

Klasterisasi Pola Penggunaan Internet Mahasiswa Menggunakan Pendekatan *Principal Component Analysis* dan *K-Means Clustering*

Maxtulus Junedy Nababan^{1*}, Anil Hakim Syofra², Desniarti³

¹Program Studi Matematika, Universitas Mataram, Indonesia

²Program Studi Pendidikan Matematika, Universitas Asahan, Indonesia

³Program Studi Pendidikan Matematika, Universitas Muslim Nusantara Al-Washliyah, Indonesia

*Penulis Korespondensi: maxtnab@staff.unram.ac.id

Abstract. *This study aims to map and identify patterns of internet usage behavior based on usage functions and time management. As internet utilization continues to increase among students, a deeper understanding of user behavior characteristics is essential to formulate effective policies that encourage productive internet use. This research employed a quantitative approach supported by statistical analysis and machine learning techniques. Principal Component Analysis (PCA) was applied to reduce data dimensionality, while K-Means Clustering was utilized to objectively classify internet users into groups with similar behavioral characteristics. The PCA results successfully reduced 12 manifest indicators into two principal components: the dimension of structured academic utility and entertainment, and the dimension of usage intensity and dependence on internet facilities. These two latent components explained 61.5% of the total variance in internet usage behavior indicators. Furthermore, the K-Means Clustering algorithm identified three optimal clusters: academic-functional internet users ($n = 85$), consumptive and dependent internet users ($n = 88$), and independent and economical internet users ($n = 87$). The findings indicate that internet users exhibit diverse behavioral patterns, requiring cluster-based intervention strategies tailored to the characteristics of each group. Therefore, implementing cluster-oriented policies within educational institutions is crucial for improving the effectiveness, efficiency, and productivity of internet utilization while promoting more targeted and responsible digital behavior.*

Keywords: *Internet Usage; K-Means Clustering; Machine Learning; Principal Component Analysis; User Segmentation.*

Abstrak. Penelitian ini bertujuan untuk memetakan dan mengidentifikasi pola perilaku penggunaan internet berdasarkan fungsi pemanfaatan serta manajemen waktu pengguna. Seiring meningkatnya penggunaan internet di kalangan mahasiswa, diperlukan pemahaman yang lebih mendalam mengenai karakteristik perilaku pengguna agar dapat dirumuskan kebijakan yang tepat dalam mendukung pemanfaatan internet secara produktif. Penelitian ini menggunakan pendekatan kuantitatif dengan bantuan analisis statistik dan metode *machine learning*. Teknik *Principal Component Analysis (PCA)* digunakan untuk mereduksi dimensi data, sedangkan *K-Means Clustering* diterapkan untuk melakukan segmentasi pengguna internet secara objektif berdasarkan karakteristik perilaku yang serupa. Hasil analisis PCA berhasil mereduksi 12 indikator manifest menjadi dua komponen utama, yaitu dimensi utilitas akademik dan hiburan yang terstruktur serta dimensi intensitas penggunaan dan ketergantungan terhadap fasilitas internet. Kedua komponen tersebut mampu menjelaskan sebesar 61,5% varians total dari seluruh indikator perilaku penggunaan internet. Selanjutnya, hasil analisis *K-Means Clustering* menghasilkan tiga kluster optimal, yaitu pengguna internet fungsional akademik ($n = 85$), pengguna internet konsumtif dan bergantung ($n = 88$), serta pengguna internet mandiri dan hemat ($n = 87$). Temuan penelitian menunjukkan bahwa karakteristik perilaku pengguna internet memiliki pola yang berbeda sehingga memerlukan strategi intervensi yang disesuaikan dengan karakteristik masing-masing kelompok. Oleh karena itu, implementasi kebijakan berbasis kluster di lingkungan institusi pendidikan menjadi langkah yang penting untuk meningkatkan efektivitas, efisiensi, dan produktivitas penggunaan internet secara tepat sasaran.

Kata kunci: *K-Means Clustering; Machine Learning; Penggunaan Internet; Principal Component Analysis; Segmentasi Pengguna.*

1. LATAR BELAKANG

Perkembangan teknologi informasi memberikan dampak yang sangat besar bagi kehidupan. Informasi telah menjadi komoditi yang sangat penting dan bisa diperoleh secara cepat dengan ketersediaan internet. Penggunaan internet sangat pada saat ini sudah

merupakan salah satu kebutuhan karena adanya sifat perilaku manusia yang memiliki ketertarikan untuk mengetahui kondisi atau keadaan sekitar (Maharani, 2021; Syukri, 2025). Kelompok masyarakat yang memiliki peluang besar untuk menggunakan internet adalah mahasiswa karena fasilitas internet dalam lingkungan pendidikan sudah harus tersedia untuk meningkatkan kualitas pelaksanaan pendidikan (Alvaridzi, 2025; Alpianoor, 2025). Dalam dunia pendidikan internet bukan hanya sebagai fasilitas pendukung tetapi menjadi infrastruktur primer dalam akademik dan personal mahasiswa. Saat ini mahasiswa dalam lingkungan universitas sudah harus terhubung dengan penggunaan internet baik itu mengakses platform pembelajaran, penelusuran artikel ilmiah, berinteraksi dengan dosen maupun rekan belajar.

Salah satu factor yang membentuk karakteristik penggunaan internet adalah manajemen waktu. Tingginya penggunaan internet saat ini ditandai dengan mobilitas Hp yang sangat tinggi yang mengakibatkan meningkatnya akses terhadap internet melebihi 5 jam perhari (Putri, 2025). Disatu sisi kecenderungan akses internet yang berlebihan dapat menurunkan dan mengganggu konsentrasi belajar mahasiswa (Susanti, 2025). Ketika tidak ada kontrol terhadap penggunaan teknologi juga secara tidak langsung dapat merusak kesehatan fisik mahasiswa karena aktifitas siber yang sampai larut malam dan mengganggu waktu istirahat (Marciando, 2022). Disisi lain mahasiswa dengan perilaku konsumtif internet yang tinggi tetapi aktivitasnya didominasi akses internet riset akademik merupakan perilaku yang positif (Fuada, 2023). Penggunaan internet yang rendah tetapi dilakukan pada jam kuliah dengan tujuan hiburan maka hal tersebut menjadi perilaku yang negatif. Untuk mengurangi sisi negatif dari perilaku penggunaan internet oleh mahasiswa adalah kontrol diri dari mahasiswa tersebut (Febrina, 2025).

Factor lain yang membentuk pola perilaku penggunaan internet mahasiswa adalah aspek finansial dan ketersediaan infrastruktur. Fenomena perilaku mahasiswa dalam menggunakan internet sangat variatif. Ada mahasiswa secara mandiri selalu memastikan ketersediaan paket internet agar dapat mengakses internet kapanpun dan dimanapun. Tetapi sebagian mahasiswa ada yang mempunyai ketergantungan terhadap akses internet dengan fasilitas kampus dimana mahasiswa rela menunda mendownload file akademik sebelum terhubung dengan wifi universitas (Listya, 2023).

Perilaku penggunaan internet cenderung sulit untuk direduksi menjadi variable-variable fisik yang dapat diukur. Perilaku penggunaan internet pada umumnya melibatkan banyak variable yang saling berkorelasi yang kemungkinan saling memiliki multikolinieritas dalam pemodelan statistika kuantitatif. Untuk menyederhakan dan mengelompokkan

variable-variabel tersebut secara subjektif dilakukan dengan mengintegrasikan dua algoritma multivariate dan *unsupervised machine learning*. Pendekatan yang diaplikasikan adalah *Principal Component Analysis* untuk mereduksi dimensi dari variable-variabel yang ada kemudian dilakukan pengklasteran dengan pemodelan K-means clustering.

2. KAJIAN TEORITIS

Principal Component Analysis (PCA)

Principal Component Analysis adalah salah satu metode analisis multivariate yang berfungsi untuk menyederhanakan data dari beberapa variable menjadi variable baru yang lebih sedikit dengan tetap mempertahankan informasi dari data awal sehingga mudah untuk dianalisis (Migenda, 2021; Hamina, 2024). Langkah pertama dari *Principal Component Analysis* adalah menghitung matriks kovarians kemudian menentukan *eigen value* dan *eigen vector* (Nguyet, 2025; Dinanti, 2023).

Perhitungan matriks kovarians dilakukan dengan persamaan berikut

$$M_{ij} = \frac{1}{n-1} \sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j) \dots \dots \dots (i)$$

Keterangan

- M_{ij} : Matriks kovarians
- x_{ki} : data ke-k pada fitur ke-i
- x_{kj} : data ke-k pada fitur ke-j
- $\bar{x}_i$: rata-rata dari fitur ke-i
- $\bar{x}_j$: rata-rata dari fitur ke-j
- n : jumlah sampel

Eigen value ditentukan dengan persamaan berikut

$$\det(M - \lambda I) = 0 \dots \dots \dots (ii)$$

Keterangan

- M : Matriks kovarians
- λ : Eigen value
- I : Matriks identitas

Eigen vector ditentukan dengan persamaan berikut

$$(M - \lambda I)V = 0 \dots \dots \dots (iii)$$

Keterangan

- V : Eigen vektor

Eigen value yang digunakan dalam mereduksi dimensi data adalah yang bernilai ≥ 1 (You dan Hung, 2021; Cruz, 2025). Proses reduksi data dilakukan dengan persamaan berikut

$$x_{reduced} = \text{Eigen vektor Subset}^T \times x^T \dots\dots\dots (iv)$$

Keterangan

$x_{reduced}$: Data hasil reduksi

Eigen vektor Subset : Eleman eigen vector

x : Data asli

K-Means Clustering

K-Means Clustering digunakan untuk mengelompokkan sekumpulan objek menjadi beberapa kelompok berdasarkan kesamaan karakteristik dan objek dengan karakteristik yang berbeda dikelompokkan dalam cluster yang berbeda (Apriyanto, 2025). Langkah pertama dari algoritma K-Means Clustering yaitu menentukan jumlah kluster (k) dimana setiap kluster akan diwakili oleh sebuah centroid (Bimantara, 2023). Data dikelompokkan kedalam kluster berdasarkan centroid terdekat. Centroid akan terus diperbaharui berdasarkan rata-rata posisi data sampai posisi centroid stabil. Apabila centroid sudah stabil secara signifikan maka pembentukan kelompok dianggap sudah optimal (Chen, 2024).

Untuk mengevaluasi kualitas pembentukan kluster yang sudah optimal tersebut dilakukan dengan Silhouette Score dengan membandingkan jarak antara objek dan anggota dalam kelompok yang sama dengan jarak antara objek dan anggota kelompok lain yang terdekat (Juanita, 2024). Perhitungan nilai Silhouette Score dilakukan dengan menyelesaikan persamaan berikut

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \dots\dots\dots (v)$$

$a(i)$: rata-rata jarak dari titik i ke titik lain dalam kelompok yang sama

$b(i)$: rata-rata jarak dari titik i ke titik lain dalam kelompok terdekat lainnya

3. METODE PENELITIAN

Tahap awal dari penelitian ini adalah mengumpulkan data melalui *survey* kepada responden. Data yang diperoleh diolah dengan *Principal Component Analysis* untuk mereduksi dimensi data dan tetap mempertahankan informasi yang paling penting. Langkah selanjutnya yaitu mengelompokkan data yang telah direduksi menggunakan *K-Means Clustering* kedalam beberapa kluster dimana setiap kluster yang terbentuk terdiri dari data dengan karakteristik yang mirip. Data dengan karakteristik yang berbeda ditempatkan pada kluster yang lain. Proses ini dilakukan untuk menemukan struktur tersembunyi dari data.

Untuk mengevaluasi hasil klasterisasi tersebut dilakukan pengujian Silhouette Score. Nilai tertinggi dari hasil pengujian merupakan kualitas klasterisasi yang terbaik. Hasil dari evaluasi ini yang akan menentukan efektifitas dari klasterisasi yang dilakukan.

Indicator perilaku penggunaan internet:

X_1 : Akademik-Riset (mencari jurnal ilmiah, artikel guna mendukung tugas kuliah)

X_2 : Academic-System (membuka platform pembelajaran kampus)

X_3 : Academic-Diskusi (berdiskusi kelompok secara daring)

X_4 : Hiburan-Sosial (menggunakan medsos)

X_5 : Hiburan-Media (menonton streaming video dan mendengarkan music)

X_6 : Hiburan-Game (bermain game online)

X_7 : Durasi tinggi (akses internet lebih dari 5 jam)

X_8 : waktu kuliah (akses internet pada jam operasional kuliah)

X_9 : Begadang (akses internet hingga larut malam mengganggu waktu istirahat)

X_{10} : Mobilitas gadget (lebih sering akses internet lewat gadget)

X_{11} : Ketergantungan fasilitas (menunda unduh file atau dokumen sampai terhubung wifi kampus)

X_{12} : Kemandirian data (memastikan paket pribadi selalu aktif agar bisa akses internet dimana saja dan tidak bergantung wifi kampus)

4. HASIL DAN PEMBAHASAN

PCA

Berdasarkan hasil uji kelayakan antar variable melalui uji chi squared test diperoleh nilai signifikansi statistic model $\chi^2 = 223,61$ dengan nilai $p < 0,001$ yang menunjukkan adanya hubungan linear yang kuat antar indicator maka data dapat dianalisis menggunakan PCA. Hasil PCA diperoleh dua komponen utama yang memiliki nilai eigen value > 1 yaitu komponen 1 sebesar 5,473 dan komponen 2 sebesar 1,904. Kombinasi dari kedua komponen laten ini mampu menjelaskan 61,5% varians total dari keseluruhan indicator perilaku penggunaan internet responden. Distribusi loading factor dari 12 indikator utama menunjukkan pola pengelompokan yang sangat bersih yaitu sebagai berikut

Komponen 1 yaitu dimensi utilitas akademik dan hiburan yang terstruktur

Korelasi muatan yang dominan terdapat pada variable X_2 (Akses system akademik sebesar 0,825), X_6 (Akses game online sebesar 0,796), X_5 (Akses streaming video dan music sebesar 0,780), X_1 (Akses jurnal ilmiah dan riset tugas sebesar 0,727 dan X_4 (Akses medsos

sebesar 0,648). Komponen ini menjelaskan jenis aktivitas fungsional yang dilakukan oleh responden diruang digital.

Komponen 2 yaitu dimensi pola intensifitas dan ketergantungan fasilitas

Korelasi muatan yang dominan terdapat pada X_8 (Akses pada jam kuliah sebesar), X_9 (Akses sampai larut malam sebesar 0,796), X_7 (Akses lebih dari 5 jam sehari sebesar 0,790), X_{11} (menunda unduh dokumen sampai dapat wifi kampus sebesar 0,759), X_{10} (mobilitas akses lewat gadget sebesar 0734), X_{12} (kemandirian paket data pribadi sebesar 0,699). Komponen ini perilaku pernggunaan internet berdasarkan manajemen waktu, ketergantungan jaringan internet dan cara mahasiswa terhubung ke internet

K-Means Clustering

Jumlah parameter yang digunakan yaitu $k = 3$. Proporsi varians yang dijelaskan antar klaste R^2 pada model final yang terbentuk sebesar 0,673. Indeks koefisien Silhouette yang diperoleh sebesar 0,370. Hasil tersebut menunjukkan bahwa pemissahan kelompok berada pada kategori sedang.

Table 1. output K-Means Clustering

Statistic	Klaster 1	Klaster 2	Klaster 3
Jumlah Responden	85	88	77
Persentasi Proporsi	34,0%	35,2%	30,8%
Center PC1	+0,593	+0,383	-1,092
Center PC2	-0,868	+0,981	-0.163

Berdasarkan table tersebut skor positif menunjukkan bahwa tingkat kecenderungan perilaku penggunaan internet berada diatas rata-rata dan skor negative menunjukkan tingkat kecenderungan perilaku penggunaan internet berada dibawah rata-rata keseluruhan reponden.

Klaster 1. Pengguna internet fungsional akademik

Pada kelompok ini responden memanfaatkan internet secara aktif untuk riset akademik maupun unuk hiburan secara terstrutur dan memiliki control diri yang sangat baik. Dalam klaster ini responden tidak menunjukkan kecenderungan begadang dalam mengakses internet, tidak bergantung pada fasilitas wifi gratis di kampus. Kelompok ini menunjukkan responden yang memiliki tingkat literasi digital yang sehat dan fungsional.

Klaster 2. Pengguna internet yang konsumtif dan bergantung

Pada kelompok ini responden merupakan kelompok pengguna internet yang beresiko dimana responden menggunakan internet secara tidak terkendali untuk kegiatan perkuliahan dan untuk hiburan. Skor Center +0,981 menunjukkan penggunaan internet dengan durasi yang berlebih > 5 jam, sering sampai mengakses internet begadang serta menunjukkan ketergantungan pada fasilitas wifi kampus demi menghemat paket pribadi

Klaster 3. Pengguna internet mandiri dan hemat

Responden pada kelompok ini memiliki frekuensi penggunaan internet yang paling rendah. Pola akses internet responden cenderung menggunakan internet dengan paket pribadi. Responden tidak bergantung pada wifi kampus sehingga responden menggunakan paket data pribadi secara efisien serta minim terhadap akses internet untuk hiburan.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis perilaku penggunaan internet dengan melalui integrasi PCA dan *K-Means Clustering*, keragaman perilaku responden dikelompokkan menjadi 3 klaster. Klaster 1 yaitu responden mampu menyelaraskan tingginya utilitas internet akademik dengan manajemen waktu akses yang sehat sebanyak 34,0%. Klaster 2 yaitu responden dengan perilaku konsumsi internet yang intensif dan tidak sehat serta memiliki ketergantungan pada fasilitas penunjang wifi kampus sebanyak 35,2%. Klaster 3 yaitu responden yang tidak bergantung pada fasilitas wifi kampus dan pola penggunaan internet yang minim terhadap hiburan sebanyak 30,8%. Saran yang direkomendasikan untuk responden klaster 2 adalah universitas perlu melakukan pembatasan akses internet terhadap tautan link non akademik agar fasilitas kampus tidak terbuang untuk akses hiburan yang kotraproduktif. Untuk klaster 1 yang produktif dalam riset akademik, universitas perlu meningkatkan penyediaan akses jurnal ilmiah gratis untuk memfasilitasi kebutuhan riset mahasiswa.

DAFTAR REFERENSI

- Alpianoor, A. (2025). Analisis kecepatan internet dan produktivitas mahasiswa dalam pembelajaran daring. *JUTECH: Journal Education and Technology*, 6(1), 35–49. <https://doi.org/10.31932/jutech.v6i1.4750>
- Alvaridzi, R. A. D., & Ardelia, V. (2025). Dilema mahasiswa di era digital: Korelasi antara FOMO dan cyberslacking. *Character: Jurnal Penelitian Psikologi*, 12(1), 601–610. <https://doi.org/10.26740/cjpp.v12n01.p601-610>
- Apriyanto, B., & Sitio, S. L. M. (2025). Penerapan K-means dalam menganalisis pola pembelian pelanggan pada data transaksi e-commerce. *Bit-Tech (Binary Digital-Technology)*, 7(3), 791–798. <https://doi.org/10.32877/bt.v7i3.2195>
- Bimantara, I. M. S., & Widiarta, I. M. (2023). Optimization of K-means clustering using particle swarm optimization algorithm for grouping traveler reviews data on Tripadvisor sites. *Jurnal Ilmiah Kursor*, 12(1), 1–10. <https://doi.org/10.21107/kursor.v12i01.269>
- Chen, L., Roe, D. R., Kochert, M., Simmerling, C., & Miranda-Quintana, R. A. (2024). K-means NANI: A clustering algorithm for molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 20(13), 5583–5597. <https://doi.org/10.1021/acs.jctc.4c00308>
- Cruz Diaz, L. E., & Bautista Zuniga, L. D. L. C. (2025). Principal component analysis (PCA): A multivariate approach to macroeconomic indicators from Vizcarra to

- Sagasti during the COVID-19 pandemic in Peru. *Journal of Posthumanism*, 5(6), 4560–4573. <https://doi.org/10.63332/joph.v5i6.2661>
- Dinanti, A., & Purwadi, J. (2023). Analisis performa K-nearest neighbor dan reduksi dimensi menggunakan principal component analysis. *Jambura Journal of Mathematics*, 5(1), 16–29. <https://doi.org/10.34312/jjom.v5i1.17098>
- Febrina, S., & Sabrina, P. R. N. (2025). Cyberslacking di kalangan mahasiswa: Peran self-control dalam mengatur perilaku digital. *Jurnal Penelitian Pendidikan, Psikologi dan Kesehatan (J-P3K)*, 6(4), 1392–1401. <https://doi.org/10.51849/j-p3k.v6i4.871>
- Fuada, N., Dwatra, F. D., Susanti, R. E., & Magistarina, E. (2023). Kontribusi self-regulated learning terhadap cyberloafing mahasiswa Universitas Negeri Padang. *CAUSALITA: Journal of Psychology*, 1(3), 58–67. <https://doi.org/10.62260/causalita.v1i3.36>
- Hamina, S. (2024). Data reduction using principal component analysis: Theoretical underpinnings and practical applications in public health. *Journal of Contemporary Medical Education*, 14(1). <https://www.jcmedu.org/jcmedu-articles/data-reduction-using-principal-component-analysis-theoretical-underpinnings-and-practical-applications-in-public-health-105124.html>
- Juanita, S., & Cahyono, R. D. (2024). K-means clustering comparison of elbow and silhouette methods for medicines clustering based on user reviews. *Jurnal Teknik Informatika (JUTIF)*, 5(1), 283–289. <https://doi.org/10.52436/1.jutif.2024.5.1.1349>
- Listya, D., Fitria, D., Triyana, E., & Asni, N. (2023). Pemanfaatan wireless (Wi-Fi) di perguruan tinggi sebagai fasilitas kegiatan akademik mahasiswa. *BIODIK: Jurnal Ilmiah Pendidikan Biologi*, 9(1), 31–39. <https://doi.org/10.22437/bio.v9i1.19534>
- Maharani, D., Helmiah, F., & Rahmadani, N. (2021). Penyuluhan manfaat menggunakan internet dan website pada masa pandemi COVID-19. *Abdiformatika: Jurnal Pengabdian Masyarakat Informatika*, 1(1), 1–7. <https://doi.org/10.25008/abdiformatika.v1i1.130>
- Marciando, P. V. G., Ayuningtias, A. U. H., & Retnoningtias, D. W. (2022). Hubungan internet addiction terhadap kualitas tidur pada mahasiswa. *Jurnal Psimawa*, 5(2). <https://doi.org/10.36761/jp.v5i2.2110>
- Migenda, N., Moller, R., & Schenk, W. (2021). Adaptive dimensionality reduction for neural network-based online principal component analysis. *PLoS ONE*, 16(3), 1–32. <https://doi.org/10.1371/journal.pone.0248896>
- Nguyet, N. T. M., Huong, V. T. M., Vinh, B. Q., & Long, P. N. (2025). Cluster-informed policy design for school administrators in Vietnam: A data-driven perspective on digital readiness. *Educational Process: International Journal*, 16(3), 1–23. <https://doi.org/10.22521/edupij.2025.16.291>
- Putri, N. D. N. (2025). Pengguna internet di Indonesia: Tren, tantangan, dan peluang di era digital. *Maliki Interdisciplinary Journal (MIJ)*, 3(8), 604–608. <http://urj.uin-malang.ac.id/index.php/mij/index>
- Susanti, S., Manalu, A. H., Pakpahan, N., Pinem, S. A., & Pinem, Y. V. B. (2025). Analisis tingkat konsentrasi belajar siswa akibat penggunaan gadget di SMA Methodist 7 Medan, Jl. Madong Lubis No. 7, Medan. *JIIIC: Jurnal Intelek Insan Cendekia*, 2(6), 11658–11665. <https://jicnusanantara.com/index.php/jiic/article/view/3802>
- Syukri, A. U., & Megawaty. (2025). Dampak dunia digital terhadap kesejahteraan masyarakat di Indonesia. *Movere Journal*, 7(2), 305–319. <https://doi.org/10.53654/mv.v7i2>
- You, S. D., & Hung, M. J. (2021). Comparative study of dimensionality reduction techniques for spectral-temporal data. *Information*, 12(1), 1–12. <https://doi.org/10.3390/info12010001>