

Terbit online pada laman : <http://teknosi.fti.unand.ac.id/>

Jurnal Nasional Teknologi dan Sistem Informasi

| ISSN (Print) 2460-3465 | ISSN (Online) 2476-8812 |



Studi Kasus

Analisis Pola Belajar Mahasiswa Pada Platform Pembelajaran Daring (Studi Kasus: LeADS UPNVJ)

Nindy Irzavika ^{a,*}, Kharisma Wiati Gusti ^b, Mohamad Thoriq Abdurachman ^c, Bima Saputra ^d

^a Program Studi Sains Data, Universitas Pembangunan Nasional Veteran Jakarta, Jakarta, Indonesia

^{b,d} Program Studi Informatika, Universitas Pembangunan Nasional Veteran Jakarta, Jakarta, Indonesia

^c Program Studi Sistem Informasi, Universitas Pembangunan Nasional Veteran Jakarta, Jakarta, Indonesia

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 13 Agustus 2025

Revisi Akhir: 20 November 2025

Diterbitkan Online: 28 Desember 2025

KATA KUNCI

Big data,
learning analytics,
pola belajar,
random forest,
pembelajaran daring

KORESPONDENSI

E-mail: nindiyirzavika@upnvj.ac.id*

ABSTRACT

Kebutuhan pembelajaran jarak jauh mendorong pemanfaatan platform pembelajaran daring sebagai media utama yang mendukung proses belajar mengajar di perguruan tinggi. Salah satu tantangan dalam implementasi sistem ini adalah kemampuan untuk memahami pola belajar mahasiswa secara objektif dan berbasis data. Penelitian ini bertujuan untuk mengidentifikasi dan menganalisis pola belajar mahasiswa pada *Learning Management System* (LMS) LeADS di Universitas Pembangunan Nasional Veteran Jakarta melalui pendekatan analisis big data. Data penelitian berupa log aktivitas mahasiswa pada LMS yang mencakup interaksi akademik seperti melihat materi, mengumpulkan tugas, mengerjakan kuis, dan partisipasi diskusi selama empat semester di program studi Sistem Informasi. Metode penelitian menggunakan pendekatan kuantitatif sesuai dengan siklus hidup big data. Algoritma *Random Forest* digunakan untuk mengklasifikasikan waktu belajar mahasiswa dengan optimasi parameter menggunakan metode *Grid Search*. Hasil penelitian menunjukkan bahwa partisipasi aktif relatif rendah yang mengindikasikan pola belajar cenderung pasif. Konsistensi dan keterlibatan mahasiswa dalam LMS memiliki pengaruh yang lebih signifikan pada prestasi akademik dibandingkan waktu belajar dominan mahasiswa. Hasil *Random Forest* yang dibangun memiliki akurasi 88%, namun performanya belum optimal dalam mengklasifikasi kelas dengan jumlah terbatas. Hasil penelitian ini menunjukkan pentingnya pemanfaatan data log LMS untuk memahami perilaku belajar mahasiswa secara lebih komprehensif serta membuka peluang pengembangan strategi pembelajaran adaptif berbasis data. Selain itu, hasil penelitian ini memberikan dasar bagi institusi pendidikan untuk merancang intervensi pembelajaran yang lebih efektif dan personal sesuai kebutuhan mahasiswa.

1. PENDAHULUAN

Platform pembelajaran daring telah menjadi bagian integral dari pendidikan modern, terutama dengan meningkatnya kebutuhan akan pembelajaran jarak jauh. Pendidikan daring telah berkembang pesat serta menjadi pilar penting dalam dunia pendidikan, terutama setelah pandemi COVID-19. Dengan meningkatnya adopsi platform pembelajaran daring, muncul tantangan baru bagi pendidik, yaitu memahami pola belajar siswa dan memastikan efektivitas pembelajaran daring. Tantangan utamanya adalah pemantauan efektivitas platform pembelajaran

daring, keterlibatan mahasiswa dalam proses belajar mengajar, serta memberikan dukungan individual kepada mahasiswa.

Salah satu sistem yang dapat digunakan adalah *Learning Management System* (LMS) yang menyediakan sarana bagi perguruan tinggi untuk melaksanakan perkuliahan secara daring dan bauran. *Learning Management System* (LMS) LeADS UPNVJ merupakan sistem pengelolaan belajar daring di lingkungan Universitas Pembangunan Nasional Veteran Jakarta [1]. Platform ini dikembangkan berbasis Moodle, sebuah sistem manajemen pembelajaran *open source* yang populer digunakan diberbagai institusi pendidikan [2]. Dengan platform ini

diharapkan proses belajar dapat berlangsung efektif dan mendukung capaian pembelajaran setiap mahasiswa secara optimal.

Pembelajaran daring yang diimplementasikan melalui Learning Management System (LMS) menghasilkan jejak digital (log) yang dapat dianalisis untuk memahami perilaku belajar mahasiswa. Kajian terkait learning analytics menunjukkan bahwa interaksi mahasiswa pada LMS dapat digunakan untuk memprediksi performa akademik serta mengidentifikasi mahasiswa yang berpotensi mengalami kesulitan belajar. Setiap interaksi dan aktivitas mahasiswa pada platform LeADS menghasilkan log yang dapat dimanfaatkan untuk menganalisis pola belajar mahasiswa serta korelasinya dengan Indeks Prestasi Kumulatif (IPK) yang diperoleh oleh mahasiswa. Pola akses mahasiswa terhadap atribut pembelajaran digital secara signifikan mempengaruhi nilai akhir mereka [3]. Analisis log aktivitas mahasiswa di secara konsisten menunjukkan korelasi kuat antara keterlibatan digital, seperti frekuensi akses, partisipasi forum, pengerjaan tugas, dan prestasi akademik [4], [5]. Pola akses pada malam hari dan akhir pekan juga berhubungan positif dengan performa [4], serta model prediksi berbasis log LMS dapat digunakan untuk deteksi dini mahasiswa berisiko gagal atau dropout [6], [7].

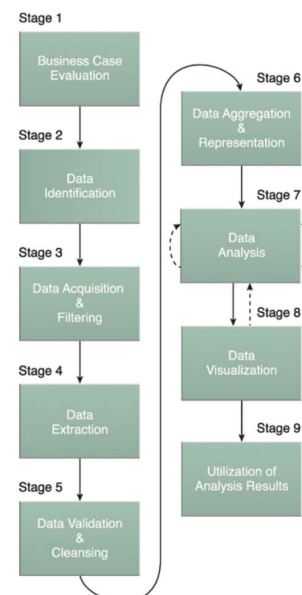
Keterlibatan siswa dengan video daring dapat lebih dipahami menggunakan data log dari sistem manajemen pembelajaran seperti Moodle dan Vimeo, tetapi waktu dan keahlian yang dibutuhkan untuk penerapannya secara luas dapat membatasi penerapannya secara luas [8]. Aspek perilaku siswa dalam hal akses sumber daya, tugas penilaian, dan tingkat keterlibatan dapat diungkapkan melalui penggalan data log tindakan dalam lingkungan belajar daring [9]. Berbagai aspek perilaku mahasiswa dalam LMS – seperti frekuensi akses sumber belajar, pengerjaan tugas penilaian, dan tingkat partisipasi dalam forum diskusi – dapat diungkap melalui penambahan data log aktivitas di lingkungan belajar daring [10]. Pendekatan ini dapat digunakan untuk mendapatkan wawasan mendalam mengenai pola belajar mahasiswa dengan memanfaatkan big data [11].

Random Forest menjadi salah satu algoritma yang paling populer dan efektif untuk menganalisis big data [12], serta memiliki akurasi yang tinggi, penggunaan memori yang efisien, serta ketahanan pada data tidak seimbang [13] [14]. *Random Forest* merupakan algoritma yang banyak digunakan untuk klasifikasi berdasarkan perilaku pengguna. Model ini unggul dalam akurasi prediksi [5], [15], identifikasi fitur perilaku yang paling berpengaruh [16], serta ketahanan terhadap data yang tidak seimbang [17], [18]. Pada penelitian ini *Random Forest* digunakan untuk mengklasifikasikan waktu belajar mahasiswa dengan optimasi parameter menggunakan metode *Grid Search*. Algoritma ini mampu menentukan fitur perilaku yang paling berpengaruh [19], mengidentifikasi siswa yang berisiko mengalami penurunan prestasi [20], serta mengelompokkan mahasiswa berdasarkan pola belajar mereka, sehingga strategi pembelajaran dapat dipersonalisasi sesuai kebutuhan kelompok tertentu [21]. Hal ini dapat dimanfaatkan sebagai pendukung pengambilan keputusan berdasarkan data untuk meningkatkan kualitas pembelajaran daring.

Berbeda dengan penelitian sebelumnya yang hanya fokus pada evaluasi keterlibatan mahasiswa secara deskriptif [10] [11], penelitian ini mengintegrasikan siklus hidup big data secara menyeluruh dengan menerapkan algoritma *Random Forest* untuk menganalisis pola belajar mahasiswa pada LMS LeADS UPNVJ. Pendekatan ini tidak hanya memetakan pola belajar mahasiswa berdasarkan waktu, frekuensi, dan durasi aktivitas, tetapi juga mengkaji hubungan fitur perilaku tersebut dengan prestasi akademik mahasiswa secara kuantitatif. Analisis big data dapat memberikan wawasan penting bagi pendidik dalam merancang pengalaman pembelajaran yang lebih efektif. Dengan mengintegrasikan analisis big data dan model klasifikasi, pola belajar siswa dapat diidentifikasi untuk meningkatkan efektivitas pendidikan daring. Oleh karena itu, penelitian ini memberikan kontribusi dengan menggabungkan tahapan identifikasi, akuisisi, pemrosesan, hingga analisis dan visualisasi menggunakan algoritma klasifikasi yang telah dioptimasi.

2. METODE

Analisis big data berbeda dari analisis data tradisional, terutama karena karakteristik *volume*, *velocity*, dan *variety* data yang diproses [22]. Sehingga diperlukan langkah-langkah untuk mengatur aktivitas dan tugas yang terlibat dalam perolehan, pemrosesan, dan analisis data. Siklus hidup big data digunakan sebagai landasan dalam metode penelitian yang dilakukan. Siklus ini mencakup tahapan yang sistematis untuk memastikan data diolah, dianalisis, dan dimanfaatkan secara efektif untuk menghasilkan wawasan yang valid [23][24]. Siklus hidup analisis big data terdiri dari sembilan (9) tahap, yaitu (1) Business Case Evaluation, (2) Data Identification, (3) Data Acquisition and Filtering, (4) Data Extraction, (5) Data validation and Cleansing, (6) Data Aggregation and Representation, (7) Data Analysis, (8) Data Visualization, (9) Utilization of Aseperti pada Gambar 1.



Gambar 1. Siklus Hidup Analisis Big Data [22]

Gambar 1 menggambarkan sembilan tahapan terstruktur dalam siklus hidup analisis big data, yang dimulai dari Business Case Evaluation sebagai proses penetapan tujuan dan ruang lingkup

analisis, diikuti oleh Data Identification serta Data Acquisition & Filtering untuk menentukan sumber data yang relevan dan memperoleh data yang berkualitas. Tahap berikutnya mencakup Data Extraction serta Data Validation & Cleansing guna memastikan data berada dalam format yang konsisten, akurat, dan siap diproses lebih lanjut. Data yang telah tervalidasi kemudian dikonsolidasikan pada tahap Data Aggregation & Representation, sebelum dianalisis secara mendalam melalui Data Analysis. Hasil analisis disajikan dalam bentuk visual pada tahap Data Visualization, yang selanjutnya menjadi dasar bagi pengambilan keputusan strategis pada tahap akhir yaitu Utilization of Analysis Results. Diagram ini juga menekankan sifat iteratif dari proses tersebut, yang mengindikasikan bahwa beberapa tahap dapat dilakukan berulang untuk meningkatkan kualitas dan ketepatan hasil analisis.

3. HASIL DAN PEMBAHASAN

3.1. Business Case Evaluation

Evaluasi kasus dilakukan untuk mengidentifikasi tujuan serta permasalahan big data yang akan diselesaikan. Fokus utama dalam penelitian ini adalah memahami pola belajar mahasiswa melalui analisis data log dari LMS LeADS UPNVJ. Informasi yang dihasilkan diharapkan dapat digunakan oleh dosen dan pengelola LMS dalam proses pengambilan keputusan strategis. Tujuannya adalah untuk merancang strategi pembelajaran yang lebih adaptif, menghasilkan informasi yang dapat digunakan oleh pengelola pendidikan tinggi dalam pengambilan keputusan terkait, dan peningkatan kualitas pembelajaran berbasis data.

Selain itu, evaluasi kasus dilakukan untuk memastikan bahwa kebutuhan analisis selaras dengan tantangan pedagogis dan operasional yang dihadapi dalam pembelajaran daring. Pada tahap ini, diidentifikasi pula faktor-faktor eksternal yang dapat memengaruhi perilaku belajar mahasiswa, seperti beban kerja akademik, tingkat literasi digital, dan pola penggunaan LMS yang tidak merata antar mahasiswa. Dengan memahami konteks tersebut, analisis big data yang dilakukan tidak hanya menghasilkan temuan deskriptif, tetapi juga memberikan wawasan yang relevan bagi pengembangan kebijakan pembelajaran digital yang lebih efektif dan berkelanjutan.

3.2. Data Identification

Tahap ini mengidentifikasi data yang relevan dari platform pembelajaran daring. Identifikasi data yang akurat dapat mengurangi penggunaan memori, waktu komputasi, dan kebutuhan perangkat yang digunakan [25]. Identifikasi data dilakukan secara sistematis untuk menjamin bahwa fitur yang diambil merepresentasikan aktivitas akademik yang aktual. Data yang digunakan merupakan data log aktivitas mahasiswa dari LMS LeADS UPNVJ yang mencatat setiap interaksi mahasiswa dalam bentuk *timestamp (time)*, *user full name*, *event context*, *component*, *event name*, *description*, dan *IP address*. Data log tersebut meliputi waktu akses mahasiswa, frekuensi partisipasi dalam diskusi daring, dan tingkat keterlibatan mereka dalam aktivitas pembelajaran lainnya. Selain itu, aktivitas yang berkaitan langsung dengan pembelajaran seperti membuka materi, mengerjakan kuis, mengunggah tugas, dan berkomunikasi termasuk data yang dibutuhkan dalam penelitian ini.

Proses identifikasi data ini juga mempertimbangkan kualitas, kelengkapan, serta konsistensi data untuk memastikan bahwa seluruh atribut yang digunakan mampu mencerminkan perilaku belajar secara komprehensif. Setiap jenis interaksi dianalisis berdasarkan relevansinya dengan tujuan penelitian, sehingga hanya fitur yang memberikan kontribusi strategis terhadap pemahaman pola belajar yang dipertahankan. Selain itu, identifikasi data dilakukan dengan memperhatikan potensi redundansi, bias pencatatan, serta variasi aktivitas lintas mata kuliah yang dapat memengaruhi distribusi data. Dengan pendekatan ini, dataset yang dihasilkan tidak hanya representatif terhadap aktivitas mahasiswa dalam LMS, tetapi juga siap untuk diproses lebih lanjut dalam tahapan analisis big data secara efisien dan akurat.

3.3. Data Acquisition and Filtering

Pengumpulan data dilakukan dari sistem LMS LeADS UPNVJ. Data yang dikumpulkan terdiri dari 4.265.825 baris data data log mahasiswa program studi S1 Sistem Informasi selama empat (4) semester, yaitu Ganjil 2023, Genap 2023, Ganjil 2024, dan Genap 2024. Data diperoleh dalam format *.csv yang kemudian disaring untuk memastikan validitas dan relevansi. Selain itu juga digunakan dataset IPK mahasiswa dari sistem akademik UPNVJ. Data ini mencakup interaksi mahasiswa terhadap berbagai komponen LMS seperti materi kuliah, kuis, forum diskusi, dan tugas.

Filtering adalah proses menyaring, membersihkan, dan menghilangkan data yang tidak relevan, duplikat, atau *noise* sebelum data dianalisis lebih lanjut [26] [27]. Proses penyaringan yang dilakukan diantaranya adalah menghapus data NIM yang kosong, aktivitas dari akun dosen, dan administrator LMS, serta baris data yang tidak lengkap. Tujuan dari tahapan ini adalah menjamin bahwa data mahasiswa yang valid dan representatif yang akan digunakan dalam tahap berikutnya.

Skala data yang sangat besar menuntut proses akuisisi dilakukan secara terstruktur untuk memastikan setiap file log dapat diintegrasikan tanpa kehilangan informasi penting. Pada tahap ini, dilakukan pula pengecekan format dan konsistensi struktur file untuk menghindari kesalahan pembacaan selama proses pemrosesan data. Karena data berasal dari berbagai semester dan mata kuliah, langkah penyesuaian metadata dilakukan untuk menyamakan definisi atribut, mengatasi perubahan struktur log antar periode, serta memastikan bahwa seluruh variabel dapat dianalisis. Pendekatan ini memungkinkan data yang diakuisisi tidak hanya mencerminkan aktivitas mahasiswa secara individual, tetapi juga mendukung analisis tren perilaku belajar dari waktu ke waktu.

3.4. Data Extraction

Ekstraksi data dilakukan untuk memisahkan atribut kompleks menjadi fitur yang lebih terstruktur dan terukur. Pada big data ekstraksi data menjadi penting untuk dapat mengurangi dimensi data dan meningkatkan efisiensi analisis [28]. Tahap ini menentukan kualitas dan kecepatan analisis big data serta memastikan data siap digunakan untuk analisis.

Proses ekstraksi tidak hanya berfokus pada pemisahan atribut, tetapi juga pada mengubah data menjadi variabel baru yang lebih informatif dan relevan terhadap tujuan analisis. Misalnya, atribut

timestamp diolah menjadi fitur turunan seperti kategori waktu belajar, hari aktivitas, serta durasi antar aktivitas untuk menangkap pola perilaku mahasiswa secara lebih mendalam. Atribut deskriptif pada *event_name* diklasifikasikan ke dalam kelompok aktivitas seperti *view*, *submit*, *attempt*, dan *communicate* agar dapat dianalisis secara kuantitatif. Pendekatan ini memastikan bahwa data log yang semula bersifat granular dan tidak terstruktur dapat diubah menjadi dataset analitis yang mampu merepresentasikan perilaku belajar secara komprehensif.

Pada data log yang diperoleh terdapat beberapa atribut yang perlu dilakukan proses ekstraksi data. Atribut *time* berisi informasi tanggal dan waktu akses, fitur ini diubah menjadi atribut tanggal, jam, dan nama hari, serta dikategorikan menjadi waktu belajar berdasarkan waktu dominan aktivitas. Atribut *user_full_name* berisi informasi NIM dan nama lengkap, fitur ini dipisahkan menjadi NIM, nama depan, dan nama belakang untuk keperluan identifikasi unik per mahasiswa. Hasil ekstraksi data dapat dilihat pada Gambar 2. Waktu belajar dibagi menjadi empat kategori, yaitu pagi, siang, malam dan dini hari. Waktu belajar dikelompokkan berdasarkan waktu akses ke platform pembelajaran daring sebagai berikut:

- Pagi: waktu akses pukul 06:00 – 11:59
- Siang: waktu akses pukul 12:00 – 17:59
- Malam: waktu akses pukul 18:00 – 23:59
- Dini hari: waktu akses pukul 00:00 – 05.59

date	day_name	hour	waktu_belajar	nim	first_name
2025-07-28	Monday	11	Pagi	2310512092	ACHMAD
2025-07-28	Monday	11	Pagi	2310512092	ACHMAD
2025-06-26	Thursday	13	Siang	2310512117	JIHAN
2025-06-16	Monday	12	Siang	2210512166	DAVY
2025-06-16	Monday	12	Siang	2210512166	DAVY
2025-06-04	Wednesday	20	Malam	2310512109	MUHAMMAD
2025-05-20	Tuesday	9	Pagi	2310512040	ZELDA
2025-05-20	Tuesday	9	Pagi	2310512040	ZELDA
2025-05-20	Tuesday	9	Pagi	2310512040	ZELDA
2025-05-20	Tuesday	9	Pagi	2310512040	ZELDA

Gambar 2. Hasil Ekstraksi data

3.5. Data Validation and Cleansing

Validasi dan pembersihan data dilakukan untuk memastikan kualitas dan konsistensi data. Data yang telah diekstrak diperiksa dan dibersihkan untuk menghilangkan kesalahan, nilai kosong, atau inkonsistensi. Tahap ini memastikan bahwa data yang digunakan untuk analisis akurat dan dapat diandalkan. Validasi dan pembersihan data mencegah data kotor masuk ke tahap analisis, sehingga mengurangi risiko kesalahan dalam pengambilan keputusan [29] [30]. Validasi dilakukan untuk memastikan bahwa setiap entitas mahasiswa teridentifikasi secara unik melalui NIM serta semua nilai waktu valid tidak ada nilai NaT (*Not a Time*). Selain itu, atribut *event_name* diklasifikasikan ke dalam kategori aktivitas yang lebih sederhana, yaitu *View*, *Update*, *Submit*, *Attempt*, *Communicate*, *Download*, *Delete*, dan *Other* seperti pada Gambar 3. Proses ini bertujuan untuk memudahkan klasifikasi perilaku belajar mahasiswa.

activity_type	
View	3110645
Update	346806
Submit	204356
Other	148658
Attempt	22357
Delete	7440
Communicate	4869
Download	354

Gambar 3. Jumlah untuk setiap jenis aktivitas

Selain memastikan keakuratan data, tahap validasi juga mencakup pemeriksaan konsistensi lintas semester dan mata kuliah, mengingat data log diperoleh dari periode yang berbeda dan berpotensi memiliki variasi dalam format pencatatan. Proses ini melibatkan pengecekan anomali seperti lonjakan aktivitas yang tidak wajar, duplikasi log akibat refresh halaman, dan aktivitas sistem yang tidak mencerminkan perilaku mahasiswa. Teknik deteksi anomali sederhana, seperti pemeriksaan distribusi frekuensi dan identifikasi outlier berbasis aturan, digunakan untuk memastikan bahwa hanya data yang benar-benar merepresentasikan interaksi mahasiswa yang dipertahankan. Dengan demikian, tahap ini tidak hanya berfungsi sebagai pembersihan dasar, tetapi juga sebagai mekanisme kontrol kualitas agar dataset akhir memiliki reliabilitas yang tinggi untuk dianalisis lebih lanjut.

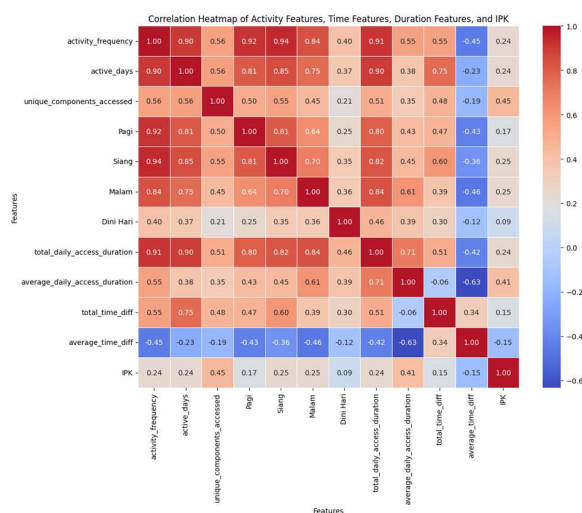
Setelah proses pembersihan dan validasi data, dataset disiapkan untuk proses pelatihan model klasifikasi. Data dibagi ke dalam dua subset menggunakan metode train-test split dengan rasio 80:20. Pembagian ini bertujuan untuk memastikan bahwa model dilatih pada sebagian besar data namun tetap dievaluasi menggunakan data baru yang belum pernah dilihat sebelumnya. Selain itu, dilakukan proses *feature_engineering* untuk menghasilkan variabel-variabel baru yang lebih informatif, seperti, total akses per hari, rata-rata selang waktu antar aktivitas, intensitas penggunaan fitur tertentu, dan proporsi aktivitas aktif dibanding pasif.

Proses ini dilakukan dengan memastikan bahwa setiap kategori aktivitas hasil ekstraksi dapat dipetakan secara konsisten ke dalam skema klasifikasi yang telah ditentukan. Hal ini penting karena klasifikasi aktivitas menjadi dasar dalam analisis frekuensi dan intensitas perilaku belajar pada tahap berikutnya. Selain itu, dilakukan penyederhana nilai-nilai atribut yang memiliki variasi penamaan atau format akibat perbedaan konfigurasi LMS, seperti perbedaan istilah pada *event_name* atau variasi penulisan deskripsi aktivitas. Penyederhana ini memastikan bahwa seluruh data berada dalam satu struktur yang telah ditetapkan sehingga proses *feature_engineering*, agregasi data, dan pemodelan dapat berjalan lebih efisien dan bebas dari bias teknis yang disebabkan oleh pencatatan yang tidak konsisten.

3.6. Data Aggregation & Representation

Data mahasiswa dari berbagai mata kuliah selama empat (4) semester digabungkan menjadi satu dataset yang terstruktur. Data kemudian direpresentasikan dalam format yang memudahkan analisis, yaitu tabel aktivitas per mahasiswa dengan kolom untuk waktu login, durasi belajar, dan frekuensi interaksi. Data dikonsolidasikan per mahasiswa untuk menghitung frekuensi aktivitas serta proporsi jenis aktivitas berdasarkan waktu belajar.. Label waktu belajar ditentukan berdasarkan waktu aktivitas

terbanyak. Representasi data dibuat dalam bentuk visualisasi *heatmap* untuk memperlihatkan korelasi fitur penting pada data log mahasiswa dengan IPK mahasiswa seperti pada Gambar 4.



Gambar 4. *Heatmap* korelasi antara fitur penting dengan IPK

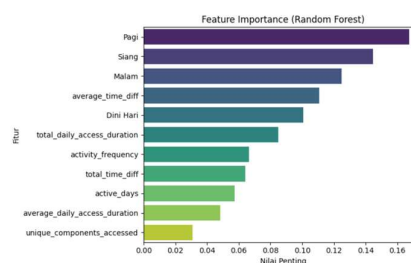
Gambar 4 menunjukkan korelasi antara fitur penting dengan IPK mahasiswa. IPK memiliki korelasi positif yang signifikan dengan fitur waktu belajar, *activity_frequency*, *active_days*, dan *total_daily_access_duration*, yang artinya frekuensi interaksi dan total waktu yang dihabiskan untuk belajar secara langsung berkorelasi positif terhadap peningkatan IPK. Fitur *average_time_diff* menunjukkan korelasi negatif, yang artinya semakin sedikit jeda antar sesi akses LMS semakin tinggi peningkatan IPK mahasiswa. Fitur *active_days* dan *activity_frequency* memiliki hubungan yang cukup kuat dengan IPK. Hal ini mengindikasikan bahwa bukan hanya waktu belajar yang menentukan keberhasilan akademik, tetapi terutama konsistensi interaksi mahasiswa dengan LMS dalam jangka Panjang. Oleh karena itu dapat disimpulkan bahwa waktu, intensitas, durasi, dan konsistensi interaksi mahasiswa dengan LMS berperan penting bagi keberhasilan akademis mahasiswa.

3.7. Data Analysis

Pada tahap ini, algoritma analitik diterapkan untuk mengidentifikasi pola belajar mahasiswa. *Random Forest* digunakan untuk mengklasifikasikan waktu belajar berdasarkan fitur frekuensi aktivitas, jenis aktivitas, dan waktu akses LMS. Model diuji dengan menggunakan metode *grid search* untuk tuning parameter. *Hyperparameter tuning* dilakukan untuk memperoleh kombinasi nilai parameter terbaik untuk meningkatkan performa model secara signifikan [31]. Metode *grid search* menghasilkan konfigurasi optimasi dengan *max_depth: None*, *min_samples_leaf: 1*, *min_samples_split: 2*, dan *n_estimators: 200*.

Model *Random Forest* memberikan interpretasi terhadap *feature importance*, yang menunjukkan bahwa fitur waktu dan jenis aktivitas tertentu memiliki pengaruh signifikan terhadap klasifikasi waktu belajar mahasiswa. Dari model *Random Forest* yang diperoleh dilakukan analisis *feature importance* seperti pada Gambar 5. Waktu belajar merupakan prediktor paling dominan dalam menentukan performa akademik mahasiswa, sedangkan korelasinya terhadap prestasi akademik relatif rendah. Fitur

lainnya seperti konsistensi akses, frekuensi aktivitas, dan durasi interaksi harian memiliki nilai penting relatif lebih rendah, meskipun memiliki peran yang lebih signifikan terhadap keberhasilan akademik mahasiswa. Temuan ini menunjukkan bahwa fitur waktu juga bergantung pada fitur lain dalam menentukan IPK mahasiswa.



Gambar 5. Feature Importance

Evaluasi pada test set dapat dilihat pada Tabel 1. Hasil evaluasi ini menunjukkan model klasifikasi waktu belajar memiliki performa yang baik, terutama pada kelas mayoritas, yaitu pagi, siang dan malam dengan nilai *precision*, *recall*, dan *f1-score* yang tinggi. Sedangkan model tidak dapat memprediksi pada kelas dini hari. Hal ini disebabkan oleh jumlah data dengan kelas dini hari terlalu sedikit. Nilai *accuracy* dan *weighted average* yang diperoleh adalah 88%, sedangkan *macro average* yang diperoleh adalah 64%. Hal ini menunjukkan bahwa model memiliki performa yang baik pada kelas mayoritas, tetapi tidak mampu menangani kelas minoritas.

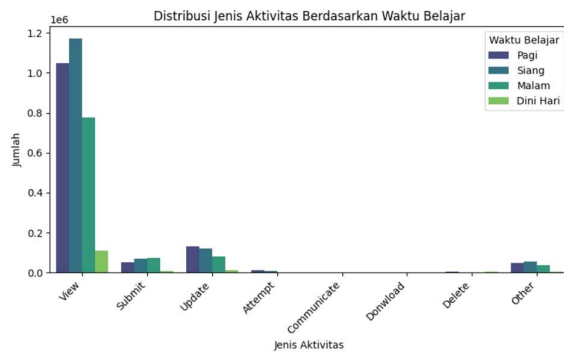
Tabel 1. Hasil evaluasi klasifikasi waktu belajar

	Precision	Recall	F1-score
Pagi	0.93	0.80	0.86
Siang	0.85	0.96	0.90
Malam	0.92	0.80	0.86
Dini Hari	0.00	0.00	0.00
Accuracy			0.88
Macro avg	0.68	0.64	0.65
Weighted avg	0.88	0.88	0.88

3.8. Data Visualization

Hasil analisis ditampilkan dalam bentuk visual, seperti grafik yang menunjukkan hubungan antara durasi belajar dengan performa akademik atau dashboard interaktif yang memvisualisasikan pola belajar mahasiswa secara keseluruhan. Visualisasi ini membantu dosen dan perguruan tinggi untuk memahami temuan dengan cepat.

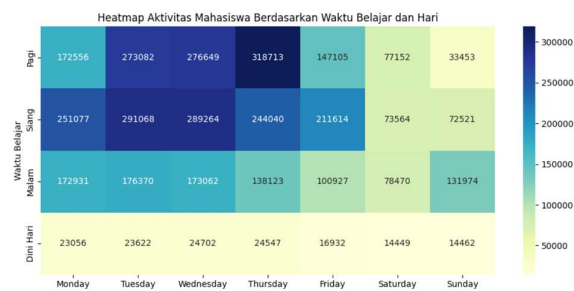
Gambar 6 menunjukkan aktivitas “View” menjadi aktivitas paling dominan pada semua waktu belajar. Aktivitas “Communicate”, “Download”, dan “Delete” memiliki jumlah sangat kecil dan kurang signifikan. Hal ini menunjukkan bahwa fitur ini jarang digunakan oleh mahasiswa. Maka dapat disimpulkan bahwa mahasiswa lebih cenderung pasif dan suka melihat materi saja dibandingkan berinteraksi langsung.



Gambar 6. Distribusi jenis aktivitas berdasarkan waktu belajar

Adanya pengaruh faktor non-teknis menyebabkan rendahnya tingkat partisipasi aktif mahasiswa, khususnya pada aktivitas diskusi dan berkomunikasi. Pada penelitian terdahulu menunjukkan bahwa hambatan seperti keterbatasan waktu, kurang percaya diri dalam forum publik, dan preferensi pembelajaran pasif dapat mempengaruhi partisipasi dalam LMS [8] [10]. Selain itu beban kuliah yang tinggi dan keterbatasan akses internet yang stabil dapat menjadi penghambat adanya interaksi aktif. Sehingga perlu strategi untuk meningkatkan partisipasi mahasiswa, seperti merancang tugas yang berorientasi pada kerja sama dan umpan balik langsung dari dosen.

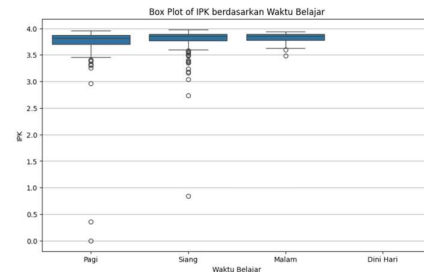
Gambar 7 menunjukkan aktivitas belajar mahasiswa tertinggi adalah pada hari Senin sampai dengan Kamis dengan waktu belajar pagi dan siang, sedangkan pada hari Jumat sampai dengan hari Minggu terjadi penurunan aktivitas. Hal ini menunjukkan bahwa Senin sampai dengan Kamis merupakan waktu paling banyak digunakan mahasiswa untuk belajar. Sedangkan waktu paling jarang digunakan mahasiswa untuk belajar adalah Sabtu dan Minggu. Hal ini menunjukkan bahwa akhir pekan bukan waktu favorit mahasiswa untuk belajar secara intensif, meskipun terjadi kenaikan aktivitas mahasiswa pada hari Minggu dan waktu belajar malam. Secara umum, waktu belajar siang merupakan waktu paling aktif terutama pada hari kerja, sedangkan waktu belajar malam merupakan waktu yang paling konsisten dengan tingkat aktivitas yang relatif stabil setiap hari.



Gambar 7. Heatmap aktivitas mahasiswa berdasarkan waktu belajar dan hari

Gambar 8 menunjukkan bahwa mahasiswa yang cenderung belajar di waktu pagi dan siang memiliki median IPK lebih tinggi dibandingkan kelompok belajar malam dan dini hari. Temuan ini mengindikasikan bahwa waktu belajar yang lebih terstruktur dan selaras dengan ritme aktivitas akademik formal berpotensi menghasilkan performa akademik yang lebih baik. Mahasiswa

yang aktif pada pagi dan siang hari kemungkinan memiliki pola belajar yang lebih konsisten, tingkat fokus yang lebih optimal, serta manajemen waktu yang lebih efektif sehingga mampu memaksimalkan pemanfaatan sumber belajar dalam LMS. Sebaliknya, mahasiswa yang dominan belajar pada malam atau dini hari mungkin menghadapi keterbatasan energi, beban aktivitas harian, atau kecenderungan menunda pekerjaan sehingga kualitas interaksi dengan materi pembelajaran menjadi kurang optimal. Secara keseluruhan, pola ini mempertegas pentingnya keteraturan waktu belajar dan konsistensi aktivitas akademik dalam mendukung pencapaian IPK yang lebih tinggi.



Gambar 8. Hubungan antara IPK dengan waktu belajar

Selain analisis berdasarkan waktu belajar, dilakukan juga pengelompokan mahasiswa berdasarkan tingkat aktivitas keseluruhan pada LMS. Mahasiswa dengan aktivitas tinggi menunjukkan kecenderungan lebih stabil dalam mengakses materi dan mengumpulkan tugas tepat waktu. Sebaliknya, mahasiswa dengan aktivitas rendah cenderung hanya mengakses LMS ketika mendekati tenggat tugas. Temuan ini sejalan dengan penelitian sebelumnya [10] yang menunjukkan bahwa mahasiswa dengan kemampuan mengatur jadwal belajar secara mandiri memperlihatkan pola akses yang lebih konsisten dan berdampak pada peningkatan performa akademik

3.9. Utilization & Analysis Results

Tahap terakhir adalah penerapan hasil analisis untuk mendukung pengambilan keputusan atau tindakan strategis. Hasil analisis digunakan untuk memberikan rekomendasi kepada dosen dan pengelola LMS untuk merancang strategi intervensi berbasis data.

Dosen dapat memanfaatkan hasil analisis untuk identifikasi mahasiswa yang berisiko mengalami kesulitan akademik, mendorong mahasiswa untuk lebih aktif menggunakan LMS, merancang aktivitas yang beragam yang dapat menarik minat mahasiswa, mengurangi beban tugas sesuai pada akhir pekan, memanfaatkan waktu malam hari untuk penguatan belajar mandiri, serta menyesuaikan waktu komunikasi untuk menyampaikan informasi penting. Selain itu, waktu belajar yang teridentifikasi juga dapat digunakan sebagai dasar personalisasi pengalaman belajar secara berkelanjutan. Pengelola LMS dapat mendukung upaya peningkatan performa akademik dengan menyediakan fitur analitik yang mudah diakses oleh dosen, memberikan notifikasi mahasiswa berisiko, serta mengoptimalkan kinerja LMS di waktu belajar mayoritas mahasiswa.

Rekomendasi ini bertujuan untuk memanfaatkan data aktivitas mahasiswa untuk mendukung dosen dalam proses belajar

mengajar dan membantu pengelola LMS dalam menyediakan lingkungan belajar digital yang lebih efektif dan suportif. Selain itu, hasil analisis ini juga dapat dimanfaatkan oleh pihak fakultas atau institusi dalam merancang kebijakan akademik berbasis data. Informasi mengenai pola belajar mahasiswa dapat digunakan untuk meninjau kembali jadwal kuliah, beban mata kuliah tertentu, atau periode penugasan agar lebih sesuai dengan ritme belajar yang telah teridentifikasi. Dengan demikian, kebijakan pendidikan tidak lagi hanya mengandalkan pendekatan normatif, tetapi didorong oleh pemahaman empiris mengenai bagaimana mahasiswa berinteraksi dengan sistem pembelajaran digital. Pendekatan ini juga membuka peluang bagi pengembangan strategi pembelajaran campuran (blended learning) yang lebih responsif terhadap kebutuhan mahasiswa.

Temuan ini juga memiliki implikasi penting dalam konteks manajemen risiko akademik. Dengan memanfaatkan pola aktivitas digital, institusi dapat mengembangkan sistem *early warning* yang mampu mendeteksi mahasiswa yang menunjukkan penurunan aktivitas atau perubahan pola belajar yang signifikan. Intervensi dapat dilakukan lebih cepat melalui konseling akademik, pendampingan belajar, atau penyesuaian strategi pengajaran. Penggunaan sistem peringatan dini berbasis analitik ini terbukti efektif dalam berbagai penelitian sebelumnya, terutama dalam meningkatkan retensi mahasiswa dan menurunkan tingkat kegagalan akademik.

Hasil analisis dapat menjadi dasar untuk pengembangan fitur personalisasi pembelajaran dalam LMS. Dengan mengetahui waktu belajar yang dominan, tingkat aktivitas, dan jenis interaksi yang paling sering dilakukan mahasiswa, sistem dapat merekomendasikan materi tambahan, mengatur pengingat tertarget, atau menyarankan strategi belajar yang sesuai dengan profil masing-masing mahasiswa. Pendekatan personalisasi ini tidak hanya meningkatkan pengalaman belajar, tetapi juga membantu mahasiswa mengembangkan kebiasaan belajar mandiri yang lebih efektif.

Pemanfaatan hasil analisis ini menegaskan peran big data sebagai alat strategis untuk meningkatkan kualitas pendidikan daring. Integrasi data log LMS dengan kebijakan akademik, desain pembelajaran, dan pengelolaan institusi memungkinkan seluruh pemangku kepentingan untuk mengambil keputusan yang lebih tepat, berbasis bukti, dan berorientasi pada peningkatan pengalaman belajar mahasiswa. Pendekatan berbasis data ini tidak hanya memperkuat efektivitas pembelajaran saat ini, tetapi juga menyediakan landasan kuat untuk pengembangan sistem pendidikan digital yang lebih adaptif dan berkelanjutan di masa mendatang.

4. KESIMPULAN

Pemanfaatan big data untuk menganalisis aktivitas mahasiswa pada platform pembelajaran daring LeADS UPNVJ telah memberikan wawasan baru tentang pola belajar mahasiswa dan hubungannya dengan keberhasilan akademik mahasiswa. Dengan mengikuti siklus hidup Big Data secara sistematis, penelitian ini tidak hanya memberikan pemahaman yang mendalam tentang pola belajar mahasiswa, tetapi juga membuka peluang untuk mengintegrasikan analitik pembelajaran ke dalam kebijakan pendidikan digital yang lebih responsif dan relevan.

Hasil analisis menunjukkan bahwa tingkat partisipasi aktif mahasiswa dalam diskusi tergolong rendah, hal ini mencerminkan pola belajar yang pasif. Total waktu yang dihabiskan untuk belajar tidak terlalu berpengaruh pada IPK mahasiswa. Namun, frekuensi aktivitas, jumlah hari aktif, komponen yang diakses, dan durasi akses harian memiliki hubungan positif dengan perolehan IPK. Oleh karena itu, konsistensi dan keterlibatan mahasiswa dalam LMS sangat penting untuk mendukung keberhasilan akademik mahasiswa.

Model klasifikasi yang dibangun menggunakan algoritma *Random Forest* dengan optimasi parameter menggunakan metode *Grid Search* memiliki akurasi 88%. Namun, performa model belum optimal dalam mengklasifikasikan kelas dengan jumlah sangat kecil, sehingga diperlukan strategi penanganan data tidak seimbang pada penelitian selanjutnya. Penelitian selanjutnya diharapkan dapat menerapkan teknik penyeimbangan data untuk meningkatkan akurasi pada kelas minoritas atau kelas dengan jumlah kecil.

Hasil penelitian ini memberikan implikasi praktis yang penting bagi berbagai pemangku kepentingan dalam ekosistem pembelajaran daring. Bagi dosen, temuan ini dapat dimanfaatkan untuk memantau tingkat keterlibatan mahasiswa secara real time, merancang strategi pembelajaran yang lebih adaptif, serta memberikan intervensi dini kepada mahasiswa dengan aktivitas rendah. Pendekatan berbasis data ini memungkinkan dosen untuk memahami pola perilaku mahasiswa secara lebih komprehensif dan menyesuaikan metode pengajaran agar lebih efektif.

Bagi pengelola LMS dan institusi pendidikan, hasil penelitian ini mendukung upaya peningkatan kualitas layanan pembelajaran digital. Pengelola LMS dapat mengoptimalkan kinerja sistem pada jam-jam dengan aktivitas tinggi, menyediakan fitur analitik otomatis, dan mengembangkan modul pelaporan yang lebih informatif bagi dosen. Di sisi institusional, temuan ini dapat menjadi dasar perumusan kebijakan pembelajaran berbasis data serta mendorong integrasi learning analytics dalam sistem penjaminan mutu internal, sehingga pengambilan keputusan pendidikan menjadi lebih responsif dan berbasis evidensi.

UCAPAN TERIMA KASIH

Terima Kasih kepada Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas Pembangunan Nasional “Veteran” Jakarta yang telah memberikan bantuan pendanaan pada skema RISTA dan semua pihak yang terlibat pada penelitian ini.

DAFTAR PUSTAKA

- [1] LeADS UPNVJ, “LMS LeADS UPNVJ.” Accessed: Jul. 30, 2025. [Online]. Available: <https://leads.upnvj.ac.id/#:~:text=Learning%20Management%20System%20,diharapkan%20dapat%20tercapai%20secara%20efektif>
- [2] Y. A. Abdillah, I. Q. Utami, W. B. M. Setiawan, M. R. Pratama, I. R. Afani, and A. Y. Pramesti, “Design and

- Development of Interactive Moodle using Design Thinking to Support Online Learning,” *Elinvo (Electronics, Informatics, and Vocational Education)*, vol. 9, no. 2, pp. 208–219, Oct. 2024, doi: [10.21831/elinvo.v9i2.75255](https://doi.org/10.21831/elinvo.v9i2.75255).
- [3] S. Muhuri and D. Mukhopadhyay, “Extracting the Relationships Among Students Based on Accessing Pattern of Digital Learning Attributes,” *IEEE Transactions on Learning Technologies*, vol. 15, no. 6, pp. 747–756, Dec. 2022, doi: [10.1109/TLT.2022.3166537](https://doi.org/10.1109/TLT.2022.3166537).
- [4] D. Shafiq, M. Marjani, R. Habeeb, and D. Asirvatham, “Digital Footprints of Academic Success: An Empirical Analysis of Moodle Logs and Traditional Factors for Student Performance,” *Educ Sci (Basel)*, vol. 15, no. 3, p. 304, Feb. 2025, doi: [10.3390/educsci15030304](https://doi.org/10.3390/educsci15030304).
- [5] E. Banjarnahor, D. P. Sibarani, B. Wibawanta, D. A. G. Sihotang, and Y. A. S. Abraham, “A Machine Learning Approach to Predicting Student Success Through Data Mining of LMS Moodle Activity Data,” in *2025 4th International Conference on Electronics Representation and Algorithm (ICERA)*, IEEE, Jun. 2025, pp. 233–238. doi: [10.1109/ICERA66156.2025.11086633](https://doi.org/10.1109/ICERA66156.2025.11086633).
- [6] M. Riestra-González, M. del P. Paule-Ruiz, and F. Ortin, “Massive LMS log data analysis for the early prediction of course-agnostic student performance,” *Comput Educ*, vol. 163, p. 104108, Apr. 2021, doi: [10.1016/j.compedu.2020.104108](https://doi.org/10.1016/j.compedu.2020.104108).
- [7] M. M. Tamada, R. Giusti, and J. F. de Magalhaes Netto, “Predicting Student Performance Based on Logs in Moodle LMS,” in *2021 IEEE Frontiers in Education Conference (FIE)*, IEEE, Oct. 2021, pp. 1–8. doi: [10.1109/FIE49875.2021.9637274](https://doi.org/10.1109/FIE49875.2021.9637274).
- [8] S. Maloney *et al.*, “Using LMS Log Data to Explore Student Engagement with Coursework Videos,” *Online Learning*, vol. 26, no. 4, Dec. 2022, doi: [10.24059/olj.v26i4.2998](https://doi.org/10.24059/olj.v26i4.2998).
- [9] R. C. Raga and J. D. Raga, “A comparison of college faculty and student class activity in an online learning environment using course log data,” in *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, IEEE, Aug. 2017, pp. 1–6. doi: [10.1109/UIC-ATC.2017.8397475](https://doi.org/10.1109/UIC-ATC.2017.8397475).
- [10] E. Araka, R. Oboko, E. Maina, and R. Gitonga, “Using Educational Data Mining Techniques to Identify Profiles in Self-Regulated Learning: An Empirical Evaluation,” *The International Review of Research in Open and Distributed Learning*, vol. 23, no. 1, pp. 131–162, Feb. 2022, doi: [10.19173/irrodl.v22i4.5401](https://doi.org/10.19173/irrodl.v22i4.5401).
- [11] M. Cantabella, R. Martínez-España, B. Ayuso, J. A. Yáñez, and A. Muñoz, “Analysis of student behavior in learning management systems through a Big Data framework,” *Future Generation Computer Systems*, vol. 90, pp. 262–272, Jan. 2019, doi: [10.1016/j.future.2018.08.003](https://doi.org/10.1016/j.future.2018.08.003).
- [12] A. Lulli, L. Oneto, and D. Anguita, “Mining Big Data with Random Forests,” *Cognit Comput*, vol. 11, no. 2, pp. 294–316, Apr. 2019, doi: [10.1007/s12559-018-9615-4](https://doi.org/10.1007/s12559-018-9615-4).
- [13] F. E. EL Habti, “Enhancing Student Performance Prediction in e-Learning Ecosystems Using Machine Learning Techniques,” *International Journal of Information and Education Technology*, vol. 15, no. 2, pp. 301–311, 2025, doi: [10.18178/ijiet.2025.15.2.2243](https://doi.org/10.18178/ijiet.2025.15.2.2243).
- [14] Y. A. Saadoon and R. H. Abdulmir, “Improved Random Forest Algorithm Performance For Big Data,” *J Phys Conf Ser*, vol. 1897, no. 1, p. 012071, May 2021, doi: [10.1088/1742-6596/1897/1/012071](https://doi.org/10.1088/1742-6596/1897/1/012071).
- [15] R. I. E. Saragih, “Predicting Student Academic Performance Using Random Forest Regression: A Case Study on LMS Behavioral Data,” *International Journal of Information System and Innovative Technology*, vol. 4, no. 1, pp. 11–14, Jun. 2025, doi: [10.63322/rm0wcg63](https://doi.org/10.63322/rm0wcg63).
- [16] A. Bessadok, E. Abouzinadah, and O. Rabie, “Exploring students digital activities and performances through their activities logged in learning management system using educational data mining approach,” *Interactive Technology and Smart Education*, vol. 20, no. 1, pp. 58–72, Feb. 2023, doi: [10.1108/ITSE-08-2021-0148](https://doi.org/10.1108/ITSE-08-2021-0148).
- [17] K. Jawad, M. A. Shah, and M. Tahir, “Students’ Academic Performance and Engagement Prediction in a Virtual Learning Environment Using Random Forest with Data Balancing,” *Sustainability*, vol. 14, no. 22, p. 14795, Nov. 2022, doi: [10.3390/su142214795](https://doi.org/10.3390/su142214795).
- [18] Á. Hernández-García, C. Cuenca-Enrique, L. Del-Río-Carazo, and S. Iglesias-Pradas, “Exploring the relationship between LMS interactions and academic performance: A Learning Cycle approach,” *Comput Human Behav*, vol. 155, p. 108183, Jun. 2024, doi: [10.1016/j.chb.2024.108183](https://doi.org/10.1016/j.chb.2024.108183).
- [19] D. Liu, “Study on Online Learning Behavior Analysis and Performance Prediction Based on Improved Random Forest Algorithm,” *Creat Educ*, vol. 14, no. 08, pp. 1527–1535, 2023, doi: [10.4236/ce.2023.148097](https://doi.org/10.4236/ce.2023.148097).
- [20] M. Adnan *et al.*, “Predicting at-Risk Students at Different Percentages of Course Length for Early Intervention Using Machine Learning Models,” *IEEE Access*, vol. 9, pp. 7519–7539, 2021, doi: [10.1109/ACCESS.2021.3049446](https://doi.org/10.1109/ACCESS.2021.3049446).
- [21] Y. Bai, M. Bosu, and D. Abuaiadah, “Predicting Learning Outcomes in an Online Learning Platform,” in *Proceedings: CITREnz 2023 Conference, Auckland, 27-29 September*, Unitec ePress, Jul. 2024, pp. 78–90. doi: [10.34074/proc.240111](https://doi.org/10.34074/proc.240111).
- [22] T. Erl, W. Khattak, and P. Buhler, *Big data fundamentals: concepts, drivers & techniques*. Prentice Hall Press, 2016.
- [23] G. G. Hamshin, A. Y. Omar, O. S. Said, and S. S. Saleh, “Leveraging big data analytics to enhance E-learning services,” *Int Neurourol J*, vol. 27, no. 4, pp. 1529–1542, 2023.
- [24] O. Ovtšarenko, “Innovative techniques for e-learning log data processing: Trends and methods,” *Journal of Innovation & Knowledge*, vol. 10, no. 5, p. 100765, Sep. 2025, doi: [10.1016/j.jik.2025.100765](https://doi.org/10.1016/j.jik.2025.100765).
- [25] H. Ahmed and M. A. Ismail, “A Structured Approach Towards Big Data Identification,” *IEEE Trans Big Data*,

- vol. 9, no. 1, pp. 147–159, Feb. 2023, doi: [10.1109/TBDATA.2021.3139069](https://doi.org/10.1109/TBDATA.2021.3139069).
- [26] D. Garcia-Gil, J. Luengo, S. García, and F. Herrera, “Enabling Smart Data: Noise filtering in Big Data classification,” *InfSci (N Y)*, vol. 479, pp. 135–152, Apr. 2019, doi: [10.1016/j.ins.2018.12.002](https://doi.org/10.1016/j.ins.2018.12.002).
- [27] Y. Li *et al.*, “Cluster-based data filtering for manufacturing big data systems,” *Journal of Quality Technology*, vol. 54, no. 3, pp. 290–302, May 2022, doi: [10.1080/00224065.2021.1889420](https://doi.org/10.1080/00224065.2021.1889420).
- [28] R. Nair and A. Bhagat, “A Life Cycle on Processing Large Dataset - LCPL,” *Int J Comput Appl*, vol. 179, no. 53, pp. 27–34, Jun. 2018, doi: [10.5120/ijca2018917382](https://doi.org/10.5120/ijca2018917382).
- [29] C. Xie, J. Gao, and C. Tao, “Big Data Validation Case Study,” in *2017 IEEE Third International Conference on Big Data Computing Service and Applications (BigDataService)*, IEEE, Apr. 2017, pp. 281–286. doi: [10.1109/BigDataService.2017.44](https://doi.org/10.1109/BigDataService.2017.44).
- [30] C. O’HIGGINS, C. KENT, D. HESTER, and S. TAYLOR, “Validation Of Data For Use In Civil Infrastructure Big Data Applications,” in *Proceedings of the 14th International Workshop on Structural Health Monitoring*, Destech Publications, Inc., Sep. 2023. doi: [10.12783/shm2023/36757](https://doi.org/10.12783/shm2023/36757).
- [31] P. P. Ippolito, “Hyperparameter Tuning,” 2022, pp. 231–251. doi: [10.1007/978-3-030-88389-8_12](https://doi.org/10.1007/978-3-030-88389-8_12).