sikap, dan Status Drop Out (Ya/Tidak). Pemilihan variabel ini didasarkan pada penelitian sebelumnya yang menunjukkan bahwa kehadiran dan prestasi akademik merupakan indikator penting dalam mendeteksi risiko putus sekolah (Purnama & Apsiswanto, 2025).

b. Pra-pemrosesan Data (Preprocessing)

Tahap ini meliputi *Data Cleaning* (menghapus duplikasi), *Handling Missing Values* (mengisi nilai kosong), dan *Normalization & Categorization* (transformasi data numerik ke kategori). Tahap ini penting untuk memastikan dataset bersih dan konsisten (Lestari & Rochmawati, 2024).

c. Klasifikasi (Data Mining)

Tahapan inti penelitian adalah penerapan algoritma *Decision Tree* C4.5. Proses pembentukan pohon keputusan dilakukan dengan langkah-langkah berikut:

- 1) Hitung Entropy dari dataset:

$$Entropy(S) = \sum_i^c -p_i \log_2 p_i.$$

di mana p_i adalah proporsi data dari kelas ke- i .

2) Hitung Information Gain untuk setiap atribut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i).$$

3) Hitung Gain Ratio untuk setiap atribut:

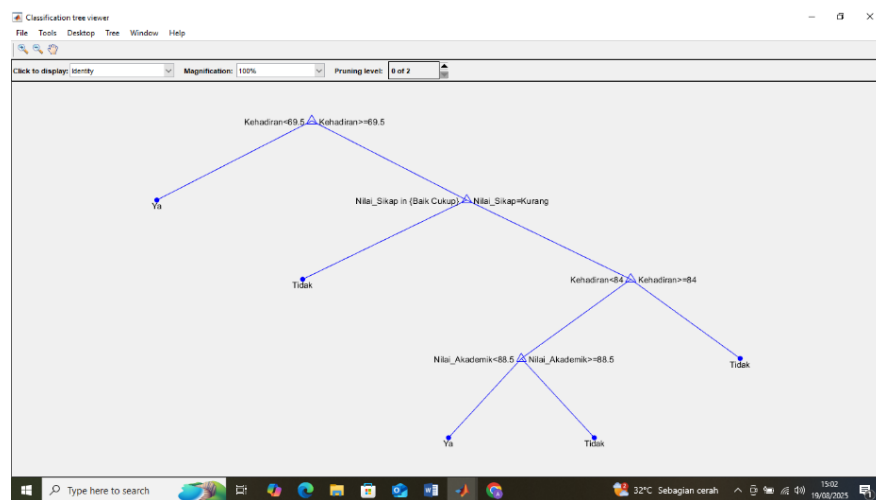
$$Gain\ Ratio = \frac{Gain}{SplitInfo}$$

4. HASIL DAN PEMBAHASAN

Berdasarkan analisis tingkat kehadiran siswa menggunakan algoritma Decision Tree C4.5, diperoleh hasil sebagai berikut:

a. Pohon Keputusan Algoritma C4.5

Hasil implementasi algoritma C4.5 menghasilkan sebuah pohon keputusan dengan Kehadiran sebagai atribut utama pada simpul akar. Siswa dengan tingkat kehadiran rendah cenderung diklasifikasikan sebagai *Drop Out*, sementara siswa dengan tingkat kehadiran tinggi masuk kategori *Tidak Drop Out*.

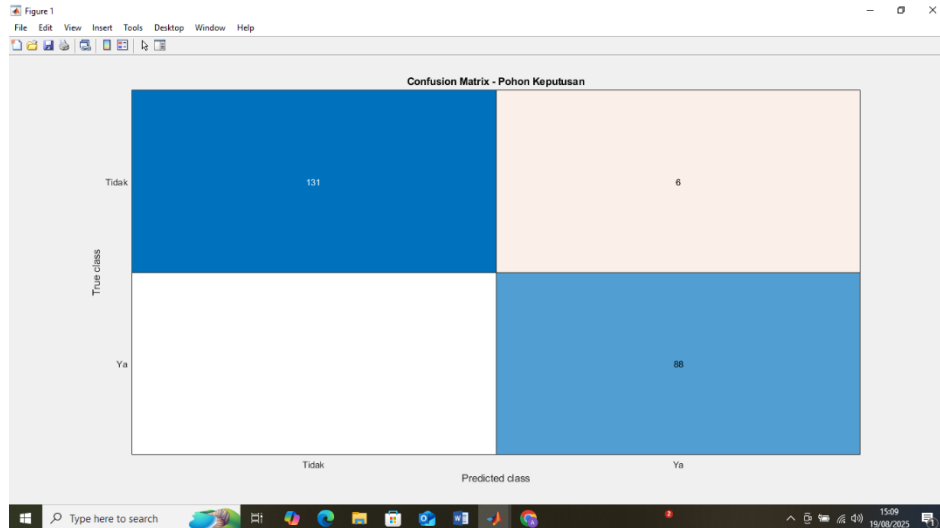


Gambar 1. Pohon Keputusan

b. Evaluasi Performa Model (Confusion Matrix)

Pengujian menggunakan teknik 10-fold cross validation menghasilkan akurasi model sebesar 85,33%, Confusion matrix menunjukkan:

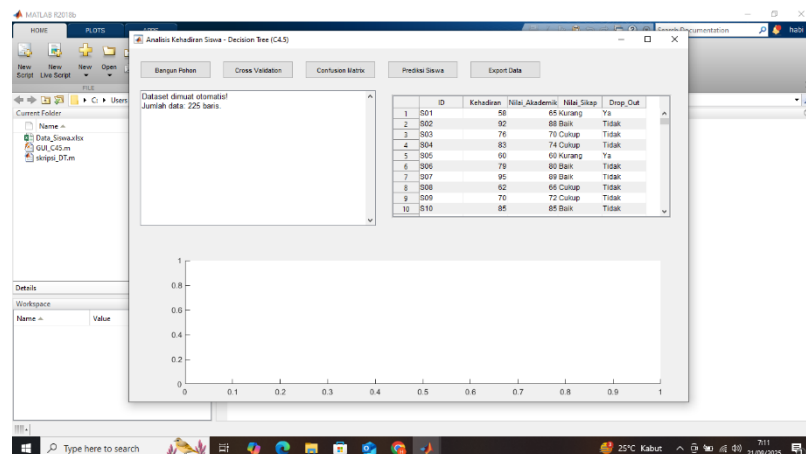
- 1) 190 siswa diklasifikasikan benar sebagai *Tidak Drop Out*
- 2) 32 siswa diklasifikasikan benar sebagai *Drop Out*
- 3) 20 siswa salah diklasifikasikan sebagai *Drop Out* (false positive)
- 4) 8 siswa salah diklasifikasikan sebagai *Tidak Drop Out* (false negative)



Gambar 2. Confusion Matrix

c. Implementasi Aplikasi GUI

Sebagai implementasi praktis, model dikembangkan dalam bentuk Graphical User Interface (GUI) berbasis MATLAB. GUI ini menampilkan input data siswa, hasil pohon keputusan, confusion matrix, serta nilai akurasi model. Dengan adanya GUI, pihak sekolah dapat dengan mudah menganalisis data siswa tanpa perlu melakukan perhitungan manual.



Gambar 3. Tampilan GUI

4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma Decision Tree C4.5 efektif digunakan untuk menganalisis tingkat kehadiran siswa dalam mendeteksi pola dan risiko *Drop Out* di SMK Swasta Tunas Bangsa, dengan tingkat akurasi mencapai 85,33%. Faktor Kehadiran terbukti menjadi penentu utama status siswa, diikuti oleh Nilai Akademik dan Nilai Sikap. Hasil ini mengindikasikan bahwa pola absensi dapat dijadikan indikator awal untuk mendeteksi risiko *Drop Out*. Namun, performa model masih lebih baik pada kelas mayoritas (*Tidak Drop Out*) dibanding kelas minoritas (*Drop Out*). Oleh karena itu, penelitian selanjutnya disarankan menambahkan variabel pendukung lain seperti motivasi belajar dan faktor ekonomi, serta menerapkan metode balancing data agar hasil klasifikasi lebih merata.

5. DAFTAR PUSTAKA

- Al-Dawood, A. (2020). The Impact of Manual and Automated Attendance System on Student Performance. *International Journal of Modern Education and Computer Science*, 12(1), 1-10.
- Amna, S. W., Sudipa, I. G. I., Putra, T. A. E., Wahidin, A. J., Sukrilla, W. A., Wardhani, A. K., Heryana, N., Indriyani, T., & Santoso, L. W. (2023). Data Mining. In *Mining of Massive Datasets*. PT GLOBAL EKSEKUTIF TEKNOLOGI.
- Anwar, S. (2025). Application of K-Means and C4.5 Algorithms for dropout risk prediction in vocational high schools. *Indonesian Journal of Multidisciplinary Science*, 4(5), 428–435. <https://doi.org/10.55324/ijoms.v4i5.1102>
- Auliani, R., Siregar, S. M., Zein, A., & Siregar, R. M. (2025). Konsep Deteksi Dini Tumbuh Kembang Anak Usia Dini. 9, 1668–1673.
- Barutu, S. (2024). Application Of C4.5 Algorithm In Disease Classification. *Journal Of Data Science*, 2(02), 58–62. <https://doi.org/10.58471/jds.v2i02.5263>
- Billa, C., Husaini, A., Kuliah, M., Risiko, M., Studi, P., Syariah, E., Islam, U., Sulthan, N., & Saifuddin, T. (2023). Pemahaman Resiko Dan Manajemen Resiko. 1(3).
- Budiningsih, R. (2025). Pengaruh Motivasi Belajar terhadap Tingkat Kehadiran dan Prestasi Siswa SMK. *Jurnal Pendidikan dan Pengajaran*, 14(1), 101–110.
- Ermilian, A., & Nugroho, K. (2024). Perancangan Model Deteksi Potensi Siswa Putus Sekolah Menggunakan Metode Logistic Regression Dan Decision Tree. *Jurnal Informatika: Jurnal Pengembangan IT*, 9(3), 281–295. <https://doi.org/10.30591/jpit.v9i3.8007>
- Fatma, Y. L., & Rochmawati, N. (2024). Prediksi Siswa Putus Sekolah Menggunakan Algoritma Decision Tree C4.5. *Journal of Informatics and Computer Science (JINACS)*, 5(04), 486–493. <https://doi.org/10.26740/jinacs.v5n04.p486-493>
- Han, J., Kamber, M., & Pei, J. (2022). *Data Mining: Concepts and Techniques* (4th ed.). Morgan Kaufmann.
- Islam, U., Sumatera, N., Matematika, P., Islam, U., & Sumatera, N. (2023). Konsep Dasar Algoritma Pemrograman. 1(6).
- Lestari, S., & Dina, B. (2023). Klasifikasi Ketepatan Kelulusan Siswa Pada Smk Yadika 9 Bintara Jaya Kota Bekasi Menggunakan Algoritma C4.5. *Jurnal Indonesia: Manajemen Informatika Dan Komunikasi*, 4(3), 1618–1632. <https://doi.org/10.35870/jimik.v4i3.389>
- Lestari, N., & Rochmawati, I. (2024). Data Preprocessing untuk Peningkatan Kualitas Dataset dalam Proses Knowledge Discovery in Databases (KDD). *Jurnal Teknologi dan Sistem Informasi*, 10(3), 233–241.
- Mienye, I. D., & Jere, N. (2024). A Survey of Decision Trees: Concepts, Algorithms, and Applications. *IEEE Access*, 12, 86716–86727. <https://doi.org/10.1109/ACCESS.2024.3416838>
- Muhamad, M., Windarto, A. P., & Suhada, S. (2019). Penerapan Algoritma C4.5 Pada Klasifikasi Potensi Siswa Drop Out. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 3(1), 753–760. <https://doi.org/10.30865/komik.v3i1.1688>
- Ojoboh, T. M., & Igben, H. G. O. (2024). Impact of Research Methodology on Data Quality and Research Findings. *JPPUMA Jurnal Ilmu Pemerintahan Dan Sosial Politik Universitas Medan Area*, 12(1), 34–42. <https://doi.org/10.31289/jppuma.v12i1.11793>
- Pratama, F. W. (2022). Tingkat Pemahaman Konsep Siswa pada Materi Pola Bilangan Berdasarkan Teori APOS Mosharafa: *Jurnal Pendidikan Matematika*. Mosharafa:

- Jurnal Pendidikan Matematika, 11, 237–246.
- Purnama, D. L. S., & Apsiswanto, U. (2025). Analysis of C4.5 Algorithm Performance for Predicting Student Achievement Based on Socio-Economic Status, Motivation, Discipline, and Past Achievement. *Journal of Computer Networks, Architecture and High Performance Computing*, 7(1), 190–199. <https://doi.org/10.47709/cnahpc.v7i1.5143>
- Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann.
- Rosnani, S. (2023). Pengaruh Kehadiran Dosen dan Proses Pembelajaran Terhadap Prestasi Akademik Mahasiswa. *Makalah Ilmiah*.
- Rumberger, R. W., & Lim, S. A. (2009). Why students drop out of school. *Journal of Education for Students Placed at Risk*, 14(4), 519–540.
- Saparwadi, L. (2023). Peningkatan Prestasi Belajar Siswa Melalui Pendekatan Kontekstual. *Jurnal Pendidikan Dasar FONDASI*, 2(1), 22–29.
- Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181(2019), 526–534. <https://doi.org/10.1016/j.procs.2021.01.199>
- Sosu, E. M., Shapira, M., & Priestley, M. (2021). Patterns of School Attendance and Impact on Student Outcomes: A Longitudinal Study. *Journal of Educational Research*, 114(5), 589–604. <https://doi.org/10.1080/00220671.2021.1877372>
- Suhendar, B. (2021). Analisa Sistem Informasi Manajemen Berbasis Komputer Dalam Proses Sistem Pengambilan Keputusan (Spk). *Jurnal Ilmiah Sistem Informasi*, 1(1), 46–56. <https://doi.org/10.46306/sm.v1i1.7>
- Surip, A., Pratama, M. A., Ali, I., Dikananda, A. R., & Purnamasari, A. I. (2021). Penerapan Machine Learning menggunakan algoritma C4.5 berbasis PSO dalam Menganalisa Data Siswa Putus Sekolah. *INFORMATICS FOR EDUCATORS AND PROFESSIONAL: Journal of Informatics*, 5(2), 147. <https://doi.org/10.51211/itbi.v5i2.1530>
- Susilawati, A., Al Obaidi, A. S. M., Abduh, A., Irwansyah, F. S., & Nandiyanto, A. B. D. (2025). How to do research methodology: From Literature Review, Bibliometric, Step-by-step Research Stages, to Practical Examples in Science and Engineering Education. *Indonesian Journal of Science and Technology*, 10(1), 1–40. <https://doi.org/10.17509/ijost.v10i1.78637>
- Wulandari, N., & Sholihin, H. (2024). Analisis Komponen Utama Dalam Sistem Informasi Manajemen: Konsep, Fungsi, Dan Implementasi. *Jurnal Ilmiah Pendidikan Dasar*, 9(4), 872–890.
- Zalukhu, A., Singly, P., & Darma, D. (2023). Perangkat Lunak Aplikasi Pembelajaran Flowchart. *Jurnal Teknologi, Informasi Dan Industri*, 4(1), 61–70.
- Zhao, L. (2021). Event Prediction in the Big Data Era: A Systematic Survey. *ACM Computing Surveys*, 54(5), 1–37. <https://doi.org/10.1145/3450287>