

# **PREDIKSI PERFORMA AKADEMIK MAHASISWA DENGAN METODE REGRESI LINEAR BERBASIS DATA KEBIASAAN HARIAN**

Nur Akhmad Van Jouvi<sup>1</sup>, Judika Marpaung<sup>2</sup>

Program Studi Teknik Informatika<sup>1,2</sup>

[nur.akhmad24@perbanas.id](mailto:nur.akhmad24@perbanas.id)<sup>1</sup>, [judika.marpaung21@perbanas.id](mailto:judika.marpaung21@perbanas.id)<sup>2</sup>

Perbanas Institute

## **ABSTRAK**

Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor utama yang memengaruhi prestasi akademik mahasiswa berdasarkan kebiasaan harian, kondisi psikologis, serta pola hidup sehari-hari. Sebuah model regresi linear diterapkan pada dataset komprehensif yang mencakup 1000 mahasiswa dan 15 variabel independen, termasuk durasi belajar, paparan media sosial, status pekerjaan paruh waktu, kualitas tidur, dan indikator kesehatan psikologis. Proses pra-pemrosesan data dan validasi silang menghasilkan model prediktif yang kuat, dibuktikan dengan nilai koefisien determinasi ( $R^2$ ) sebesar 0,89. Analisis lebih lanjut mengidentifikasi durasi belajar harian, status kesehatan mental, dan frekuensi aktivitas fisik sebagai prediktor positif yang paling signifikan terhadap nilai ujian. Sebaliknya, waktu yang dihabiskan untuk media sosial dan layanan *streaming* seperti Netflix menunjukkan korelasi negatif yang signifikan. Implikasi dari temuan ini dapat menjadi landasan bagi institusi pendidikan dan mahasiswa dalam merancang strategi intervensi untuk optimalisasi prestasi akademik.

Kata-kata kunci: Performa Akademik, Kesehatan Mental, Gaya Hidup Mahasiswa, Regresi Linear, Durasi Belajar

## **ABSTRACT**

*This study investigates the determinants of college students' academic performance by analyzing factors from the domains of daily habits, mental health, and lifestyle. A linear regression model was applied to a comprehensive dataset of 1000 students and 15 independent variables, including study duration, social media exposure, part-time job status, sleep quality, and psychological health indicators. Data pre-processing and cross-validation resulted in a robust predictive model, evidenced by a coefficient of determination ( $R^2$ ) value of 0.89. Further analysis identified daily study duration, mental health status and frequency of physical activity as the most significant positive predictors of test scores. In contrast, time spent on social media and streaming services such as Netflix showed significant negative correlations. The implications of these findings can serve as a foundation for educational institutions and students in designing intervention strategies for the optimization of academic performance.*

*Keywords: Academic Performance, Mental Health, Student Lifestyle, Linear Regression, Study Duration*

## 1. PENDAHULUAN

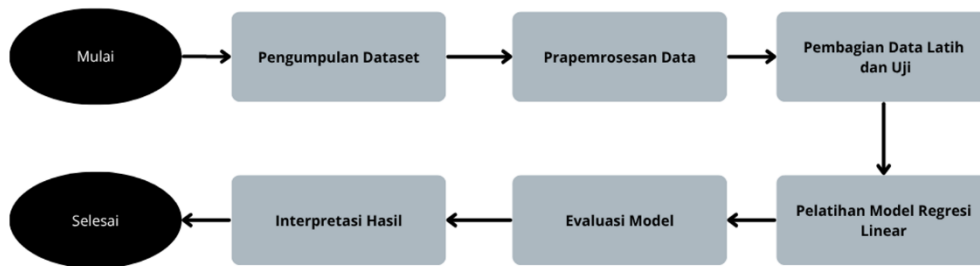
Performa akademik mahasiswa menjadi indikator penting dalam menilai keberhasilan pembelajaran di perguruan tinggi. Berbagai studi menunjukkan bahwa faktor psikososial, gaya hidup, dan kebiasaan belajar memainkan peran signifikan dalam menentukan prestasi akademik mahasiswa. Aydin (2017) mengungkapkan bahwa efikasi diri, manajemen waktu, serta keterlibatan mahasiswa dalam aktivitas akademik berkorelasi dengan kesuksesan studi. Sementara itu, Arofah et al. (2020) menjelaskan bahwa prestasi akademik dipengaruhi oleh faktor fisiologis, psikologis, keluarga, dan lingkungan.

Di tengah perkembangan teknologi digital, tantangan baru muncul. Studi oleh Azizah & Anshori (2025) menyoroti bagaimana penggunaan media sosial yang berlebihan, seperti TikTok, dapat menurunkan fokus dan produktivitas mahasiswa. Di sisi lain, kesehatan mental dan kualitas tidur juga menjadi isu krusial yang berdampak langsung pada kemampuan belajar. Suardiaz-Muro et al. (2023) menunjukkan bahwa kurang tidur, terutama saat menghadapi ujian, secara signifikan menurunkan performa akademik. Hal ini didukung oleh studi lokal dari Melly et al. (2021), yang menemukan tingginya prevalensi kualitas tidur buruk di kalangan mahasiswa kedokteran.

yang menemukan tingginya prevalensi kualitas tidur buruk di kalangan mahasiswa kedokteran. Meskipun penelitian-penelitian tersebut telah mengidentifikasi berbagai faktor secara terpisah, masih terbatas studi yang mengintegrasikan variabel kebiasaan harian, gaya hidup digital, dan kesehatan mental secara komprehensif dalam satu model prediktif kuantitatif. Celah inilah yang berusaha diisi oleh penelitian ini. Oleh karena itu, penelitian ini bertujuan untuk memodelkan hubungan antara kebiasaan harian mahasiswa dan nilai akademik menggunakan regresi linear multivariat.

Penelitian ini bertujuan untuk memodelkan hubungan antara kebiasaan harian mahasiswa dan nilai akademik menggunakan regresi linear multivariat. Dengan menggabungkan analisis kuantitatif dan referensi dari berbagai literatur, penelitian ini diharapkan dapat memberikan kontribusi dalam pemahaman serta pengembangan strategi peningkatan prestasi akademik mahasiswa.

## 2. METODE



Gambar.1 Tahapan Penelitian

Flowchart ini menunjukkan tahapan penelitian mulai dari pengumpulan data mahasiswa, prapemrosesan (pembersihan data, encoding, dan scaling), pembagian data latih dan uji, pelatihan model regresi linear, evaluasi dengan metrik statistik, hingga interpretasi hasil berdasarkan fitur yang paling berpengaruh.

### 2.1. Sumber Data

Penelitian ini menggunakan dataset primer yang dihimpun dari 1000 responden mahasiswa. Variabel yang dikumpulkan mencakup data demografis (usia, jenis kelamin), variabel perilaku (durasi belajar, penggunaan media sosial, jam tidur, pola makan, frekuensi olahraga, kehadiran), serta indikator psikologis (kondisi kesehatan mental). Variabel target yang menjadi proksi performa akademik adalah skor ujian akhir mahasiswa.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
student_id	age	gender	study_hou	social_mei	netflix_hoi	part_time	attendanc	sleep_hou	diet_qualit	exercise_f	parental_ε	internet_q	mental_he	extracurric	exam_score	
S1000	23	Female	0	1.2	1.1	No	85	8	Fair	6	Master	Average	8	Yes	56.2	
S1001	20	Female	6.9	2.8	2.3	No	97.3	4.6	Good	6	High Schoc	Average	8	No	100	
S1002	21	Male	1.4	3.1	1.3	No	94.8	8	Poor	1	High Schoc	Poor	1	No	34.3	
S1003	23	Female	1	3.9	1	No	71	9.2	Poor	4	Master	Good	1	Yes	26.8	
S1004	19	Female	5	4.4	0.5	No	90.9	4.9	Fair	3	Master	Good	1	No	66.4	
S1005	24	Male	7.2	1.3	0	No	82.9	7.4	Fair	1	Master	Average	4	No	100	
S1006	21	Female	5.6	1.5	1.4	Yes	85.8	6.5	Good	2	Master	Poor	4	No	89.8	
S1007	21	Female	4.3	1	2	Yes	77.7	4.6	Fair	0	Bachelor	Average	8	No	72.6	
S1008	23	Female	4.4	2.2	1.7	No	100	7.1	Good	3	Bachelor	Good	1	No	78.9	
S1009	18	Female	4.8	3.1	1.3	No	95.4	7.5	Good	5	Bachelor	Good	10	Yes	100	
S1010	19	Female	4.6	3.7	0.8	No	77.6	5.8	Fair	1	None	Good	3	No	63.3	
S1011	23	Male	3.9	2.4	2.5	No	71.7	7.9	Fair	2	Bachelor	Average	1	No	74.4	
S1012	19	Female	3.7	2.1	0.4	Yes	81.1	4.5	Fair	1	Bachelor	Good	9	No	76.9	
S1013	19	Female	3.4	2.7	2.7	No	89.3	4.7	Fair	4	Bachelor	Good	10	No	75.8	
S1014	24	Male	2.4	1.5	0.7	No	87.4	6.7	Poor	6	Bachelor	Average	9	No	78.9	
S1015	21	Male	3.1	5	1	No	97.5	6.5	Good	6	High Schoc	Average	7	No	74	
S1016	20	Male	1	0.6	0.2	No	92.9	5.6	Poor	3	High Schoc	Poor	8	Yes	55.2	
S1017	24	Female	3.4	2.7	1.2	No	94.7	7.5	Poor	0	High Schoc	Average	1	Yes	70.8	
S1018	24	Other	2	4.9	2.9	Yes	88.3	7.1	Good	2	High Schoc	Good	5	No	43.9	
S1019	19	Male	1.8	2.5	2.4	No	71.1	7.5	Fair	1	High Schoc	Average	2	No	45.3	
S1020	22	Female	3.8	2.3	2.9	No	83	6.4	Good	3	Master	Good	1	No	58.5	
S1021	21	Male	5.6	2.1	2.4	No	95.6	7.2	Fair	1	High Schoc	Good	3	No	82.5	
S1022	18	Other	4.9	2.3	0.6	No	84.5	6	Fair	3	High Schoc	Average	7	No	98.7	
S1023	24	Female	1.1	4.1	1.4	Yes	90	9	Fair	6	Bachelor	Good	1	No	43.7	
S1024	20	Female	2	0	0.9	Yes	81.8	5.5	Fair	4	Bachelor	Average	2	No	54.9	
S1025	22	Male	4.9	4.3	3.3	No	74.7	9	Fair	1	High Schoc	Average	2	Yes	69.9	
S1026	22	Male	2	0.8	0.5	No	83.8	6.5	Good	4	Bachelor	Poor	4	No	73.5	
S1027	18	Male	3.2	2.2	2.8	Yes	88.1	4.8	Fair	5	Bachelor	Average	3	No	71.1	
S1028	24	Male	4.3	2	2.8	Yes	78.4	8.1	Good	5	High Schoc	Poor	3	Yes	82.8	
S1029	20	Female	2	3.2	3.8	Yes	82.6	6.7	Poor	6	High Schoc	Poor	10	Yes	75.7	

## Gambar.2 Dataset Mentah

### 2.2. Pra-pemrosesan Data

Tahapan pra-pemrosesan data diawali dengan penanganan nilai yang hilang (*missing values*). Kolom `parental_education_level`, yang memiliki 91 data kosong, dieksklusi dari analisis. Selanjutnya, transformasi data dilakukan pada variabel kategorikal seperti jenis kelamin, status pekerjaan paruh waktu, kualitas diet, dan kualitas internet, dengan mengaplikasikan metode *one-hot encoding*. Untuk memastikan komparabilitas skala antar variabel, fitur-fitur numerik seperti durasi belajar dan jam tidur distandarisasi menggunakan algoritma `StandardScaler` sehingga memiliki rata-rata nol dan standar deviasi satu. Proses ini menghasilkan dataset final yang terdiri dari 909 observasi (baris) dan 18 fitur numerik yang siap untuk dimodelkan.

### 2.3. Pemodelan Regresi

Untuk mengkuantifikasi hubungan antara variabel prediktor dan performa akademik, digunakan model regresi linear multivariat. Pendekatan ini dipilih karena kemampuannya dalam menjelaskan hubungan linier secara simultan, tetapi juga karena tingkat interpretasi koefisiennya yang tinggi. Keunggulan ini memungkinkan identifikasi faktor-faktor paling berpengaruh secara jelas dan kuantitatif, yang menjadi tujuan utama dari penelitian ini.. Dataset dibagi secara acak menjadi data latih (80%) dan data uji (20%). Kinerja model dievaluasi berdasarkan metrik *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), dan koefisien determinasi ( $R^2$ ). Guna memastikan robustisitas dan menghindari *overfitting*, stabilitas model diuji lebih lanjut menggunakan prosedur validasi silang *5-fold cross-validation*.

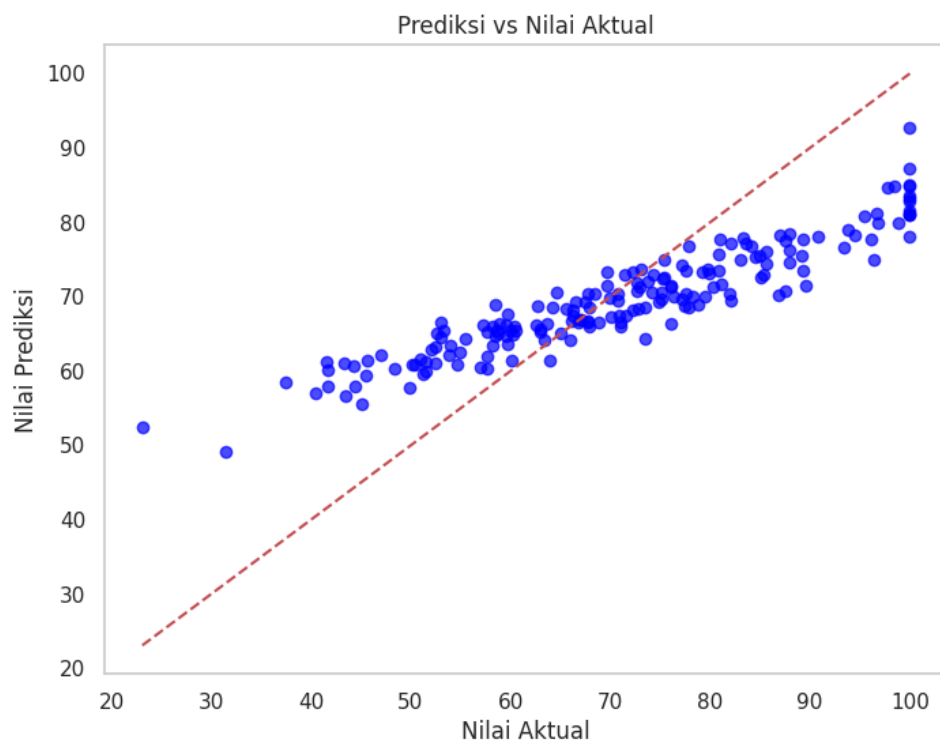
### 3. HASIL DAN PEMBAHASAN

#### 3.1. Evaluasi Kinerja Model

Kinerja prediktif model regresi dievaluasi pada data uji dan menunjukkan hasil yang sangat memuaskan, dengan metrik sebagai berikut:

➡	MSE	: 30.01
	RMSE	: 5.48
	MAE	: 4.24
	$R^2$	: 0.89

Nilai  $R^2$  sebesar 0,89 mengindikasikan bahwa model yang dikembangkan mampu menjelaskan 89% variabilitas dalam skor ujian mahasiswa. Tingkat kesalahan prediksi, sebagaimana tecermin dari nilai MSE, RMSE, dan MAE yang relatif rendah, menunjukkan akurasi model yang tinggi. Lebih lanjut, hasil validasi silang menghasilkan rata-rata  $R^2$  yang konsisten (0.89), memberikan bukti kuat bahwa model ini bersifat stabil dan tidak mengalami *overfitting*.



**Gambar.3 Visualisaisi Hasil**

Gambar menunjukkan sebaran antara nilai prediksi model dan nilai aktual dari skor ujian mahasiswa. Titik-titik biru merepresentasikan pasangan nilai

aktual dan prediksi untuk masing-masing mahasiswa, sementara garis putus-putus merah menggambarkan garis ideal (prediksi = aktual). Sebagian besar titik berada dekat dengan garis diagonal, yang menandakan bahwa model memiliki akurasi prediksi yang baik dan konsisten. Hal ini juga selaras dengan hasil evaluasi model yang menunjukkan nilai  $R^2$  sebesar 0.89, menandakan bahwa 89% variasi skor ujian dapat dijelaskan oleh model regresi linear.

### 3.2.Analisis Faktor-Faktor Signifikan

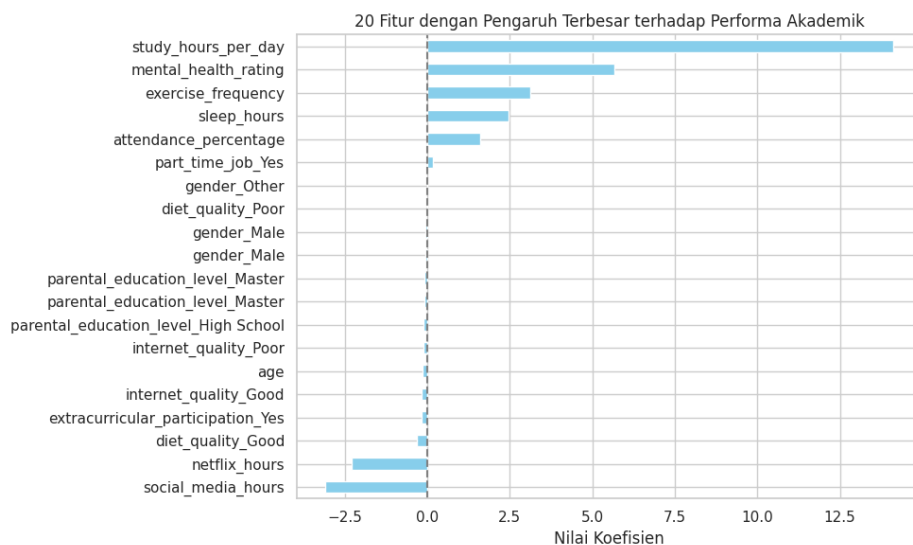
Analisis terhadap koefisien regresi memungkinkan identifikasi variabel-variabel yang paling berpengaruh terhadap performa akademik. Koefisien ini merepresentasikan besaran perubahan pada skor ujian untuk setiap kenaikan satu unit pada variabel prediktor.

Tabel.1: Lima Prediktor Positif Teratas

Fitur	Koefisien
study_hours_per_day	+14.11
mental_health_rating	+5.65
exercise_frequency	+3.11
Sleep_hours	+2.46
attendance_percentage	+1.59

Tabel.2: Lima Prediktor Negatif Teratas

Fitur	Koefisien
social_media_hours	-3.08
netflix_hours	-2.30
diet_quality_good	-0.32
extracurricular_participation_Yes	-0.18
internet_quality_Good	-0.17



### Gambar.3 Koefisien Regresi

Gambar di atas menampilkan 20 fitur dengan nilai koefisien terbesar berdasarkan model regresi linear. Fitur `study_hours_per_day` memiliki pengaruh positif paling kuat terhadap performa akademik, diikuti oleh `mental_health_rating`, `exercise_frequency`, `sleep_hours`, dan `attendance_percentage`. Artinya, peningkatan pada faktor-faktor tersebut cenderung meningkatkan skor ujian mahasiswa.

Sebaliknya, fitur seperti `social_media_hours`, `netflix_hours`, dan `diet_quality_Good` memiliki koefisien negatif, menandakan bahwa peningkatan nilai pada variabel-variabel ini cenderung menurunkan skor ujian. Visualisasi ini memperkuat temuan bahwa kebiasaan belajar, kesehatan fisik dan mental memiliki peran besar dalam pencapaian akademik, sementara distraksi digital berkontribusi terhadap penurunan performa

### 3.3.Diskusi Temuan dan Keterkaitan dengan Literatur

Temuan kuantitatif di atas konsisten dan memperkuat berbagai hasil penelitian sebelumnya. Pengaruh positif terkuat dari durasi belajar (`study_hours_per_day`) sejalan dengan argumen Aydin (2017) mengenai pentingnya manajemen waktu. Selanjutnya, pengaruh signifikan dari kesehatan mental (`mental_health_rating`) menegaskan kembali temuan Hysenbegasi et al. (2005) tentang dampak kondisi depresi. Peran krusial tidur (`sleep_hours`), seperti yang ditunjukkan oleh Suardiaz-Muro et al. (2023) dan Melly et al. (2021), juga terkonfirmasi dalam model ini. Di sisi lain, dampak negatif dari penggunaan media sosial (`social_media_hours`) mendukung kesimpulan Azizah & Anshori (2025) terkait penurunan fokus belajar.

### 3.4.Keterbatasan Penelitian

Perlu diakui bahwa penelitian ini memiliki beberapa keterbatasan. Pertama, data yang digunakan bersifat *self-reported* (laporan diri) dari mahasiswa, yang berpotensi mengandung bias subjektivitas. Kedua, dataset berasal dari

satu lingkungan institusi sehingga generalisasi hasil ke populasi mahasiswa yang lebih luas perlu dilakukan dengan hati-hati. Ketiga, model regresi linear yang digunakan mengasumsikan adanya hubungan linier antar variabel; hubungan non-linear yang mungkin ada tidak dapat ditangkap oleh model ini. Terakhir, studi ini menunjukkan korelasi yang kuat, namun tidak secara definitif membuktikan hubungan sebab-akibat (kausalitas).

### 3.5. Implikasi Penelitian

Secara praktis, hasil penelitian ini mengimplikasikan bahwa intervensi untuk meningkatkan performa akademik mahasiswa akan lebih efektif jika berfokus pada beberapa area kunci. Bagi mahasiswa, temuan ini menggarisbawahi pentingnya alokasi waktu belajar yang disiplin, pemeliharaan kesehatan mental, serta menjaga kualitas tidur. Bagi institusi pendidikan, hasil ini menyarankan urgensi penyediaan layanan dukungan yang komprehensif, seperti program pembinaan gaya hidup sehat dan penguatan layanan konseling psikologis, sebagai bagian integral dari strategi akademik.

## 4. KESIMPULAN

Penelitian ini berhasil mengembangkan sebuah model regresi linear multivariat yang mampu memprediksi performa akademik mahasiswa dengan tingkat akurasi yang tinggi, sebagaimana ditunjukkan oleh nilai koefisien determinasi ( $R^2$ ) sebesar 0,89. Model ini mengonfirmasi bahwa kebiasaan harian, kesehatan mental, dan gaya hidup merupakan determinan signifikan bagi keberhasilan studi. Secara spesifik, durasi belajar harian (`study_hours_per_day`) teridentifikasi sebagai prediktor positif paling kuat, diikuti oleh rating kesehatan mental (`mental_health_rating`) dan frekuensi olahraga (`exercise_frequency`). Temuan ini memperkuat argumen bahwa alokasi waktu untuk belajar dan menjaga kesehatan psikofisik adalah investasi utama untuk mencapai prestasi akademik.



## DAFTAR PUSTAKA

- Arofah, I., Ningsi, B.A. & Masyhudi, L., 2020. *Analisis Faktor-Faktor yang Mempengaruhi Prestasi Akademik Mahasiswa*. Jurnal Binawakya, 15(5), pp.4511–4513. Available at: <http://ejurnal.binawakya.or.id/index.php/MBI>.
- Azizah, I.Q. & Anshori, I., 2025. *Pengaruh Penggunaan TikTok sebagai Media Sosial terhadap Fokus Belajar dan Produktivitas Mahasiswa*. JASIMA: Jurnal Komunikasi Korporasi dan Media, 6(1), pp.88–92.
- Ginting, F., Buulolo, E. & Siagian, E.R., 2019. *Implementasi Algoritma Regresi Linear Sederhana dalam Memprediksi Besaran Pendapatan Daerah*. KOMIK: Konferensi Nasional Teknologi Informasi dan Komputer, 3(1), pp.274–279. doi:10.30865/komik.v3i1.1602.
- Melly, M., Lubis, L.D., Daulay, M. & Adella, C.A., 2021. *Hubungan Kualitas Tidur dengan Fungsi Kognitif pada Mahasiswa Fakultas Kedokteran Universitas Sumatera Utara*. JIMKI: Jurnal Ilmiah Mahasiswa Kedokteran Indonesia, 9(2), pp.27–29. doi:10.53366/jimki.v9i2.474.
- Mokodompit, R.S. et al., 2025. *Meningkatkan Prestasi Akademik Mahasiswa Teknik Informatika UNSRAT Melalui Optimasi Pembelajaran dengan Random Forest*. JAMASTIKA, 4(1), pp.201–203.
- Rahmawati, A. et al., 2025. *Gangguan Kesehatan Mental terhadap Performa Akademik Mahasiswa Tingkat Awal*. Educate: Journal of Education and Learning, 3(1), pp.1–15. doi:10.61994/educate.v3i1.329.
- Aydin, G., 2017. *Personal Factors Predicting College Student Success*. Eurasian Journal of Educational Research, 69, pp.93–112. doi:10.14689/ejer.2017.69.6.