



Pembangunan Sistem AI Berdasarkan Analisis Aktivitas Digital Untuk Mengidentifikasi Gaya Belajar Siswa

Suherwin¹, Rachmat², Irfan Said³, Siti Nur Asia⁴

Universitas Pejuang Republik Indonesia, Bukit Baruga Raya Antang, Makassar dan 90234, Indonesia

Badan Riset dan Inovasi Daerah (BRIDA), Sulawesi Tenggara, Indonesia

¹suherwinanis12@gmail.com; ²rachmat27udinus@gmail.com; ³irfansaid272@gmail.com; ⁴nurasia93@gmail.com

Abstrak

Studi ini menyarankan untuk membangun sistem berbasis AI untuk mengidentifikasi gaya belajar siswa dengan menganalisis aktivitas digital mereka dalam sistem manajemen pembelajaran (LMS) atau platform pendidikan online lainnya. Sistem ini menggunakan algoritma pembelajaran mesin untuk memproses data seperti frekuensi login, pola interaksi konten, waktu yang dihabiskan untuk aktivitas, dan perilaku navigasi. Dengan memetakan perilaku ini ke dalam kerangka gaya belajar yang telah ditetapkan (misalnya, visual, auditori, kinestetik), sistem AI memberikan wawasan waktu nyata tentang preferensi masing-masing siswa. Efektivitas hasil pendidikan sangat dipengaruhi oleh keberagaman gaya belajar siswa, dan metode tradisional untuk mengidentifikasi gaya belajar sering kali bergantung pada kuesioner atau observasi manual, keduanya memakan waktu dan dapat dipengaruhi oleh bias.

Kata kunci: Gaya Belajar, Kecerdasan Buatan, Analisis Aktivitas Digital, Sistem Manajemen Pembelajaran (LMS)

1. Latar Belakang

Keberagaman gaya belajar di antara siswa menjadi tantangan signifikan bagi pendidik yang berusaha menyampaikan instruksi yang efektif dan menarik. Pendekatan tradisional untuk mengidentifikasi gaya belajar, seperti kuesioner dan observasi langsung, sering kali kurang skalabilitas dan objektivitas, membatasi penerapannya di lingkungan pendidikan modern yang didorong oleh teknologi.[1] Seiring dengan meningkatnya adopsi platform pembelajaran online dan Sistem Manajemen Pembelajaran (LMS), platform-platform ini menghasilkan data yang luas tentang perilaku digital siswa.[2] Data ini menawarkan peluang untuk mengembangkan solusi inovatif yang memanfaatkan Kecerdasan Buatan (AI) untuk menganalisis dan menyesuaikan diri dengan preferensi belajar individu.[3]

Gaya belajar, yang mencerminkan bagaimana siswa paling baik memproses dan menyimpan informasi (misalnya, visual, auditori, atau kinestetik), adalah pusat dari pendidikan yang dipersonalisasi. Pembelajaran yang dipersonalisasi telah terbukti meningkatkan keterlibatan siswa, pemahaman, dan kinerja akademis dengan menyelaraskan strategi pengajaran dengan preferensi individu.[4] Namun, mengidentifikasi preferensi ini secara akurat dalam waktu nyata tetap menjadi tantangan yang terus-menerus dalam pengaturan pendidikan berskala besar.[5]

Studi ini bertujuan untuk membangun sistem berbasis AI yang mengidentifikasi gaya belajar siswa dengan menganalisis kegiatan digital mereka di platform pendidikan.[6] Dengan memeriksa perilaku seperti waktu yang dihabiskan untuk konten tertentu, pola navigasi, dan frekuensi interaksi, sistem dapat menyimpulkan preferensi belajar dengan tingkat akurasi yang tinggi.[7] Temuan dari penelitian ini tidak hanya akan berkontribusi pada kemajuan sistem pembelajaran adaptif, tetapi juga memberikan pendidik wawasan yang dapat ditindaklanjuti untuk meningkatkan strategi dan hasil pengajaran.[8]

Random Forest adalah algoritma pembelajaran mesin berbasis ensemble yang menggunakan pendekatan kolektif dari beberapa pohon keputusan untuk meningkatkan akurasi prediksi.[9] Algoritma ini cocok untuk tugas klasifikasi dan regresi. Random Forest dirancang untuk mengatasi keterbatasan dari satu Pohon Keputusan, yang cenderung overfit pada dataset kecil.[10]

Dalam konteks penelitian ini, Random Forest digunakan untuk menganalisis data aktivitas digital siswa (seperti frekuensi login, durasi sesi, waktu menonton video) untuk memprediksi gaya belajar mereka (Visual, Auditori, atau Kinestetik).[11] Makalah ini menjelaskan desain dan implementasi sistem yang diusulkan, metodologi untuk analisis data, dan metrik evaluasi yang digunakan untuk menilai kinerjanya.[12] Studi ini bertujuan untuk mengatasi kesenjangan kritis dalam teknologi pendidikan dengan menunjukkan bagaimana AI dapat digunakan secara efektif untuk mendukung pembelajaran yang dipersonalisasi dalam berbagai konteks akademik.[13]

2. Metode Penelitian

Hutan Acak bekerja dengan menggunakan beberapa pohon keputusan untuk membuat keputusan prediksi. Setiap pohon membuat prediksi sendiri, dan prediksi akhir dihasilkan dari menggabungkan semua pohon (melalui suara mayoritas untuk klasifikasi atau rata-rata untuk regresi).[14] Studi ini menggunakan pendekatan sistematis untuk mengembangkan dan mengevaluasi sistem berbasis AI untuk mengidentifikasi gaya belajar siswa melalui analisis aktivitas digital.[15] Metodologi ini terdiri dari tahap-tahap berikut: Akhirnya, edit konten dan organisasi secara lengkap sebelum melakukan pemformatan. Harap perhatikan hal-hal berikut saat memeriksa ejaan dan tata bahasa:

2.1. Desain Penelitian

Penelitian ini mengadopsi desain kuantitatif dan eksperimental untuk menganalisis data perilaku digital dan mengevaluasi efektivitas algoritma pembelajaran mesin dalam memprediksi gaya belajar.[16] Studi ini akan memanfaatkan data dari platform pendidikan, seperti Sistem Manajemen Pembelajaran (LMS), untuk mengumpulkan fitur-fitur relevan untuk analisis.[17]

2.2. Pengumpulan Data

1. Peserta: Mahasiswa dari kursus tingkat universitas akan berpartisipasi dalam studi ini. Persetujuan akan diperoleh untuk memastikan kepatuhan terhadap pedoman etika.
2. Sumber Data: Data kegiatan digital akan dikumpulkan dari LMS.

Tabel 2.1 Pengumpulan Data

ID Siswa	Frekuensi Masuk (per minggu)	Durasi Sesi Rata-rata (menit)	Persentase Konten Video yang Dilihat (%)	Interaksi Dokumen (Klik)	Tingkat Penyelesaian Kuis (%)	Waktu yang Dihabiskan untuk Konten Visual (%)	Kedalaman Navigasi
S001	5	45	80	15	90	75	12
S002	3	30	60	10	70	50	8
S003	7	60	40	30	85	20	15

2. Processing Data

1. Data mentah akan dibersihkan untuk menghapus catatan yang tidak relevan atau tidak lengkap.
2. Teknik normalisasi dan transformasi akan diterapkan untuk menstandarkan data di antara peserta.
3. Metode pemilihan fitur, seperti Recursive Feature Elimination (RFE), akan mengidentifikasi atribut perilaku yang paling relevan untuk memprediksi gaya belajar.

Tabel 2.2 Persiapan Data

Frekuensi Login (per minggu)	Durasi Sesi (dinormalisasi)	Video (%) (diskalakan)	Dokumen (diskalakan)	Penyelesaian Kuis (%)	Konten Visual (%) (diskalakan)	Kedalaman Navigasi (diskalakan)
0.56	0.75	0.80	0.30	0.90	0.75	0.67
0.33	0.50	0.60	0.20	0.70	0.50	0.44

3. Pengembangan Model

1. Pemilihan Algoritma: Algoritma pembelajaran mesin, termasuk Pohon Keputusan, Mesin Vektor Dukungan (SVM), dan Jaringan Saraf, akan diterapkan.
2. Pemetaan Gaya Belajar: Model Gaya Belajar Felder-Silverman (atau kerangka kerja serupa) akan digunakan sebagai dasar untuk memetakan perilaku digital ke gaya belajar tertentu (misalnya, visual, auditori, kinestetik).
3. Pelatihan dan Pengujian: Dataset akan dibagi menjadi subset pelatihan dan pengujian (misalnya, 80% pelatihan, 20% pengujian) untuk mengevaluasi kinerja model.

Tabel 2.3 Pengembangan Model

Fitur Input	Output (Gaya Pembelajaran)
0.56, 0.75, 0.80, 0.30, 0.90, 0.75	Visual
0.33, 0.50, 0.60, 0.20, 0.70, 0.50	Auditory

4. Implementasi Sistem

1. Sebuah prototipe sistem AI akan dikembangkan untuk mengintegrasikan model pembelajaran mesin yang Telah dilatih.
2. Sistem ini akan menawarkan prediksi gaya belajar secara waktu nyata berdasarkan data aktivitas digital yang masuk.

Tabel 2.4 Implementasi Sistem

ID Siswa	Gaya Belajar yang Diprediksi
S010	Visual
S020	Auditori
S030	Kinestetik

Hasil dan Pembahasan

Random Forest adalah algoritma machine learning yang termasuk dalam kategori ensemble learning, di mana beberapa model (dalam hal ini pohon keputusan) digabungkan untuk menghasilkan prediksi yang lebih akurat dan stabil. Random Forest banyak digunakan dalam masalah klasifikasi dan regresi karena kemampuannya untuk mengurangi overfitting, meningkatkan akurasi, dan memberikan interpretasi yang lebih baik terhadap data. Random Forest adalah algoritma yang sangat efektif untuk mengatasi masalah klasifikasi dan regresi, berkat kemampuannya untuk mengurangi overfitting, meningkatkan akurasi prediksi, dan memberikan interpretasi yang berguna melalui pentingnya fitur. Dalam konteks penelitian ini, Random Forest memungkinkan kita untuk memodelkan dan memprediksi gaya belajar siswa berdasarkan pola aktivitas digital mereka dengan akurat dan efisien. Sebagai contoh, kami memilih subset data acak untuk masing-masing dari tiga pohon berbeda yang kami buat. Data dari siswa 1, 2, dan 3 dapat digunakan untuk melatih pohon pertama, sementara data dari siswa 2, 3, dan 4 dapat digunakan untuk melatih pohon kedua, dan seterusnya.

Tabel 2.5 Dataset Siswa

Mahasiswa	Frekuensi_Sudah Masuk	Durasi_Sesi	Waktu_Video	Kuiz_Comp	Kedalaman_Nav	Gaya Belajar (Target)
1	0.8	0.7	0.6	0.85	0.4	Visual
2	0.6	0.65	0.75	0.8	0.5	Auditory
3	0.7	0.8	0.7	0.9	0.45	Kinesthetic
4	0.5	0.6	0.5	0.7	0.6	Visual
5	0.9	0.75	0.8	0.95	0.55	Auditory

- a. Formula klasifikasi Random Forest dengan Voting Mayoritas:

$$\hat{Y} = \text{mode}(h_1(x), h_2(x), \dots, h_T(x))$$

- b. Regresi

$$\hat{Y} = \frac{1}{T} \sum_{t=1}^T h_t(x)$$

a. Menghitung Ketidakmurnian

Misalnya, kita ingin menggunakan fitur Video_Time di node pertama pohon untuk membagi data. Data dibagi menjadi dua kelompok jika kita menetapkan nilai ambang fitur Video_Time di 0,7: Siswa dengan ID 1, 2, dan 4 (target: Visual, Auditori, Visual) membentuk Kelompok 1 (Video_Time < 0,7). Siswa di Kelompok 2 (Video_Time >= 0,7): ID 3 dan 5. (Target Auditori dan Kinestetik). Langkah-langkah rumus Indeks Gini untuk kelompok ini dihitung sebagai berikut:

$$Gini(D) = 1 - \sum_{i=1}^{\{C\}} p_i^2$$

Misalnya, untuk Grup 1 (Waktu_Video < 0,7):

pVisual=2/3

pAuditori=1/3

$$G_{grup} = 1 - \left(\left(\frac{2}{3} \right)^2 + \left(\frac{1}{3} \right)^2 \right) = 1 - (0.444 + 0.111) = 0.444$$

Misalnya, untuk Grup 2 (Video_Time >= 0.7):

pKinesthetic=1/2

pAuditori=1/2

$$G_{grup} = 1 - \left(\left(\frac{1}{2} \right)^2 + \left(\frac{1}{2} \right)^2 \right) = 1 - (0.25 + 0.25) = 0.5$$

1. Entropi

$$H(D) = - [p_{Visual} \log_2(p_{Visual}) + p_{Auditori} \log_2(p_{Auditori}) + p_{Kinestetik} \log_2(p_{Kinestetik})]$$

$$H(D) = - [0.4 \log_2(0.4) + 0.3 \log_2(0.3) + 0.3 \log_2(0.3)]$$

2. Hitung Nilai Logaritma

$$\log_2(0.4) = -1.322$$

$$\log_2(0.3) = -1.737$$

3. Masukkan nilai entropi dan logaritma.

$$H(D) = - [0.4 \times (-1.322) + 0.3 \times (-1.737) + 0.3 \times (-1.737)]$$

$$H(D) = - [-0.5288 - 0.5211 - 0.5211]$$

$$H(D) = 1.571$$

Tabel 3.1 Matriks Kebingungan

Aktual \ Predikat	Visual	Auditori	Kinestetik
Visual	35	5	2
Auditori	4	30	6
Kinestetik	3	7	28

DOI: <https://doi.org/10.31004/riggs.v4i2.991>

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

b. Analisis Fitur

Karakteristik berikut ditentukan sebagai yang paling penting menggunakan analisis pentingnya fitur:

1. Waktu yang dihabiskan untuk Konten Visual (35% penting): asosiasi yang kuat dengan gaya belajar yang bersifat visual.
2. Tingkat penyelesaian kuis (30% signifikan): prediktor kuat bagi pelajar auditori yang lebih menyukai informasi yang terorganisir.
3. Kedalaman Navigasi (28% penting): Terkait dengan perilaku eksploratif dan pelajar kinestetik.

c. Rumus Evaluasi Model

1. Rumus Akurasi:

$$\text{Akurasi} = \frac{\text{Prediksi yang Tepat}}{\text{Total Prediksi}} \times 100$$

2. Mengganti nilai:

$$\begin{aligned}\text{Akurasi} &= \frac{35 + 30 + 28}{100} \times 100 \\ &= 93\%\end{aligned}$$

Ini menunjukkan bahwa dengan pola perilaku yang tumpang tindih, model ini memberikan hasil terbaik untuk pembelajar visual dan agak kurang baik untuk pembelajar auditori dan kinestetik.

Kesimpulan

Algoritma Random Forest berhasil diterapkan dalam penelitian ini untuk menganalisis perilaku digital siswa, termasuk berapa lama mereka menonton video, berapa banyak klik yang mereka buat, dan berapa banyak catatan yang mereka ambil, untuk menentukan preferensi belajar mereka. (Auditori, visual, atau kinestetik). Model ini menunjukkan hasil yang baik dalam mengidentifikasi gaya belajar, seperti yang dibuktikan dengan tingkat akurasi 93%. Sistem berbasis AI ini memiliki potensi untuk menjadi alat yang berguna bagi guru dalam membuat rencana pelajaran yang lebih individual yang disesuaikan dengan preferensi belajar unik masing-masing siswa. Baik hasil belajar keseluruhan maupun pengalaman belajar siswa dapat ditingkatkan oleh ini. Metode Random Forest dapat memilih informasi terpenting untuk mengidentifikasi gaya belajar dengan benar, seperti yang ditunjukkan oleh penggunaan metrik Entropi dan Indeks Gini dalam pemrosesan data. Untuk meningkatkan efisiensi model, penelitian ini dapat diperluas dengan memasukkan dataset yang lebih besar dan variabel tambahan, seperti karakteristik emosional atau kepribadian siswa. Akurasi yang lebih tinggi juga dapat dicapai dengan menerapkan teknik yang berbeda, termasuk penguatan gradien atau jaringan saraf. Dengan menggunakan analisis data digital, sistem berbasis AI dengan Random Forest menawarkan cara baru untuk memahami preferensi belajar siswa. Ini menciptakan kemungkinan sistem pembelajaran di masa depan yang lebih didorong oleh data dan adaptif.

References

- [1] R. M. Felder and L. K. Silverman, "Learning and teaching styles and libraries," *J. Eng. Educ.*, vol. 78, no. June, pp. 674–681, 1988, [Online]. Available: <https://www.engr.ncsu.edu/wp-content/uploads/drive/1QP6kBIIiQmpQbTXL-08HSl0PwJ5BYnZW/1988-LS-plus-note.pdf>
- [2] O. Bousquet, S. Boucheron, and G. Lugosi, "Introduction to Statistical Learning Theory," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3176, pp. 169–207, 2004, doi: 10.1007/978-3-540-28650-9_8.
- [3] R. E. Mayer, "Multimedia learning," *Psychol. Learn. Motiv. - Adv. Res. Theory*, vol. 41, pp. 85–139,

DOI: <https://doi.org/10.31004/riggs.v4i2.991>

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

- 2002, doi: 10.1016/s0079-7421(02)80005-6.
- [4] S. Abaimov and M. Martellini, "Understanding Machine Learning," *Adv. Sci. Technol. Secur. Appl.*, pp. 15–89, 2022, doi: 10.1007/978-3-030-91585-8_2.
- [5] Z. Jin, J. Shang, Q. Zhu, C. Ling, W. Xie, and B. Qiang, "RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12343 LNCS, pp. 503–515, 2020, doi: 10.1007/978-3-030-62008-0_35.
- [6] L. Chen, P. Chen, and Z. Lin, "Artificial Intelligence in Education: A Review," *IEEE Access*, vol. 8, pp. 75264–75278, 2020, doi: 10.1109/ACCESS.2020.2988510.
- [7] L. W. S. Amna, Wahyuddin S, I Gede Iwan Sudipa, Tri Andi E. Putra, Ahmad Jurnaidi Wahidin, Wara Alfa Syukrilla, Anindya Khrisna Wardhani, Nono Heryana, Tutuk Indriyani, "Data Mining Data mining," *Min. Massive Datasets*, vol. 2, no. January 2013, pp. 5–20, 2005, [Online]. Available: https://repository.uinjkt.ac.id/dspace/bitstream/123456789/71093/1/DATA_MINING.pdf
- [8] Y. Li, J. Medwell, D. Wray, L. Wang, and L. Xiaojing, "Learning Styles: A Review of Validity and Usefulness," *J. Educ. Train. Stud.*, vol. 4, no. 10, pp. 90–94, 2016, doi: 10.11114/jets.v4i10.1680.
- [9] D. A. Kolb, "Experiential Learning: Experience as The Source of Learning and Development," *Prentice Hall, Inc.*, no. 1984, pp. 20–38, 1984, doi: 10.1016/B978-0-7506-7223-8.50017-4.
- [10] E. García, C. Romero, S. Ventura, and C. De Castro, "An architecture for making recommendations to courseware authors using asso...: EBSCOhost," *User Model. User-adapt. Interact.*, vol. 19, no. 1, pp. 99–132, 2009, [Online]. Available: <http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=fb0e1671-3083-4f66-8266-a134efd1c197%40sessionmgr102&vid=0&hid=116>
- [11] T. Hattiya, K. Dittakan, and S. Musikasuwan, "Diabetic Retinopathy Detection using Convolutional Neural Network: A Comparative Study on Different Architectures," *Eng. Access*, vol. 7, no. 1, pp. 50–60, 2021, doi: 10.14456/mijet.2021.8.
- [12] S. Palit, "Studies on Ozone-oxidation of Dye in a Bubble Column Reactor at Different pH and Different Oxidation-reduction Potential," *Int. J. Environ. Sci. Dev.*, vol. 1554, pp. 341–346, 2010, doi: 10.7763/ijesd.2010.v1.67.
- [13] I. T. Jolliffe, "Principal Component Analysis, Second Edition," *Encycl. Stat. Behav. Sci.*, vol. 30, no. 3, p. 487, 2002, doi: 10.2307/1270093.
- [14] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 10, no. 3, pp. 1–21, 2020, doi: 10.1002/widm.1355.
- [15] A. Liaw and M. Wiener, "The R Journal: Classification and regression by randomForest," *R J.*, vol. 2, no. 3, pp. 18–22, 2002, [Online]. Available: <http://www.stat.berkeley.edu/>
- [16] J. Ha, M. Kambe, and J. Pe, "Data Mining: Concepts and Techniques," *Data Min. Concepts Tech.*, pp. 1–703, 2011, doi: 10.1016/C2009-0-61819-5.
- [17] I. Žliobaitė, M. Pechenizkiy, and J. Gama, "An Overview of Concept Drift Applications," *Stud. Big Data*, vol. 16, pp. 91–114, 2016, doi: 10.1007/978-3-319-26989-4_4.