
DIGITALISASI PENDIDIKAN: ANALISIS DAN PREDIKSI POLA PEMBELAJARAN DIGITAL PADA LMS EDLINK MENGGUNAKAN ALGORITMA NAIVE BAYES

Zamzami¹, Didik Siswanto², Lasri Nijal³
Universitas Lancang Kuning, Pekanbaru

e-mail: ¹ zamzami@unilak.ac.id , ² didik@unilak.ac.id , ³ lasrinijal@unilak.ac.id

Abstract: *Learning Management Systems (LMS) produce large volumes of learning activity data, yet at Universitas Lancang Kuning these data have so far been analysed only descriptively. This study builds a Naive Bayes model to classify and predict digital learning patterns on the Edlink LMS and designs an Early Warning System (EWS). Following the Knowledge Discovery in Databases framework, Edlink monitoring and evaluation reports for the odd and even semesters of 2025/2026 were extracted, cleaned, discretised, and labelled into three classes: Proactive, Standard, and Passive (At Risk). The dataset comprised 1,014 lecturer–study programme records (353 unique lecturers; 6,380 course classes). Categorical Naive Bayes with Laplace smoothing was evaluated using stratified 10-fold cross validation. Model 1 (within-semester classification) achieved a mean accuracy of 86.78%, with a recall of 95.11% and an F1-score of 95.47% for the Passive class. Model 2 (predicting the next semester, $n = 333$) achieved 73.27% accuracy and a Passive-class recall of 90.61%, comparable to a simple persistence rule (73.87%). Learning Assessment and Lesson Planning were the most influential variables. In the high-risk EWS group, 92.81% of lecturers were indeed Passive in the following semester. Edlink use increased significantly between semesters (Wilcoxon $p < 0.001$), although 63.98% of even-semester records remained Passive.*

Keyword: *digitalisation of education; digital learning patterns; Naive Bayes; educational data mining; early warning system.*

Abstrak: *Learning Management System (LMS) menghasilkan jejak data aktivitas pembelajaran dalam jumlah besar, tetapi di Universitas Lancang Kuning data tersebut selama ini hanya dianalisis secara deskriptif. Penelitian ini membangun model Naive Bayes untuk mengklasifikasikan dan memprediksi pola pembelajaran digital pada LMS Edlink serta merancang Early Warning System (EWS). Mengikuti kerangka Knowledge Discovery in Databases, laporan Monitoring dan Evaluasi Edlink Semester Ganjil dan Genap 2025/2026 diekstraksi, dibersihkan, didiskretisasi, dan dilabeli ke dalam tiga kelas: Proaktif, Standar, dan Pasif (Berisiko). Dataset terdiri atas 1.014 rekaman dosen–program studi (353 dosen unik; 6.380 kelas mata kuliah). Categorical Naive Bayes dengan Laplace smoothing dievaluasi menggunakan stratified 10-fold cross validation. Model 1 (klasifikasi dalam semester) mencapai akurasi rata-rata 86,78%, dengan recall 95,11% dan F1-score 95,47% pada kelas Pasif. Model 2 (prediksi semester berikutnya, $n = 333$) mencapai akurasi 73,27% dan recall kelas Pasif 90,61%, setara dengan aturan sederhana “kelas tetap” (73,87%). Learning Assessment dan Lesson Planning merupakan variabel paling berpengaruh. Pada kelompok risiko Tinggi EWS, 92,81% dosen benar-benar berada pada kelas Pasif di semester berikutnya. Penggunaan Edlink meningkat signifikan antarsemester (uji Wilcoxon $p < 0,001$), tetapi 63,98% rekaman Semester Genap masih berada pada kelas Pasif.*

Kata kunci: *digitalisasi pendidikan; pola pembelajaran digital; Naive Bayes; educational data mining; early warning system.*

PENDAHULUAN

Digitalisasi pendidikan telah menjadikan teknologi bukan sekadar alat bantu, melainkan ekosistem utama pembelajaran. *Learning Management System* (LMS) mencatat setiap aktivitas pembelajaran, mulai dari pembagian bahan ajar, interaksi forum, hingga pemberian tugas dan kuis, sehingga menghasilkan jejak data dalam jumlah besar (Han et al., 2022).

Namun, data aktivitas tersebut sering kali hanya menjadi tumpukan angka tanpa tindak lanjut. Di kelas digital, pola pembelajaran menjadi “tak kasat mata” karena tersimpan dalam log aktivitas, sehingga tanpa analisis yang memadai sulit mengetahui kelas atau dosen yang belum memanfaatkan LMS secara optimal. Bidang *Educational Data Mining* (EDM) dan *Learning Analytics* menawarkan pendekatan untuk mengubah data aktivitas tersebut menjadi pengetahuan yang dapat ditindaklanjuti (Baker & Yacef, 2009; Siemens & Baker, 2012; Romero & Ventura, 2020). Macfadyen dan Dawson (2010) menunjukkan bahwa data LMS dapat dimanfaatkan sebagai dasar sistem peringatan dini bagi pendidik, dan sistem semacam Course Signals telah diterapkan untuk meningkatkan keberhasilan belajar (Arnold & Pistilli, 2012).

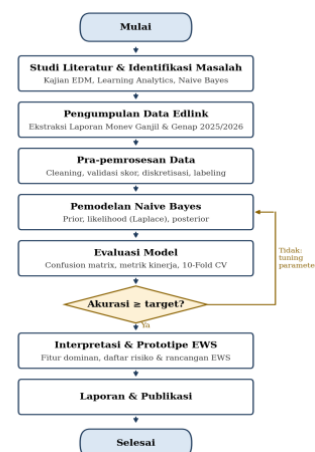
Salah satu algoritma klasifikasi yang banyak digunakan adalah Naive Bayes, yaitu pengklasifikasi probabilistik berbasis Teorema Bayes yang efisien secara komputasi dan menghasilkan probabilitas yang mudah ditafsirkan (Murphy, 2012; Witten et al., 2017). Naive Bayes telah diterapkan untuk mengklasifikasikan gaya belajar mahasiswa dengan akurasi 88,89% (Lisnawita et al., 2022), memetakan pola prestasi siswa (Fajri & Rusdi, 2025), mengklasifikasikan penggunaan akses internet siswa (Ramadhan & Suarna, 2022), dan memprediksi kelulusan mahasiswa (Zulfa & Utami, 2023). Pendekatan klasifikasi berbasis pembelajaran mesin juga telah digunakan

peneliti Universitas Lancang Kuning untuk analisis sentimen dengan K-NN dan SVM (Siswanto et al., 2022). Meskipun demikian, pemanfaatan data log LMS institusi untuk memprediksi pola pembelajaran digital dan menyusun daftar prioritas pendampingan masih jarang dilakukan.

Berdasarkan celah tersebut, penelitian ini bertujuan (1) mengidentifikasi variabel aktivitas pembelajaran digital yang paling berpengaruh, (2) membangun model klasifikasi pola pembelajaran digital menggunakan Naive Bayes, (3) mengukur kinerja model melalui *accuracy*, *precision*, *recall*, dan *F1-score* dengan *10-fold cross validation*, serta (4) merancang *Early Warning System* (EWS) berbasis hasil prediksi.

METODE

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan probabilitas dan statistik yang mengikuti kerangka *Knowledge Discovery in Databases* (KDD), yaitu seleksi data, pra-pemrosesan, transformasi, penambangan data, serta interpretasi dan evaluasi (Fayyad et al., 1996), sebagaimana ditunjukkan pada Gambar 1.



Gambar 1 Diagram Alir Penelitian

Data bersumber dari Laporan Monev Pembelajaran melalui Edlink yang disusun Pusat Pengembangan

Pembelajaran Universitas Lancang Kuning untuk Semester Ganjil dan Genap Tahun Akademik 2025/2026. Laporan memuat skor setiap dosen per program studi pada sepuluh fakultas/program, terdiri atas komponen *Lesson Planning* (bobot 30), *Learning Process* (bobot 40), dan *Learning Assessment* (bobot 30). Tabel skor diekstraksi dari berkas PDF menggunakan pdfplumber, nama dosen diganti dengan kode anonim, dan hasil ekstraksi divalidasi terhadap rata-rata resmi pada laporan. Rata-rata hasil ekstraksi Semester Ganjil sama persis dengan angka resmi, sedangkan pada Semester Genap selisihnya paling besar 0,05 poin.

Pra-pemrosesan meliputi pembuangan 21 baris tabel “Top Skor” yang mengulang data dan 11 baris ganda akibat tabel terpotong halaman, penyeragaman nama fakultas, pencocokan nama dosen antarsemester, serta penghitungan ulang skor total sebagai jumlah ketiga komponen karena 23 baris Semester Ganjil tidak konsisten. Atribut kemudian didiskretisasi dan kelas target ditentukan berdasarkan skor total sesuai Tabel 1. Dataset akhir terdiri atas 1.014 rekaman (345 Ganjil dan 669 Genap) dari 353 dosen unik.

Tabel 1 Variabel Penelitian dan Aturan Diskretisasi

Kode	Variabel	Kategori
X1	Lesson Planning	Nol / Rendah (< 0–30) 15) / Tinggi (≥ 15)
X2	Learning Process	Nol / Rendah (< 0–40) 10) / Tinggi (≥ 10)
X3	Learning Assessment	Nol / Rendah (< 0–30, 15) / Tinggi (≥ 15)
X4	Beban kelas	1–3 / 4–8 / ≥ 9 kelas
X5	Jenjang program studi	S1/D4; S2/S3
X6	Fakultas	10 fakultas/program
Y	Kelas pola pembelajaran digital	Proaktif (≥ 40) / Standar (20–< 40) / Pasif (< 20)

Kelas Y ditentukan dari skor total (0–100).

Model dibangun dengan menghitung probabilitas prior setiap kelas dan likelihood setiap kategori atribut dengan *Laplace smoothing*, lalu menetapkan kelas berdasarkan posterior terbesar, $\hat{C} = \text{argmax } P(C) \prod P(x_i | C)$ (Murphy, 2012). Implementasi menggunakan CategoricalNB pada pustaka scikit-learn (Pedregosa et al., 2011), dengan GaussianNB pada data numerik tanpa diskretisasi sebagai pembanding. Parameter *smoothing* (alpha) diuji pada nilai 0,1; 0,5; 1,0; dan 2,0.

Dua skenario diuji. Model 1 mengklasifikasikan kelas setiap rekaman berdasarkan atribut pada semester yang sama ($n = 1.014$). Model 2 menggunakan atribut Semester Ganjil untuk memprediksi kelas dosen pada Semester Genap ($n = 333$ dosen yang tercatat pada kedua semester), sehingga menguji kemampuan peringatan dini. Kinerja dievaluasi dengan *confusion matrix*, *accuracy*, *precision*, *recall*, dan *F1-score* melalui *stratified 10-fold cross validation* (Witten et al., 2017), dengan target akurasi minimal 80%. Kedua model dibandingkan dengan pembanding sederhana, yaitu menebak kelas mayoritas dan, untuk Model 2, aturan “kelas tetap seperti semester lalu”. Pengaruh variabel diukur dengan *mutual information* dan penurunan akurasi ketika variabel dihapus (*ablation*), sedangkan perubahan skor antarsemester diuji dengan uji Wilcoxon *signed-rank*.

HASIL DAN PEMBAHASAN

Deskripsi Data dan Distribusi Kelas

Sebaran skor sangat menceng ke kanan: median skor total hanya 8,50 pada Semester Ganjil dan 10,00 pada Semester Genap. Komponen *Learning Process* mengalami kenaikan rata-rata terbesar, dari 2,40 menjadi 7,68, sedangkan *Lesson Planning* (6,43 dan 5,88) dan *Learning Assessment* (6,95 dan 7,31) relatif tetap. Pada Semester Genap, 52,02% rekaman bernilai nol pada *Lesson Planning* dan 57,70% bernilai nol pada *Learning*

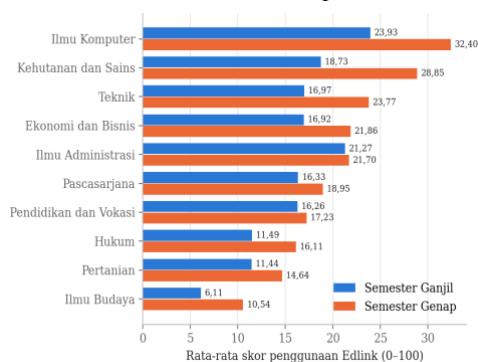
Assessment. Distribusi kelas disajikan pada Tabel 2.

Tabel 2 Distribusi Kelas Pola Pembelajaran Digital

Kelas	Ganjil	%	Genap	%	Total	%
Proaktif	40	11,59	122	18,24	162	15,98
Standar	79	22,90	119	17,79	198	19,53
Pasif (Berisik o)	226	65,51	428	63,98	654	64,50
Total	345	100	669	100	1.014	100

Sebagian besar rekaman berada pada kelas Pasif, sedangkan proporsi kelas Proaktif meningkat dari 11,59% menjadi 18,24%. Karena distribusi kelas tidak seimbang, evaluasi tidak hanya bertumpu pada akurasi, tetapi juga pada metrik per kelas, terutama kelas Pasif.

Rata-rata skor seluruh fakultas meningkat dari Semester Ganjil ke Genap (Gambar 2). Kenaikan terbesar terjadi pada Fakultas Kehutanan dan Sains (18,73 menjadi 28,85) dan Fakultas Ilmu Komputer (23,93 menjadi 32,40), sedangkan Fakultas Ilmu Budaya memiliki skor terendah (10,54). Uji Wilcoxon *signed-rank* terhadap 333 dosen menunjukkan kenaikan rata-rata dari 16,15 menjadi 20,79 (median 8,80 menjadi 11,92) yang signifikan secara statistik ($W = 13.219$; $p < 0,001$); 234 dosen mengalami kenaikan, 90 dosen penurunan, dan 9 dosen tetap.



Gambar 2 Rata-rata Skor Penggunaan Edlink per Fakultas

Kinerja Model 1: Klasifikasi Pola Pembelajaran Digital

Nilai α 0,1; 0,5; 1,0; dan 2,0 menghasilkan akurasi 86,69%; 86,69%; 86,79%; dan 86,69%, sehingga $\alpha = 1,0$ digunakan. Probabilitas prior kelas Proaktif, Standar, dan Pasif masing-masing 0,1598; 0,1953; dan 0,6450. Likelihood memperlihatkan pola yang jelas: peluang dosen kelas Pasif tidak memiliki aktivitas *Learning Assessment* sama sekali adalah 0,7900, sedangkan pada kelas Proaktif hanya 0,0424; sebaliknya, 90,30% dosen kelas Proaktif memiliki *Learning Assessment* kategori Tinggi. Hasil perhitungan manual pada contoh rekaman identik dengan keluaran program, sehingga implementasi dinyatakan benar.

Tabel 3 Kinerja Klasifikasi per Kelas (Model 1)

Kelas	Precision	Recall	F1-score	Support
Proaktif	78,81	73,46	76,04	162
Standar	64,95	70,20	67,48	198
Pasif (Berisik o)	95,84	95,11	95,47	654
Macro average	79,87	79,59	79,66	1.014
Weighted average	87,09	86,79	86,90	1.014

Nilai dalam persen; akurasi keseluruhan 86,79% (880 dari 1.014 rekaman).

Sebanyak 880 dari 1.014 rekaman terklasifikasi dengan benar (Gambar 3). Model tidak pernah menukar kelas Proaktif dan Pasif, sehingga kesalahan hanya terjadi antara kelas yang berdekatan. Kelas Pasif dikenali dengan sangat baik (*precision* 95,84%; *recall* 95,11%), sedangkan kelas Standar merupakan kelas tersulit (*F1-score* 67,48%) karena berada di antara dua kelas lain. Akurasi rata-rata 10-fold *cross validation* adalah 86,78% (simpangan baku 3,21), dan seluruh *fold* berada di atas target 80%. Nilai ini 22,29 poin lebih tinggi daripada pembandingan kelas mayoritas (64,50%). GaussianNB pada data numerik memperoleh akurasi lebih

tinggi (92,11%; *recall* Pasif 93,88%), yang menunjukkan bahwa diskretisasi menghilangkan sebagian informasi; namun, CategoricalNB dipertahankan karena probabilitasnya lebih mudah ditafsirkan oleh pimpinan.

Perlu dicatat bahwa kelas pada Model 1 diturunkan dari skor total, sedangkan skor total merupakan jumlah ketiga komponen yang juga menjadi atribut. Oleh karena itu, akurasi Model 1 sebagian besar mencerminkan kemampuan model merekonstruksi aturan penilaian dari data yang telah didiskretisasi. Kemampuan prediksi yang sesungguhnya diuji pada Model 2.

		Model 1					Model 2 (EWS)		
Aktual	Proaktif	119 73%	43 27%	0 0%	Aktual	Proaktif	32 61%	17 31%	4 7%
	Standar	32 16%	139 70%	27 14%		Standar	14 21%	18 27%	34 52%
	Pasif	0 0%	32 5%	622 95%		Pasif	8 4%	12 6%	193 91%
		Prediksi					Prediksi		

Gambar 3 Confusion Matrix Model 1 dan Model 2

Kinerja Model 2: Prediksi Semester Berikutnya

Model 2 mencapai akurasi 73,27% (rata-rata *fold* 73,28%; simpangan baku 3,80), di bawah target 80%. Kelas Pasif tetap dapat diprediksi dengan baik (*recall* 90,61%; *F1-score* 86,94%), tetapi kelas Standar sulit diprediksi (*F1-score* 31,86%) karena banyak dosen berpindah kelas antarsemester. Perbandingan dengan pembandingan sederhana disajikan pada Tabel 4.

Tabel 4. Perbandingan Model 2 dengan Pembandingan Sederhana

Model	akurasi	<i>recall</i> Pasif	<i>recall</i> Standar	<i>recall</i> Pasif
Naive Bayes (Model 2)	3,27	3,55	0,61	6,94
Aturan “kelas tetap”	3,87	7,50	8,73	8,11
Kelas mayoritas (Pasif)	3,96	-	100,00	-

Nilai dalam persen; $n = 333$ dosen (Ganjil $\rightarrow$ Genap).

Aturan sederhana “kelas semester depan sama dengan kelas semester ini” menghasilkan akurasi 73,87%, setara dengan Naive Bayes. Naive Bayes unggul tipis pada *recall* kelas Pasif (90,61% dibanding 88,73%), sedangkan aturan sederhana unggul pada *precision* (87,50% dibanding 83,55%). Matriks perpindahan kelas menunjukkan bahwa 189 dari 216 dosen Pasif pada Semester Ganjil tetap Pasif pada Semester Genap, sedangkan 26 dosen Standar naik menjadi Proaktif. Dengan data dua semester ini, nilai tambah Naive Bayes untuk peringatan dini terletak pada probabilitas risiko yang dihasilkannya, bukan pada akurasi yang lebih tinggi.

Variabel Paling Berpengaruh dan Rancangan EWS

Tabel 5 Peringkat Pengaruh Variabel

Variabel	Mutual informasi	Penurunan akurasi (poin)
X3 Learning Assessment	0,3605	9,28
X1 Lesson Planning	0,2937	7,89
X6 Fakultas	0,0609	0,30
X2 Learning Process	0,0461	-0,29
X4 Beban kelas	0,0098	0,30
X5 Jenjang	0,0029	-0,59

Penurunan akurasi Model 1 apabila variabel dihapus (ablation).

Learning Assessment dan *Lesson Planning* merupakan variabel paling berpengaruh (Tabel 5). Menghapus X3 menurunkan akurasi 9,28 poin dan menghapus X1 menurunkan 7,89 poin. *Learning Process* berpengaruh kecil, bahkan penghapusannya sedikit menaikkan akurasi, karena pada Semester Genap hampir semua dosen memiliki nilai komponen ini di atas nol sehingga kurang membedakan kelas. Fakultas, beban kelas, dan jenjang hampir tidak berpengaruh. Dengan demikian, keberhasilan pola pembelajaran digital paling ditentukan oleh dua kebiasaan dasar, yaitu

menyiapkan bahan ajar di setiap sesi dan memberikan tugas atau kuis melalui Edlink. Temuan ini sejalan dengan Romero dan Ventura (2020) yang menekankan pentingnya indikator perilaku pada LMS dalam analitik pembelajaran.

Tabel 6 Sebaran Tingkat Risiko Hasil Prediksi EWS

Risiko	P(Pasif)	Dose n	Rekomenda si tindakan
Tinggi	$\geq 80\%$	153	Pendampingan langsung (<i>coaching clinic</i>) sejak awal semester
Sedang	$50\% - < 80\%$	64	Pengingat berkala dan pelatihan fitur Edlink
Rendah	$< 50\%$	116	Pemantauan rutin

$n = 333$ dosen.

Probabilitas kelas Pasif dari Model 2 dikelompokkan menjadi tiga tingkat risiko (Tabel 6). Dari 153 dosen pada tingkat risiko Tinggi, 92,81% benar-benar berada pada kelas Pasif di Semester Genap, sehingga daftar ini layak menjadi prioritas pendampingan oleh Pusat Pengembangan Pembelajaran. Rancangan dasbor EWS mencakup ringkasan jumlah dosen per tingkat risiko untuk setiap fakultas dan program studi, daftar dosen berisiko beserta probabilitas dan komponen terendahnya, riwayat skor per semester, serta notifikasi kepada ketua program studi ketika tingkat risiko meningkat, sebagaimana prinsip sistem peringatan dini berbasis data LMS.

Penelitian ini memiliki beberapa keterbatasan. Data tersedia pada tingkat dosen per program studi, bukan log aktivitas setiap mahasiswa; hanya tersedia dua semester sehingga prediksi antarsemester baru diuji satu kali; skor Edlink merupakan ringkasan, bukan log mentah seperti durasi akses; dan terdapat ketidakkonsistenan skor pada sebagian

data Semester Ganjil. Selain itu, asumsi independensi Naive Bayes tidak sepenuhnya terpenuhi karena ketiga komponen saling berkorelasi, yang turut menjelaskan kesalahan pada kelas Standar.

SIMPULAN

Algoritma Naive Bayes dapat diterapkan melalui tahapan KDD untuk mengklasifikasikan pola pembelajaran digital pada LMS Edlink ke dalam kelas Proaktif, Standar, dan Pasif (Berisiko). Klasifikasi dalam semester mencapai akurasi rata-rata 86,78% dengan *recall* kelas Pasif 95,11%, melampaui target 80%, sedangkan prediksi semester berikutnya mencapai akurasi 73,27% dengan *recall* kelas Pasif 90,61%, setara dengan aturan sederhana “kelas tetap”. *Learning Assessment* dan *Lesson Planning* merupakan variabel paling berpengaruh. Probabilitas risiko dari model dapat digunakan sebagai dasar EWS; pada kelompok risiko Tinggi, 92,81% dosen benar-benar berada pada kelas Pasif di semester berikutnya. Penggunaan Edlink meningkat signifikan antarsemester, tetapi 63,98% rekaman Semester Genap masih berada pada kelas Pasif.

Pendampingan dosen sebaiknya difokuskan pada kebiasaan melengkapi bahan ajar setiap sesi dan memberikan tugas atau kuis melalui Edlink. Pusat Pengembangan Pembelajaran disarankan menyediakan ekspor data pada tingkat mahasiswa dan log mentah, serta menelusuri ketidakkonsistenan skor pada sistem monev. Penelitian selanjutnya disarankan menguji ulang model dengan data semester berikutnya, menambahkan seleksi fitur berbasis algoritma genetika (*Hybrid Naive Bayes-GA*), serta membandingkannya dengan algoritma lain seperti *Random Forest*.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih

kepada Universitas Lancang Kuning yang telah mendanai penelitian ini melalui APBU berdasarkan surat perjanjian pelaksanaan penelitian, kepada LPPM Universitas Lancang Kuning atas pendampingannya, serta kepada Pusat Pengembangan Pembelajaran Universitas Lancang Kuning yang telah menyediakan data Monev Edlink.

DAFTAR PUSTAKA

- Arnold, K. E., & Pistilli, M. D. (2012). Course Signals at Purdue: Using learning analytics to increase student success. Dalam *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (hlm. 267–270). ACM.
- Baker, R. S. J. d., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3–17.
- Fajri, M., & Rusdi, M. (2025). Implementasi metode Naive Bayes pada pola prestasi siswa dengan menggunakan RapidMiner. *Jurnal Mahasiswa Teknik Informatika (JAMASTIKA)*, 1(1), 320–328.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37–54.
- Han, J., Pei, J., & Tong, H. (2022). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann.
- Lisnawita, Guntoro, & Musfawati. (2022). Implementation of Naïve Bayes for classification of learning types. *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi*, 13(1),
- Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an “early warning system” for educators: A proof of concept. *Computers & Education*, 54(2), 588–599.
- Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press.
- Nijal, L., Zamzami, Z., Siswanto, D., & Febriadi, B. (2024). The role of multimedia in religious learning for students with disabilities in the inclusive system (case study: Lancang Kuning University). *IJISTECH (International Journal of Information System and Technology)*, 8(4), 211–217.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Ramadhan, A., & Suarna, I. W. (2022). Penerapan model klasifikasi metode Naive Bayes terhadap penggunaan akses internet siswa. *JURSISTEKNI*, 4(1), 1–8.
- Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3), e1355.
- Sahputri, D. N., Siswanto, D., Zamzami, Z., Nijal, L., Febriadi, B., & Agusviyanda. (2024). Creative design training in the Gen Z era: Teacher training at vocational schools using Canva for innovative learning media. *Dinamisia: Jurnal Pengabdian Kepada Masyarakat*, 8(5), 1515–1522.
- Siemens, G., & Baker, R. S. J. d. (2012). Learning analytics and educational data mining: Towards communication and collaboration. Dalam *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (hlm. 252–254). ACM.
- Siswanto, D., Zamzami, Nijal, L., & Rajab, S. (2022). Analisa sentimen publik mengenai perekonomian Indonesia pada masa pandemi Covid-19 di Twitter menggunakan metode

-
- klasifikasi K-NN dan SVM. *Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence)*, 2(1), 1–9.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). *Data mining: Practical machine learning tools and techniques* (4th ed.). Morgan Kaufmann.
- Zamzami, Siswanto, D., & Nijal, L. (2024). Pelatihan pemanfaatan Edpuzzel dalam peningkatan teknopreneurship pembelajaran secara interaktif bagi peserta didik di SMAN 7 Pekanbaru. *J-COSCIS: Journal of Computer Science Community Service*, 4(1), 72–78.
- Zamzami, Siswanto, D., & Sadar, M. (2021). Training of trainer e-learning bagi admin di Sekolah Tinggi Agama Islam Lukman Edy (STAILE) Pekanbaru. *J-COSCIS: Journal of Computer Science Community Service*, 1(2), 123–131.
- Zulfa, M., & Utami, R. (2023). Educational data mining for predicting student graduation using the Naïve Bayes classifier algorithm. *International Journal of Emerging Technologies in Learning*, 15(4)