



Penerapan Metode PCA dan Apriori dalam Deteksi Pola Penggunaan Internet Dikalangan Mahasiswa

Aero Dhimas Syarif^{1*}, Hanif Rabbani², Armanto Hakim³

¹⁻³Teknologi Informasi, Universitas Bina Sarana Informatika, Indonesia

*Penulis Korespondensi: aerowolfzzz@gmail.com

Abstract. *University students generate extensive digital footprints that form complex, high-dimensional datasets reflecting diverse patterns of online behavior. Understanding these patterns requires analytical methods capable of handling large and interrelated variables. This study aims to map internet usage trends among university students using an integrated approach that combines Principal Component Analysis (PCA) and the Apriori Algorithm. PCA is employed to reduce data dimensionality by identifying the most influential components and eliminating redundant information, thereby simplifying the dataset without losing essential characteristics. Subsequently, the Apriori Algorithm is applied to uncover association rules that describe relationships between different types of digital activities. Data were collected through structured questionnaires distributed to active university students, capturing various aspects of their internet usage behavior. Through this combined methodology, the study seeks to identify the main factors that shape students' digital habits and to reveal hidden behavioral patterns that may not be evident through conventional analysis. The expected results will provide a clearer understanding of how students interact with digital platforms and online resources. Ultimately, the findings are intended to serve as an empirical basis for designing more effective, data-driven digital literacy programs and strategies in higher education environments.*

Keywords: *Apriori; Association Rules; Data Mining; Internet Behavior; PCA.*

Abstrak. Mahasiswa universitas menghasilkan jejak digital yang luas yang membentuk kumpulan data kompleks dan berdimensi tinggi yang mencerminkan beragam pola perilaku daring. Memahami pola-pola ini membutuhkan metode analitik yang mampu menangani variabel-variabel besar dan saling terkait. Studi ini bertujuan untuk memetakan tren penggunaan internet di kalangan mahasiswa universitas menggunakan pendekatan terintegrasi yang menggabungkan Analisis Komponen Utama (PCA) dan Algoritma Apriori. PCA digunakan untuk mengurangi dimensi data dengan mengidentifikasi komponen yang paling berpengaruh dan menghilangkan informasi yang berlebihan, sehingga menyederhanakan kumpulan data tanpa kehilangan karakteristik penting. Selanjutnya, Algoritma Apriori diterapkan untuk mengungkap aturan asosiasi yang menggambarkan hubungan antara berbagai jenis aktivitas digital. Data dikumpulkan melalui kuesioner terstruktur yang dibagikan kepada mahasiswa universitas yang aktif, yang menangkap berbagai aspek perilaku penggunaan internet mereka. Melalui metodologi gabungan ini, studi ini berupaya mengidentifikasi faktor-faktor utama yang membentuk kebiasaan digital mahasiswa dan mengungkap pola perilaku tersembunyi yang mungkin tidak terlihat melalui analisis konvensional. Hasil yang diharapkan akan memberikan pemahaman yang lebih jelas tentang bagaimana mahasiswa berinteraksi dengan platform digital dan sumber daya daring. Pada akhirnya, temuan ini dimaksudkan untuk menjadi dasar empiris dalam merancang program dan strategi literasi digital yang lebih efektif dan berbasis data di lingkungan pendidikan tinggi.

Kata kunci: Apriori; Aturan Asosiasi; Data Mining; PCA; Perilaku Internet.

1. LATAR BELAKANG

Akselerasi inovasi teknologi memposisikan koneksi digital sebagai instrumen vital bagi sivitas akademika dalam menunjang aspek edukasi maupun personal. (Sangotra et al., 2026). Beragam studi sebelumnya menggarisbawahi kontribusi signifikan internet dalam memperlancar transfer ilmu. (Lukianova et al., 2024), interaksi pendidikan, serta perolehan data saintifik secara instan. Kendati demikian, pemanfaatan yang masif turut membawa dampak buruk, mulai dari pecahnya konsentrasi saat belajar hingga kecanduan gawai dan degradasi kinerja akademik. (Gao, 2021). Masalah-masalah seperti kelelahan mental akibat

layar (*screen fatigue*) dan kebiasaan berpindah-pindah tugas secara ekstrem menjadi tantangan nyata di era saat ini. (Kumar et al., 2022).

Rekaman aktivitas daring mahasiswa biasanya melibatkan sekumpulan parameter yang luas, mulai dari rentang waktu selancar, kategori perangkat lunak yang digunakan, hingga alasan di balik pemilihan platform tertentu. Padatnya informasi ini sering kali memicu fenomena "kutukan dimensi" (*curse of dimensionality*), di mana keberlimpahan variabel justru menghambat ketajaman analisis karena adanya data yang saling tumpang tindih atau redundan. (Alias et al., 2024).

Guna memitigasi persoalan tersebut, skema *Principal Component Analysis* (PCA) diterapkan sebagai solusi penyederhanaan data tanpa mengabaikan inti dari informasi asli. (Ramasubramanian et al., 2024). Pendekatan ini dinilai mumpuni dalam membedah susunan tersembunyi dari tindakan pengguna serta memetakan faktor-faktor yang paling berpengaruh. (G. Martín et al., 2021). Pasca tahap penyusutan dimensi, identifikasi keterkaitan antar-perilaku dilakukan lewat mekanisme Algoritma Apriori dalam kerangka *Association Rule Mining* (ARM). (Babitha et al., 2024). Teknik ini bekerja dengan mencari himpunan data yang sering muncul guna menyusun pola relasi berdasarkan parameter nilai *support*, *confidence*, dan *lift*. (Kulkarni et al., 2022).

Studi ini berfokus pada upaya merampingkan variabel perilaku digital melalui PCA, membedah korelasi antar-aktivitas dengan teknik Apriori. serta memetakan tren utama dalam kebiasaan berinternet mahasiswa. (Meena et al., 2023). Melalui pendekatan hibrida ini, gambaran utuh mengenai karakteristik penggunaan teknologi oleh kelompok pelajar dapat dipahami secara lebih mendalam dan akurat. (Zhou & Deng, 2024).

2. KAJIAN TEORITIS

Pendekatan *Principal Component Analysis* (PCA) memiliki kapabilitas yang mumpuni dalam menyingkap kerangka tersembunyi dari kebiasaan pelajar pada sekumpulan data dengan variabel yang massif. (Nurputra et al., 2025). Secara teknis, apabila terdapat sebuah matriks informasi X berukuran $n \times p$ yang telah melalui proses normalisasi (titik tengah), maka kalkulasi terhadap matriks kovariansi (S) dilakukan guna memetakan interaksi antar-variabel. (Taleb et al., 2025). Tahapan krusial berikutnya adalah mengekstraksi nilai eigen (λ) serta vektor eigen (v) melalui penyelesaian formulasi karakteristik: (Liu, 2022)

$$Sv = \lambda v$$

Sementara itu, mekanisme Algoritma Apriori diimplementasikan untuk menggali tren serta keterkaitan dalam aktivitas ruang siber mahasiswa. Validitas dan kekuatan dari pola hubungan yang ditemukan diukur menggunakan tiga indikator fundamental: *Support*, *Confidence*, dan *Lift Ratio*. Keberadaan skor *Lift* yang melampaui angka satu menjadi parameter kunci bahwa korelasi antar-perilaku yang terbentuk bersifat signifikan, valid, dan bukan merupakan kejadian kebetulan semata. (Saputra et al., 2025)

3. METODE PENELITIAN

Desain Penelitian

Studi ini mengadopsi kerangka kuantitatif deskriptif yang berorientasi pada teknik penggalian data (*data mining*). Alur kerja dilakukan secara sistematis, dimulai dari fase akuisisi informasi, pembersihan data, penyusutan dimensi melalui PCA, hingga tahap akhir berupa identifikasi relasi antar-elemen menggunakan metode Apriori. Strategi analitis ini dipilih karena keandalannya dalam mendeteksi pola-pola laten yang terkubur di dalam kumpulan data dengan parameter yang sangat luas. (Barun & Önder, n.d.).

Teknik Pengumpulan

Data Data diperoleh melalui instrumen kuesioner digital (*Google Form*) yang ditargetkan bagi mahasiswa yang masih menempuh masa studi. Berbagai indikator yang diobservasi meliputi durasi berselancar harian, preferensi platform dan perangkat lunak, fokus kegiatan utama saat daring, periode waktu akses, serta penilaian subjektif mengenai efisiensi kerja digital. Guna menjamin hasil komponen yang konsisten dan tepercaya pada tahap reduksi, kuota sampel ditetapkan sekurang-kurangnya 100 responden. (Limbong et al., 2025).

Tahapan Pengolahan Awal Prosedur prapemrosesan data mencakup empat tindakan krusial:

- 1) Pembersihan – mengeleminasi data ganda, memulihkan informasi yang kosong, serta mereduksi gangguan (*noise*).
- 2) Transformasi Kode – mengubah data kualitatif menjadi representasi numerik melalui teknik *label encoding* atau penggunaan skala Likert.
- 3) Standardisasi – menyelaraskan rentang variabel agar berada pada skala yang seimbang dan dapat diperbandingkan.
- 4) Diskretisasi – mengategorikan keluaran komponen PCA ke dalam klasifikasi tingkat rendah, moderat, dan tinggi.

Penyusutan Variabel via PCA Implementasi PCA dijalankan dengan menyusun matriks korelasi, menghitung nilai eigen serta vektor eigen, dan menyeleksi komponen utama berdasarkan kriteria Kaiser (nilai eigen > 1). Inti dari tahap ini adalah membedah dan memahami komponen yang memberikan kontribusi variansi paling signifikan terhadap keseluruhan data. (Yata & Aoshima, 2022).

Eksplorasi Aturan Asosiasi (Apriori) Mekanisme algoritma Apriori dilakukan dengan menetapkan ambang batas dukungan (*minimum support*) dan tingkat kepercayaan (*minimum confidence*) tertentu. (Hikmawati et al., 2021). Melalui ekstraksi kumpulan data yang sering muncul (*frequent itemset*), logika keterkaitan disusun dalam format JIKA–MAKA (premis-konsekuensi) dengan merujuk pada skor rasio daya angkat (*lift ratio*) sebagai ukuran validitasnya. (Ahmad et al., 2021).

4. HASIL DAN PEMBAHASAN

Profil dan Statistik Deskriptif Responden \

Melalui keterlibatan seratus subjek dari kalangan mahasiswa, studi ini berhasil memetakan pola pemanfaatan teknologi informasi secara mendetail. Mayoritas partisipan, yakni delapan dari sepuluh orang (80%), lebih mengandalkan ponsel pintar sebagai instrumen utama untuk terhubung ke jagat maya, dengan dukungan konektivitas yang didominasi oleh fasilitas Wi-Fi di area tempat tinggal (84%). Fenomena durasi penggunaan yang ekstrem terlihat dari fakta bahwa lebih dari separuh responden (51%) berselancar lebih dari tujuh jam setiap harinya, di mana puncak kepadatan trafik ditemukan terjadi pada waktu larut malam (77%). Di sisi lain, konsumsi konten media sosial menjadi prioritas kegiatan utama (52%), mengungguli sektor permainan daring (22%). Di antara berbagai platform yang tersedia, TikTok mendominasi preferensi pengguna (41%), yang kemudian disusul oleh Instagram (32%) serta WhatsApp (20%).

Tabel 1. Distribusi Profil dan Pola Penggunaan Responden.

Kategori	Item Dominan	Persentase (%)
Perangkat Akses	Smartphone	80%
Jenis Koneksi	Wi-Fi Rumah	84%
Durasi Penggunaan	>7 Jam per hari	51%
Waktu Akses Tertinggi	Malam Hari	77%
Aktivitas Utama	Media Sosial	52%
Platform Favorit	Tiktok	41%

Analisis Reduksi Dimensi (PCA)

Berdasarkan pembedahan terhadap keterkaitan antar-variabel dalam instrumen penelitian, teknik *Principal Component Analysis* (PCA) sukses merampingkan kerumitan data menjadi dua pilar perilaku yang dominan. Pilar pertama, yang diidentifikasi sebagai dimensi Gaya Hidup dan Rekreasi, mencakup elemen akses media sosial yang intensif, durasi penggunaan yang sangat panjang, serta kebiasaan mengakses konten pada malam hari untuk mengusir kejenuhan. Sementara itu, pilar kedua merepresentasikan dimensi Akademik dan Multi-tugas, di mana aktivitas pendidikan sering kali berjalan secara simultan dengan penggunaan platform sosial. Hasil analisis juga mengindikasikan adanya persepsi kuat dari para responden mengenai fenomena "kehilangan waktu" yang timbul akibat pola interaksi digital tersebut.

Tabel 2. Matriks Komponen Rotasi (Hasil PCA).

Variabel Kuesionner	(Atribut Sosial)	Komponen 1 (Lifestyle & Hiburan)	Komponen 2 (Akademik & Multitasking)
Akses Media Sosial (Tiktok/IG)		0.892	0.120
Durasi Akses >7 Jam		0.845	0.150
Waktu Akses Malam Hari		0.780	-0.210
Aktivitas Saat Bosan (Scroll Sosmed)		0.755	0.310
Multitasking (Belajar + Sosmed)		0.350	0.815
Frekuensi Akses Akademik		-0.110	0.760
Persepsi Kehilangan Waktu		0.640	0.520

Analisis Pola Asosiasi (Algoritma Apriori)

Investigasi lebih dalam menggunakan metode Apriori menyingkap rangkaian pola relasi yang signifikan dalam rutinitas harian mahasiswa. Ditemukan keterkaitan yang sangat erat antara kondisi bangun tidur dengan kecenderungan subjek untuk segera membuka aplikasi WhatsApp (R1). Selain itu, terdeteksi tren "akademik multitasking" di mana kegiatan belajar memiliki probabilitas tinggi untuk dibarengi dengan pembukaan media sosial secara bersamaan (R2). Perasaan jenuh juga teridentifikasi sebagai pemicu utama bagi mayoritas responden untuk melakukan aktivitas *scrolling* di ruang siber (R3). Terakhir, durasi sesi media sosial yang melampaui batas 30 menit terbukti memiliki korelasi linear dengan gangguan terhadap produktivitas, yang menegaskan risiko dari penggunaan platform sosial yang tidak terkendali (R4).

Tabel 3. Aturan Asosiasi (Association Rules) dari Algoritma Apriori.

ID	Antecedent (Jika...)		Consequent (Maka...)		Support	Confidence	Keterangan Data
R1	Kondisi Bangun Tidur	=	Buka Aplikasi WhatsApp	=	60%	85%	60% responden membuka WA saat bangun
R2	Aktivitas Belajar	=	Aktivitas Sosmed	=	Buka 41%	75%	Kombinasi aktivitas tertinggi
R3	Perasaan Bosan	=	Aktivitas Media Sosial	=	Scroll 62%	90%	62% memilih scroll sosmed saat bosan
R4	Durasi Sosmed = >30 Menit		Perasaan Produktivitas Terganggu	=	41%	65%	41% durasi sesi sangat lama

Pembahasan

Poin paling esensial yang muncul dari observasi ini adalah perkembangan fenomena pengerjaan tugas secara simultan atau "Academic Multitasking" di kalangan pelajar. Walaupun secara statistik porsi kegiatan utama didominasi oleh jejaring sosial sebesar 52%, para partisipan riset cenderung memiliki persepsi bahwa durasi pemakaian internet mereka tetap berada dalam ambang batas yang wajar. Namun, hasil identifikasi pola asosiasi dengan nilai dukungan 41% menyingkap bahwa pemisah antara kebutuhan studi dan hiburan menjadi sangat tidak jelas. Para mahasiswa sering kali kesulitan mempertahankan fokus utuh saat menempuh pembelajaran daring karena adanya distraksi dari platform sosial yang diakses secara bersamaan. Kondisi tersebut diperburuk oleh fakta bahwa 46% subjek penelitian sesekali merasa kehilangan kendali atas waktu mereka saat berinternet. Selain itu, tingginya minat terhadap TikTok (41%) dan kebiasaan aktif di malam hari (77%) memiliki kaitan erat dengan pemakaian internet intensif yang melewati durasi 7 jam, yang secara langsung mengancam stabilitas serta kualitas waktu istirahat mereka.

5. KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini mengonfirmasi bahwa sinergi antara teknik Principal Component Analysis (PCA) dan Algoritma Apriori sangat andal dalam membedah seluk-beluk perilaku daring di kalangan pelajar perguruan tinggi. Melalui mekanisme penyederhanaan variabel, PCA sukses mentransformasi kompleksitas data digital menjadi dua poros utama, yaitu kecenderungan gaya hidup yang berorientasi pada hiburan serta praktik belajar yang dilakukan dengan pola *multitasking*. Lebih jauh lagi, identifikasi melalui aturan asosiasi menyingkap

keterkaitan khusus yang menunjukkan bahwa aktivitas edukasi di ruang siber sering kali berlangsung simultan dengan penggunaan jejaring sosial, yang pada akhirnya memicu terbentuknya sebuah siklus ketergantungan teknologi yang bersifat sistemis.

Saran

Sebagai landasan untuk menyempurnakan eksplorasi ilmiah di masa yang akan datang, para ilmuwan selanjutnya sangat didorong untuk memperluas instrumen kuesioner dengan mengintegrasikan spektrum variabel psikologis yang lebih variatif. Selain itu, disarankan pula untuk melakukan pengujian perbandingan terhadap metodologi lain, seperti algoritma FP-Growth, dengan tujuan membedah efisiensi durasi eksekusi data agar diperoleh pemahaman yang lebih tajam mengenai kinerja komputasi dalam analisis data tersebut.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Allah SWT, dosen pembimbing, serta seluruh pengajar Program Studi Teknologi Informasi, Bina Sarana Informatika. Penulis juga berterima kasih kepada rekan-rekan mahasiswa yang turut memberikan masukan serta bantuan dalam proses pengumpulan dan pengolahan data.

DAFTAR REFERENSI

- Ahmad, H. I., Sim, A. T. H., Shah, S. M. A., Abrar, M., & Gul, A. (2021). Mining predicate rules without minimum support threshold. *Kuwait Journal of Science*, 48(4). <https://doi.org/10.48129/kjs.v48i4.9782>
- Alias, M. A. H., Aziz, M. A. A., Hambali, N., & Taib, M. N. (2024). Student performance classification: A comparison of feature selection methods based on online learning activities. *International Journal of Electrical and Computer Engineering*, 14(4), 4675–4685. <https://doi.org/10.11591/ijece.v14i4.pp4675-4685>
- Babitha, B. S., Behera, S., & Sharma, M. K. (2024). Applying association rule mining to understand customer buying behaviour. In *Proceedings of the International Conference on Computing, Communication and Networking Technologies*. <https://doi.org/10.1109/icccnt61001.2024.10725121>
- Barun, M. N., & Önder, E. (2024). Unlocking the multidisciplinary potential of data science: Insights from Apriori analysis. *Politeknik Dergisi*. <https://doi.org/10.2339/politeknik.1432158>
- Gao, T. (2021). How social media “reverse domesticate” and “make the best use of everything” from students’ academic achievement. In *E3S Web of Conferences*. <https://doi.org/10.1051/e3sconf/202125303084>
- Hikmawati, E., Maulidevi, N. U., & Surendro, K. (2021). Minimum threshold determination method based on dataset characteristics in association rule mining. *Journal of Big Data*, 8, Article 138. <https://doi.org/10.1186/s40537-021-00538-3>
- Kumar, N., Jain, S., & Bhargava, L. (2023). To study the impact of screen time on IT job professionals in India. *Journal of Community Health Management*. <https://doi.org/10.18231/j.jchm.2023.024>

- Limbong, L., Suhud, U., & Wibowo, A. U. A. (2025). Analysis of factors affecting internet consumption among upper secondary students. In *Proceedings of the International Student Conference on Business, Education, Economics, Accounting, and Management*. <https://doi.org/10.21009/isc-beam.013.133>
- Liu, J. (2022). A direct method to calculate characteristic forms. *Partial Differential Equations in Applied Mathematics*, 6, Article 100308. <https://doi.org/10.1016/j.padiff.2022.100308>
- Lukianova, L., Symela, K., & Ovcharuk, O. V. (2024). Exploring the potential of online scientific research to meet modern needs: International perspective. *Naukovi Zapysky Maloi Akademii Nauk Ukrainy*. <https://doi.org/10.51707/2618-0529-2024-30-13>
- Market basket analysis (association rule mining). (2022). In *Pro Apache Spark*. https://doi.org/10.1007/978-1-4842-8954-9_2
- Martín, A., Fernández-Isabel, A., Martín de Diego, I., & Beltrán, M. (2021). A survey for user behavior analysis based on machine learning techniques: Current models and applications. *Applied Intelligence*, 51, 640–665. <https://doi.org/10.1007/s10489-020-02160-x>
- Meena, R. E., Kavitha, T., S., A., Mathew, D. M., & Anusuya, R. (2023, January 5). Extracting behavioral characteristics of college students using data mining on big data. In *Proceedings of the International Conference on Electronics and Communication Systems*. <https://doi.org/10.1109/ICECONF57129.2023.10084276>
- Nurputra, W. R., Witanti, W., & Komarudin, A. (2025). Principal component analysis (PCA) untuk meningkatkan hasil klasterisasi penjualan video game menggunakan algoritma K-means. *Jurnal Locus Penelitian dan Pengabdian*, 4(8). <https://doi.org/10.58344/locus.v4i8.4151>
- Ramasubramanian, S., C. N., S., Athreya, A. J., Devarajan, A., Shankar, A., & Kumar, R. P. (2024). Data dimensionality reduction using principal component analysis: A case study. In *Proceedings of the International Conference on Communication, Computing and Signal Processing*. <https://doi.org/10.1109/incccs60947.2024.10593421>
- Sangotra, D. I., Vasundhara, S., Kaur, R., Srivastava, A., & Sambathkumar, M. (2025). Digital innovation in higher educational system. In *Advances in Computational Intelligence and Robotics* (Book series). <https://doi.org/10.4018/979-8-3373-0035-1.ch011>
- Saputra, H., Uneputty, R. A., Simyapen, L. A., Mustamir, M. F. B., Ikawanti, F. A., & Kusumawati, S. P. (2025). Analisis pola kegiatan belajar mahasiswa terhadap keberhasilan akademik menggunakan algoritma Apriori. *Deleted Journal*. <https://doi.org/10.62357/jsit.v4i2.567>
- Taleb, N. N., Zalloua, P., Elbassioni, K., Henschel, A., & Platt, D. E. (2023). Informational rescaling of PCA maps with application to genetic distance. *arXiv*. <https://doi.org/10.48550/arXiv.2303.12654>
- Yata, K., & Aoshima, M. (2022). Automatic sparse PCA for high-dimensional data. *Statistica Sinica*, 32(4). <https://doi.org/10.5705/ss.202022.0319>
- Zhou, Y., & Deng, L. (2024). Lost in multitasking: An exploration of Chinese university students' in-class smartphone multitasking patterns using multiple approaches. *Australasian Journal of Educational Technology*. <https://doi.org/10.14742/ajet.8941>