

Analisa Klaster Pola Karakteristik Pengunjung *Event* di DKI Jakarta Menggunakan Metode *K-Means*

Nadia Umaira¹, Witri Astuti², Regina Ayu Puspita Sari³, Sheili Rika Karlina⁴, Ika Selviana Putri⁵
Universitas Bina Sarana Informatika, Jakarta, Indonesia
19231806@bsi.ac.id¹, 19232090@bsi.ac.id², 19231051@bsi.ac.id³, 19230132@bsi.ac.id⁴,
19230989@bsi.ac.id⁵

Abstrak— Kota Jakarta sebagai pusat kegiatan ekonomi, budaya, dan pariwisata nasional menyelenggarakan berbagai jenis *event* yang berpotensi memengaruhi tingkat kunjungan masyarakat. Penelitian ini bertujuan untuk mengelompokkan 133 *event* pariwisata dan ekonomi kreatif di DKI Jakarta tahun 2024–2025 guna mengidentifikasi pola karakteristik pengunjung. Proses analisis mengikuti tahapan *Knowledge Discovery in Databases (KDD)* dan menghasilkan tiga kluster utama dengan pola yang berbeda. Hasil evaluasi menggunakan *Silhouette Score* menunjukkan nilai rata-rata 0,65 yang menandakan kualitas pengelompokan pada kategori cukup baik. Temuan penelitian mengungkap adanya perbedaan yang jelas antara *event* berskala kecil dan besar serta menunjukkan dominasi *event* dengan jumlah pengunjung yang relatif rendah. Hasil ini dapat menjadi dasar dalam merancang strategi pengembangan *event* yang lebih seimbang dan mendukung peningkatan daya tarik wisata daerah.

Abstract— Jakarta, as Indonesia's economic, cultural, and tourism hub, hosts a wide range of events that influence visitor dynamics throughout the year. This study aims to cluster 133 tourism and creative economy events held in DKI Jakarta during 2024–2025 to identify patterns in visitor characteristics. The analysis follows the *Knowledge Discovery in Databases (KDD)* process and produces three main clusters with distinct patterns. Evaluation using the *Silhouette Score* yields an average value of 0.65, indicating a moderately good clustering quality. The findings reveal a clear distinction between small-scale and large-scale events and show the dominance of events with relatively low visitor numbers. These insights can support the development of more balanced event strategies to enhance the region's tourism appeal.

Keywords: *Klustering, K-Means, tourism events, DKI Jakarta*

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-nc/4.0/) license.



1. Pendahuluan

Sebagai ibu kota negara dan pusat kegiatan ekonomi serta budaya, DKI Jakarta memiliki jumlah penyelenggaraan *event* yang sangat tinggi setiap tahunnya. Kegiatan *event* di wilayah DKI Jakarta merupakan salah satu bentuk aktivitas sosial yang menggambarkan minat masyarakat terhadap berbagai bidang, seperti budaya, ekonomi kreatif, pariwisata, dan hiburan. Meskipun demikian, setiap *event* tidak selalu memiliki tingkat kunjungan yang sama. Faktor seperti tema acara, jadwal pelaksanaan, dan jenis kegiatan dapat memengaruhi jumlah pengunjung. Kondisi ini menimbulkan kebutuhan untuk memahami pola partisipasi masyarakat secara lebih terukur, terutama karena tingginya intensitas penyelenggaraan *event* membuat analisis manual menjadi tidak efektif. Oleh karena itu, diperlukan pendekatan berbasis data untuk mengidentifikasi pola kunjungan dan karakteristik *event* di DKI Jakarta.

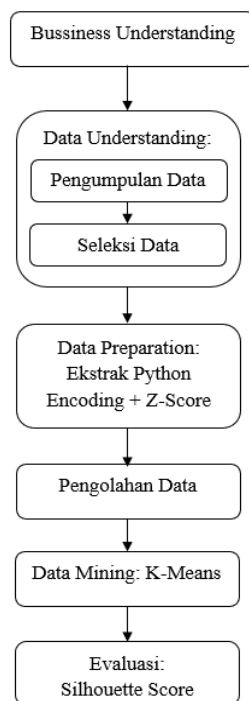
Event merupakan suatu kegiatan yang diselenggarakan untuk memperingati hal-hal penting, baik secara individu atau kelompok yang terkait secara adat, budaya, tradisi, atau agama untuk tujuan tertentu serta melibatkan lingkungan masyarakat pada waktu tertentu [1]. Pendapat lain mengatakan *event* adalah kegiatan khusus yang diselenggarakan di dalam negeri maupun di tingkat nasional dengan tujuan tertentu dan melibatkan banyak orang [2]. Tak jarang pula masyarakat menjadikan *event* sebagai objek pariwisata sebagai hiburan. Pariwisata merupakan kegiatan perjalanan wisatawan mengunjungi tempat lain yang bersifat sementara tidak menetap [3]. Pemerintah provinsi bersama pelaku industri kreatif secara rutin mengadakan berbagai *event* untuk meningkatkan partisipasi publik dan memperkuat citra kota sebagai pusat kegiatan nasional. Namun, tidak semua *event* memperoleh tingkat kunjungan yang sama. Beberapa jenis *event* bisa jadi lebih ramai dikunjungi dibanding lainnya, tergantung pada waktu, lokasi, maupun tema acara. Tingginya frekuensi penyelenggaraan *event* juga menimbulkan tantangan tersendiri dalam memahami pola kunjungan masyarakat. Oleh karena itu, dibutuhkan analisis berbasis data untuk memahami tren partisipasi masyarakat terhadap *event* yang diselenggarakan di wilayah DKI Jakarta.

Berbagai penelitian sebelumnya telah menunjukkan bahwa algoritma *K-Means* efektif digunakan untuk pengelompokan data pada domain sosial maupun pariwisata. Penelitian di Kabupaten Cirebon [4] menggunakan *K-Means* untuk mengelompokkan kecamatan berdasarkan produktivitas padi, sedangkan beberapa penelitian lain memanfaatkan algoritma yang sama untuk menganalisis karakteristik pengunjung mall dalam konteks pemasaran [5][6]. Dalam bidang pariwisata, penelitian mengenai pengelompokan tempat wisata di Jakarta dan Yogyakarta [7][8] juga menunjukkan bahwa *K-Means* mampu memberikan segmentasi yang informatif bagi pengembangan destinasi wisata. Selain itu, penelitian *hybrid K-Means* dan *Decision Tree* [9] membuktikan bahwa pendekatan klustering dapat digunakan untuk mendukung kebijakan publik berbasis data.

Dengan demikian, dari berbagai penelitian tersebut menunjukkan bahwa belum ada penelitian yang secara khusus membahas tentang pengelompokan pola karakteristik pengunjung *event* di DKI Jakarta dengan menggunakan algoritma *K-Means Klustering*. Hasil pengelompokan ini diharapkan dapat memberikan gambaran mengenai tren minat masyarakat terhadap berbagai jenis *event* di DKI Jakarta. Selain itu, penelitian ini juga dapat membantu pemerintah daerah dan pihak penyelenggara dalam mengambil keputusan strategis yang lebih tepat, misalnya dalam menentukan prioritas *event* yang berpotensi menarik lebih banyak pengunjung. Dengan demikian, penelitian ini tidak hanya memberikan nilai akademis dalam menerapkan metode data mining, tetapi juga memiliki nilai praktis dalam mendukung kebijakan berbasis data untuk pengembangan kegiatan publik DKI Jakarta. Pendekatan ini krusial mengingat data acara dari Portal Satu Data Jakarta mencakup 133 entri dengan fitur numerik dan kategorikal campuran, yang memerlukan praproses intensif seperti pengkodean *one-hot* dan normalisasi *Z-Score* untuk menghindari bias pengelompokan. Penelitian ini dapat membuka peluang integrasi dengan teknik pembelajaran mesin canggih, seperti prediksi kehadiran pengunjung, untuk memperkuat ekosistem pariwisata berkelanjutan di ibu kota.

2. Metodologi Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining berbasis proses *Knowledge Discovery in Database* (KDD) yang meliputi tahap seleksi data, *preprocessing*, transformasi data, *data mining*, dan evaluasi hasil [4]. Tujuan utama penelitian adalah mengidentifikasi pola karakteristik pengunjung *event* di DKI Jakarta untuk mendukung proses perencanaan program oleh Dinas Pariwisata dan Ekonomi Kreatif DKI Jakarta sebagai pemangku kepentingan. Analisis difokuskan pada tiga atribut utama, yaitu jumlah pengunjung, jenis *event*, dan periode hari, sehingga diperoleh representasi pola partisipasi masyarakat terhadap *event* sepanjang tahun.



Gambar 1. Tahapan *Knowledge Discovery in Database* (KDD)

Business understanding tahap ini dilakukan untuk memahami tujuan analisis, kebutuhan informasi, serta konteks permasalahan yang ingin diselesaikan agar arah pengolahan data sesuai dengan tujuan penelitian. Pemerintah Provinsi DKI Jakarta secara rutin menyelenggarakan *event* budaya, ekonomi kreatif, dan pariwisata, tetapi tidak semua *event* mendapatkan tingkat kunjungan yang setara. Masalah utama yang ingin diselesaikan adalah bagaimana mengelompokkan *event* berdasarkan pola pengunjung untuk mendukung perencanaan *event* yang lebih strategis, seperti penentuan prioritas promosi, pemilihan kategori *event*, dan evaluasi efektivitas program. Hasil klastering diharapkan memberikan gambaran empiris yang dapat dimanfaatkan oleh pembuat kebijakan.

Data understanding, data diperoleh dari Portal Satu Data Jakarta dan berisi 133 *event* yang diselenggarakan pada tahun 2024 hingga pertengahan 2025. Dataset mencakup atribut periode_data, semester, nama_event, penyelenggara, tanggal, lokasi, jenis_event, dan jumlah_pengunjung. Tiga atribut dipilih sebagai fitur utama yaitu jumlah_pengunjung, periode_hari, dan jenis_event. Sedangkan atribut lain digunakan sebagai metavariabel.

Data preparation, tahap ini mencakup proses pembersihan, transformasi, dan standarisasi data agar sesuai dengan kebutuhan algoritma *K-Means*. Karena *K-Means* tidak dapat langsung memproses data kategorikal [10], maka dilakukan tahap ekstraksi periode_hari dari kolom tanggal untuk mendapatkan durasi penyelenggaraan *event*. Konversi jenis_event menjadi variabel kategorikal yang nantinya diubah menjadi numerik menggunakan *one-hot encoding*. *One-hot encoding* diterapkan untuk menghasilkan representasi biner dari jenis_event. Normalisasi *Z-Score* dilakukan pada seluruh fitur numerik agar skala data seimbang dan tidak ada atribut yang mendominasi proses pengelompokan. Pemeriksaan missing value dan konsistensi tipe data dilakukan pada tahap akhir untuk memastikan data valid sebelum dianalisis. Tahap ini memberikan dataset final yang sepenuhnya numerik dan siap digunakan oleh algoritma *K-Means*.

Data mining, dalam konteks *data mining*, proses penemuan pola informasi yang bermanfaat dari data besar dilakukan secara semi-otomatis dengan memanfaatkan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* [11]. Untuk melakukan analisis tersebut, digunakan pendekatan data mining dengan metode *K-Means* Klastering. *K-Means* salah satu algoritma pengelompokan data (*klaster*) non hirarki yang mempartisi data yang ada ke dalam bentuk satu atau lebih *klaster*/kelompok [12].

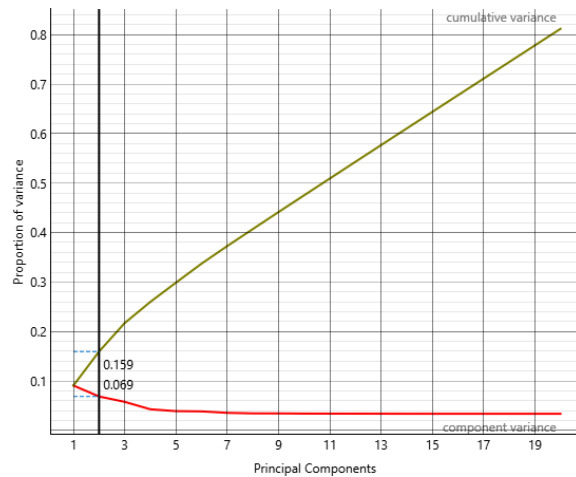
Proses pengelompokan dilakukan menggunakan algoritma *K-Means* karena sifatnya yang sederhana, efisien, dan cocok untuk data berukuran kecil hingga menengah. Penentuan jumlah klaster (*K*) dilakukan dengan mempertimbangkan konteks domain (*event* besar, *event* menengah, dan *event* kecil). Menganalisis *Silhouette Score* pada beberapa nilai *K*. Mengevaluasi kestabilan hasil setelah transformasi data. Berdasarkan analisis tersebut, nilai *K*=3 dipilih sebagai konfigurasi yang paling sesuai secara empiris dan konseptual.

Evaluation, untuk menilai kualitas hasil klaster, digunakan *Silhouette Score* sebagai metode evaluasi internal. Matriks ini dipilih karena mampu mengukur kohesi antar data dalam klaster yang sama, separasi antar klaster, dan stabilitas pembentukan klaster. Skor rata-rata *Silhouette* digunakan untuk menilai kualitas keseluruhan model, sedangkan skor per-klaster digunakan untuk mengidentifikasi klaster yang paling representatif atau yang berpotensi berisi *outlier*.

3. Hasil dan Pembahasan

Dataset berisi 133 *event* pariwisata dan ekonomi kreatif DKI Jakarta tahun 2024–2025. Tiga fitur utama digunakan untuk analisis pola pengunjung, yaitu jumlah pengunjung, durasi acara (hari), jenis acara (25 kategori, hasil *one-hot encoding*). Seluruh data valid tanpa *missing value*, sehingga dapat langsung diproses menggunakan *K-Means*. Transformasi *Z-Score* memastikan semua fitur berada pada skala yang sama, meskipun menghasilkan beberapa nilai ekstrim pada kategori acara langka. Hal ini penting karena dapat mempengaruhi posisi *centroid*.

Visualisasi PCA digunakan untuk memvisualisasikan distribusi *event* setelah *one-hot encoding*. PCA merupakan teknik statistik multivariat yang secara linear mengubah bentuk sekelompok variabel asli menjadi sekumpulan variabel lebih kecil yang tidak berkorelasi yang dapat mewakili informasi dari kumpulan variabel asli [13]. Dua komponen utama (PC1 dan PC2) menjelaskan sekitar 15% variansi, menunjukkan bahwa variasi data tersebar di banyak dimensi. Rendahnya variansi menjelaskan bukan berarti dataset “*sparse*”, namun mengindikasikan bahwa fitur kategori memiliki distribusi tinggi, representasi dua dimensi tidak mencerminkan keseluruhan struktur namun *cukup membantu* untuk melihat pola pemisahan klaster.

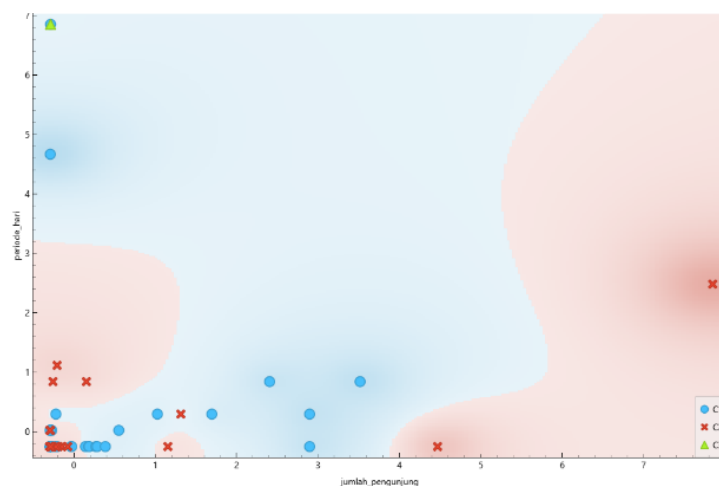


Gambar 2. Scatter Plott PCA

Pada scatter plot PCA, terlihat kecenderungan tiga kelompok besar dan satu *event* yang terisolasi sehingga dicurigai sebagai *outlier*. Klasterisasi *K-Means* dengan $K = 3$ (ditentukan berdasarkan analisis silhouette dan eksplorasi visual) menghasilkan pembagian sebagai berikut:

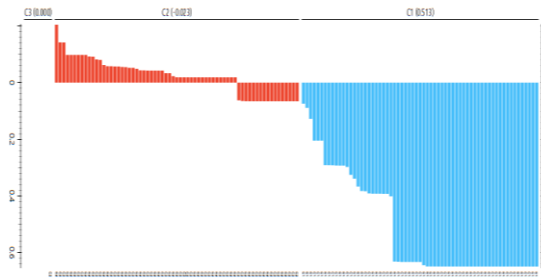
Tabel 1. Karakteristik Klaster

Klaster	Karakteristik
1	Mayoritas <i>event</i> berskala kecil-menengah, jumlah pengunjung rendah, durasi 1–5 hari.
2	Pengunjung jauh lebih besar, mayoritas pada kategori musik/festival, durasi 5–7 hari.
3	Memiliki jumlah pengunjung ekstrem tinggi dan durasi lebih panjang dari rata-rata.



Gambar 3. Visualisasi Hasil Klastering Menggunakan Scatter Plott

Scatter Plot pada gambar X menampilkan distribusi tiga klaster hasil *K-Means* berdasarkan dua fitur utama, yaitu *jumlah_pengunjung* (sumbu X) dan *periode_hari* (sumbu Y). Titik berwarna biru (C1) mendominasi area kiri bawah, menunjukkan kelompok acara kecil dengan jumlah pengunjung rendah dan durasi pendek. Titik merah (C2) berada pada area kanan grafik, menggambarkan kelompok acara besar dengan jumlah pengunjung yang jauh lebih tinggi. Sementara satu titik berwarna hijau (C3) tampak terpisah jauh dari pusat klaster lain, merepresentasikan satu outlier yang memiliki karakteristik ekstrem dibandingkan mayoritas data. Pola pemisahan ruang antar-klaster pada plot ini menunjukkan bahwa dua kelompok utama (acara kecil dan festival besar) terbentuk secara jelas, sedangkan keberadaan outlier mengonfirmasi adanya satu *event* yang sangat berbeda dari distribusi umum.



Gambar 3. Silhouette Plot

Evaluasi Kluster Menggunakan Silhouette Score menunjukkan Plot dengan Kluster 1 memiliki nilai Silhouette >0.6 (terbentuk jelas), Kluster 2 memiliki nilai 0.49 (cukup mirip), dan Kluster 3 memiliki nilai 0 (outlier). Nilai Silhouette rata-rata 0,65 termasuk kategori cukup baik. Plot ini mengurutkan event berdasarkan nilai silhouette dari tertinggi ke terendah, dengan C1 mendominasi bagian kanan, diikuti C2, dan C3 di bagian kiri. Hasil ini mengonfirmasi bahwa pola event kecil dan festival besar terdeteksi dengan akurat.

Tabel 2. Nilai Rata-Rata Silhouette Score per Kluster

Kluster	Nama	Jumlah Event	Rata-rata Silhouette	Interpretasi
1	Event Kecil	105	0.682	Cukup Baik
2	Event Besar	23	0.492	Cukup
3	Outlier	1	0.000	Netral
Rata-rata		133	0.650	Cukup Baik

Tabel 2 menunjukkan rata-rata nilai Silhouette Score untuk setiap kluster hasil algoritma K-Means. Kluster 1, yang berisi mayoritas event berskala kecil, memperoleh nilai rata-rata tertinggi yaitu 0,682 yang mengindikasikan kohesi internal yang kuat dan pemisahan yang jelas dari kluster lainnya. Kluster 2 memiliki nilai 0,492, yang berarti kualitas pemisahannya cukup baik meskipun tidak sekuat Kluster 1. Sementara itu, Kluster 3 sebagai outlier mendapatkan nilai 0,000 karena hanya terdiri dari satu event, sehingga tidak memiliki kedekatan dengan anggota kelompok lain. Secara keseluruhan, rata-rata Silhouette Score sebesar 0,650 menunjukkan bahwa struktur kluster yang terbentuk berada dalam kategori cukup baik dan mampu merepresentasikan pola umum dalam data.

4. Kesimpulan

Melalui proses klusterisasi menggunakan K-Means, penelitian ini berhasil mengidentifikasi dua kluster utama penyelenggaraan event serta satu kluster berisi outlier. Evaluasi menggunakan Silhouette Score menghasilkan nilai rata-rata 0,65, yang menunjukkan bahwa pembentukan kluster berada pada kategori cukup baik tanpa memberikan implikasi akurasi yang berlebihan. Analisis ini menyoroti adanya ketimpangan skala antara kelompok event kecil dan besar, serta kecenderungan dominasi event bertema musik pada kluster berskala kecil. Temuan tersebut dapat digunakan oleh pemangku kebijakan untuk menyusun strategi pengembangan event yang lebih seimbang, termasuk mendorong munculnya event berukuran menengah dan memperkaya variasi tema. Penelitian berikutnya perlu mempertimbangkan penambahan fitur seperti lokasi, kapasitas venue, atau anggaran penyelenggaraan untuk meningkatkan ketajaman analisis dan memungkinkan model prediksi kehadiran yang lebih robust.

5. Ucapan Terimakasih



Penulis mengucapkan terimakasih kepada semua pihak yang telah membantu dalam proses penelitian dan penyusunan laporan ini, khususnya pada dosen pembimbing dan Universitas Bina Sarana Informatika.

6. Daftar Pustaka

- [1] T. Yulianti, I. Komalasari, and A. Sudrajat, "Pengaruh Event Dan Kesadaran Masyarakat Karawang Terhadap Minat Donor Darah Di Pmi Kabupaten Karawang," *J. Ekon. Manaj.*, vol. 6, no. 1, pp. 48–54, 2020, doi: 10.37058/jem.v6i1.1229.
- [2] F. P. Putri and S. R. Nudin, "Strategi Marketing Event Organizer Menggunakan Metode

- K-Means Klastering Berbasis Web di Surabaya,” *J. Informatics Comput. Sci.*, vol. 05, pp. 73–82, 2023, [Online]. Available: <https://doi.org/10.26740/jinacs.v5n01.p73-82>
- [3] A. Etenia, R. S. Rukayah, and W. Wijayanti, “STRATEGI POTENSI DALAM PENGEMBANGAN WISATA ICE SKATING DI DALAM MALL JAKARTA (Studi Kasus: Sky Rink Taman Anggrek Jakarta),” *J. Arsit. ARCADE*, vol. 4, no. 1, p. 23, 2020, doi: 10.31848/arcade.v4i1.337.
- [4] K. Berdasarkan, P. Tanaman, and P. Di, “Jurnal Informatika Terpadu,” vol. 11, no. 1, pp. 12–19, 2025.
- [5] B. Y. Putra, F. Y. Azzahra, and I. A. Erlanda, “Klasterisasi Pengunjung Mall Menggunakan Algoritma K-Means Berdasarkan Pendapatan Dan Pengeluaran,” *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3s1, 2023, doi: 10.23960/jitet.v11i3s1.3392.
- [6] T. M. Dista and F. F. Abdulloh, “Klastering Pengunjung Mall Menggunakan Metode K-Means dan Particle Swarm Optimization,” *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1339, 2022, doi: 10.30865/mib.v6i3.4172.
- [7] D. A. Wasesha and S. Syafrianto, “Penerapan Algoritma K-Means Untuk Mengelompokkan Kunjungan Wisatawan Pada Dua Puluh Tempat Wisata Di Jakarta,” *Technol. J. Ilm.*, vol. 16, no. 1, p. 32, 2025, doi: 10.31602/tji.v16i1.16959.
- [8] B. Setio and P. Prasetyaningrum, “Penerapan Data Mining Dalam Mengelompokkan Kunjungan Wisatawan Di Kota Yogyakarta Menggunakan Metode K-Means,” *J. Comput. Sci. Technol.*, vol. 1, no. 1, pp. 27–32, 2021, doi: 10.54840/jcstech.v1i1.9.
- [9] A. Merdekawati and J. T. Kumalasari, “Model Hybrid K-Means dan Decision Tree untuk Penentuan Status Kemiskinan Penduduk Indonesia,” *J. Nas. Komputasi dan Teknol. Inf.*, vol. 8, no. 3, pp. 1680–1688, 2025, doi: 10.32672/jnkti.v8i3.9214.
- [10] S. Ana, R. Kurniawan, and A. Nazir, “Pengklasteran Risiko Covid-19 Di Riau Menggunakan Teknik One Hot Encoding Dan Algoritma K-Means Klastering,” *J. Inf. dan Komput.*, vol. 10, no. 1, pp. 154–163, 2022, doi: 10.35959/jik.v10i1.291.
- [11] N. Nurdin, “Analisa Data Mining Dalam Memprediksi Masyarakat Kurang Mampu Menggunakan Metode K-Nearest Neighbor,” *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, pp. 1090–1098, 2024, doi: 10.23960/jitet.v12i2.4131.
- [12] C. G. Selan, N. Jayanti, and M. W. Prihatmono, “Analisis Data Pengunjung Mall Nipah Mendukung Strategi Digital Marketing Menggunakan K-Means Klastering,” vol. 2, no. 1, pp. 209–220, 2022.
- [13] S. Manullang, D. Aryani, and H. Rusyda, “Analisis Principal Component Analysis (PCA) dalam Penentuan Faktor Kepuasan Pengunjung terhadap Layanan Perpustakaan Digilib,” *Edumatic J. Pendidik. Inform.*, vol. 7, no. 1, pp. 123–130, 2023, doi: 10.29408/edumatic.v7i1.14839.

7. Penulis

	<p>Nadia Umaira Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Jakarta, Indonesia.</p>
	<p>Witri Astuti Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Jakarta, Indonesia.</p>

	<p>Regina Ayu Puspita Sari Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Jakarta, Indonesia.</p>
	<p>Sheili Rika Karlina Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Jakarta, Indonesia.</p>
	<p>Ika Selviana Putri Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Jakarta, Indonesia.</p>