PESHUM : Jurnal Pendidikan, Sosial dan Humaniora
Vol.4, No.6, Oktober 2025

Analisis Klasterisasi Data pada Berbagai Bidang Menggunakan Algoritma *K-Means*: Studi Kasus dalam Kriminalitas, Ekonomi, Politik, Administrasi Digital, dan Perdagangan

Mohamad Rizky Naoval Pratama[1], Rayhan Apriansyah[2], Juliano Krizza Yoga[3], April Petra Malonovi Marbun[4], Salman Alfarisi[5], Zurnan Alfian[6]
[1,2,3,4,5,6] Universitas Pamulang, Indonesia
E-mail: saalmanalfrsi@gmail.com

Article History:
Received: 20 Juni 2025
Revised: 20 September 2025
Accepted: 26 September 2025

Keywords: *K-Means, Clustering, Data Mining, Multidomain, Model Evaluation.*

Abstract: *The K-Means algorithm is one of the most popular clustering techniques in data mining that is used to group data based on similarity of characteristics. This study aims to evaluate the application of the K-Means algorithm in five different domains: crime, digital administration, international trade, politics, and macroeconomics. Each case study uses specific datasets from reliable sources, such as BAPAS Purwokerto crime data, the level of digitization of institutions in Korea, Indonesia's export-import in 1988–1997, the results of political party elections, and Indonesia's macroeconomic indicators in 1980. The results of the analysis showed that K-Means was able to effectively group data and provide hidden patterns that are useful in the decision-making process in each field. Cluster quality evaluation was carried out using metrics such as the Silhouette Score and the Davies-Bouldin Index. This study confirms the flexibility of K-Means in handling data from various sectors and opens opportunities for the development of advanced clustering methods for multidomain analysis.*

PENDAHULUAN

Dalam era digital yang berkembang pesat, jumlah data yang tersedia semakin melimpah di berbagai sektor kehidupan (Aulia dkk., 2023). Data berbagai aktivitas manusia, seperti transaksi ekonomi, kegiatan politik, interaksi sosial, dan administrasi digital, sering kali tidak terstruktur dan membutuhkan analisis mendalam untuk menghasilkan wawasan yang berguna. Salah satu teknik utama untuk mengelompokkan data tanpa label adalah klasterisasi, memungkinkan identifikasi pola atau hubungan dalam kumpulan data besar. Di antara berbagai algoritma yang ada, *K-Means* terbukti menjadi salah satu metode efektif dan efisien (Syani & Wahyudi, 2024). Algoritma ini terkenal karena kesederhanaannya (Han dkk., 2011), kemampuannya untuk menangani data besar dengan cepat, dan efektivitas menghasilkan klaster homogen. Penggunaan *K-Means* yang luas di berbagai sektor, dari pemasaran hingga analisis ilmiah, yang menunjukkan potensi besar teknik ini untuk mengeksplorasi struktur tersembunyi dalam data yang kompleks (Andriyani dkk., 2024).

K-Means bekerja dengan membagi data ke sejumlah klaster yang ditentukan sebelumnya. Prosesnya melibatkan penentuan *centroid* (titik pusat) untuk masing-masing klaster, kemudian diperbarui secara iteratif untuk meminimalkan jarak rata-rata antara data dan *centroid*-nya. Proses

ini terus berlanjut hingga posisi *centroid* tidak lagi berubah, atau perubahan yang terjadi sangat kecil, menandakan konvergensi. Keunggulan utama *K-Means* terletak pada kemampuannya untuk mengelompokkan data dengan kecepatan tinggi, serta memberikan hasil mudah diinterpretasikan (Qibtiyah, 2024). Namun, meskipun algoritma ini memiliki sejumlah kelebihan, terdapat tantangan penerapannya, terutama dalam konteks data multidimensi dan domain yang beragam. Salah satu tantangan utama adalah menentukan jumlah klaster yang optimal, yang mempengaruhi kualitas hasil klasterisasi secara keseluruhan.

Pada ranah penelitian ini, *K-Means* diterapkan pada lima *domain* berbeda untuk menguji kekuatan dan keterbatasan algoritma ini dalam pengelompokan data yang sangat beragam. Kelima domain tersebut meliputi: 1) klasterisasi jenis kejahatan di Purwokerto, yang bertujuan untuk mengidentifikasi pola kriminalitas berdasarkan data historis; 2) segmentasi lembaga administratif berdasarkan digitalisasi, yang menilai sejauh mana lembaga-lembaga pemerintah bertransformasi secara digital; 3) pola ekspor-impor Indonesia, yang menganalisis fluktuasi perdagangan antar negara; 4) pengelompokan partai politik berdasarkan performa pemilu, yang mengelompokkan partai berdasarkan hasil pemilunya; dan 5) segmentasi bulan berdasarkan indikator ekonomi makro tahun 1980, yang menggambarkan fluktuasi ekonomi pada waktu tertentu. Setiap studi kasus ini menyediakan tantangan unik memungkinkan penelitian ini mengeksplorasi fleksibilitas *K-Means* dalam menangani data dari berbagai jenis dan ukuran.

Pentingnya penelitian ini terletak pada upaya untuk mengevaluasi kinerja *K-Means* dalam berbagai konteks melalui penggunaan metrik evaluasi yang dapat diandalkan, seperti *Silhouette Score* dan *Davies-Bouldin Index*. Metrik-metrik ini digunakan untuk menilai seberapa baik klaster yang dihasilkan terpisah satu sama lain dan sejauh mana data dalam satu klaster saling mirip. Dengan menganalisis hasil klasterisasi dari berbagai *domain,* penelitian ini bertujuan memberikan wawasan baru tentang aplikabilitas *K-Means* di berbagai bidang dan meningkatkan pemahaman tentang bagaimana metode ini dapat diadaptasi untuk tantangan analisis *multidomain* yang kompleks. Diharapkan, temuan dari penelitian ini dapat membuka jalan bagi penggunaan metode klasterisasi yang lebih canggih dan spesifik dalam pengelolaan data besar dan analisis yang lebih mendalam di berbagai sektor.

LANDASAN TEORI

Klasterisasi adalah metode dalam analisis data yang digunakan mengelompokkan objek atau data berdasarkan kesamaan karakteristik atau atribut tertentu tanpa memerlukan label (Putriana dkk., 2023). Metode ini yang berfungsi mengidentifikasi struktur tersembunyi dalam data tidak terstruktur. Dalam klasterisasi, data yang memiliki kesamaan akan dikelompokkan satu klaster, sementara data yang berbeda dikelompokkan di klaster yang berbeda. Berbagai teknik klasterisasi telah dikembangkan, tetapi *K-Means* menjadi salah satu algoritma yang paling banyak digunakan karena kesederhanaannya, efisiensinya, kemampuannya dalam menangani data berukuran besar. *K-Means* bekerja dengan cara membagi data dalam sejumlah klaster yang ditentukan, kemudian mengoptimalkan posisi *centroid* setiap klaster hingga proses konvergen.

Dalam penerapan *K-Means*, ada beberapa konsep yang perlu dipahami. Proses iteratif *K-Means* dimulai pemilihan sejumlah *centroid* secara acak, kemudian setiap data dikelompokkan ke dalam klaster berdasarkan kedekatannya dengan *centroid* yang ada. Setelah itu, posisi *centroid* dihitung ulang berdasarkan rata-rata data dalam klaster tersebut dan proses ini diulang hingga posisi *centroid* stabil. Salah satu tantangan utama dalam penerapan K-Means adalah penentuan jumlah klaster yang optimal, yang sering kali membutuhkan pendekatan heuristik atau penggunaan teknik evaluasi seperti *Elbow Method, Silhouette Score,* atau *Davies-Bouldin Index*.

Selain itu, *K-Means* memiliki keterbatasan dalam hal kepekaan terhadap *outlier,* yang dapat memengaruhi hasil klasterisasi jika data yang dianalisis mengandung banyak nilai ekstrim.

Penerapan *K-Means* pada berbagai domain memiliki relevansi yang tinggi, karena teknik ini memungkinkan analisis data yang mendalam dalam banyak bidang (Mirna dkk., 2025). Dalam konteks kriminologi, *K-Means* digunakan mengelompokkan jenis-jenis kejahatan berdasarkan karakteristik geografis dan temporal, yang dapat membantu dalam penentuan strategi pencegahan kejahatan yang lebih efisien. Di bidang ekonomi, *K-Means* digunakan untuk menganalisis pola ekspor-impor, sehingga membantu pengambilan keputusan kebijakan perdagangan internasional. Selain itu, dalam dunia politik, algoritma ini digunakan mengelompokkan partai-partai politik berdasarkan hasil pemilu, yang berguna dalam merumuskan strategi kampanye atau analisis tren politik. Penerapan *K-Means* dalam analisis administratif digital ini dapat memberikan wawasan tentang kesiapan lembaga pemerintahan dalam mengadopsi teknologi digital, sedangkan dalam bidang ekonomi makro, *K-Means* membantu dalam segmentasi data ekonomi untuk memahami fluktuasi yang mempengaruhi pertumbuhan ekonomi. Dengan demikian, algoritma *K-Means* tidak hanya memiliki kegunaan dalam analisis data tradisional, tetapi juga sangat fleksibel.

METODE PENELITIAN

Metodologi penelitian ini terdiri dari tiga tahap utama yang saling berkaitan: pra-pemrosesan data, klasterisasi menggunakan algoritma *K-Means*, dan evaluasi hasil klasterisasi. Tahap pertama, yaitu pra-pemrosesan data, yang bertujuan untuk mempersiapkan data agar dapat digunakan dalam proses klasterisasi. Mengingat setiap studi kasus menggunakan *dataset* dengan karakteristik yang berbeda, tahapan ini yang disesuaikan dengan jenis data yang ada. Untuk fitur numerik, dilakukan normalisasi menggunakan *StandardScaler* dari pustaka *Scikit-learn*, memastikan bahwa setiap fitur memiliki skala yang sebanding. Untuk data satu dimensi, seperti jumlah klien pada kasus-kasus kriminalitas, dilakukan klasterisasi langsung tanpa reduksi dimensi. Namun, pada data berdimensi tinggi, dilakukan reduksi dimensi menggunakan *Principal Component Analysis* (PCA) agar data dapat divisualisasikan dalam dua dimensi, yang memungkinkan mereduksi jumlah fitur menjadi lebih sedikit dengan mempertahankan sebanyak mungkin informasi yang relevan dari data asli, sehingga data tersebut dapat divisualisasikan dalam dua dimensi atau lebih, yang pada gilirannya memudahkan proses analisis dan interpretasi dengan cara yang jelas dan efisien, serta membantu mengidentifikasi pola-pola tersembunyi yang mungkin sulit untuk ditemukan.

Tahap kedua adalah proses klasterisasi dengan algoritma K-Means. Proses ini yang dimulai dengan inisialisasi *centroid* secara acak sebanyak k buah, di mana nilai k ditentukan berdasarkan analisis atau melalui metode *elbow* dan *Silhouette Score*. Setelah *centroid* terinisialisasi, setiap data dikelompokkan ke dalam klaster yang paling dekat dengan centroid, berdasarkan jarak *Euclidean.* Proses ini diulang secara iteratif hingga posisi *centroid* tidak mengalami perubahan signifikan, yang menandakan bahwa algoritma telah mencapai konvergensi. Penerapan algoritma K-Means dilakukan dengan menggunakan modul *KMeans* dari pustaka *Scikit-learn* versi 1.3 ke atas, yang telah terbukti efektif dalam menangani klasterisasi data besar dan kompleks.

Tahap terakhir adalah evaluasi hasil klasterisasi, di mana kualitas klaster yang dihasilkan akan dinilai menggunakan tiga metrik evaluasi yang relevan. Pertama, *Silhouette Score* digunakan untuk mengukur seberapa baik setiap data berada dalam klasternya sendiri dibandingkan dengan klaster lain. Nilai *Silhouette Score* berkisar -1 hingga 1, dengan nilai lebih tinggi menunjukkan kualitas klaster yang lebih baik. Kedua, *Davies-Bouldin* Index mengukur kemiripan antar klaster, di mana nilai yang lebih rendah menunjukkan klaster yang lebih terpisah

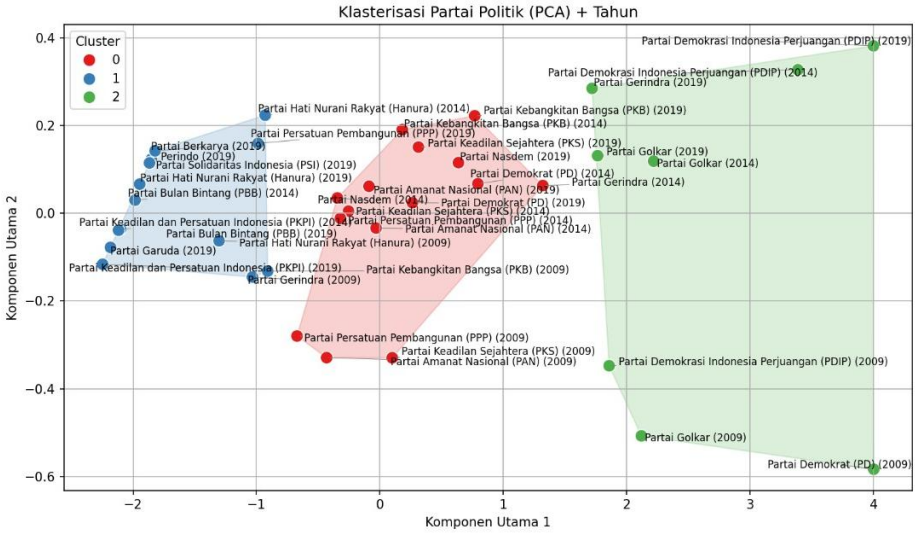
dan berkualitas. Ketiga, *Inertia* mengukur total jarak kuadrat antara data dan *centroid*-nya, dengan nilai yang lebih kecil menunjukkan bahwa titik data lebih dekat dengan *centroid,* menandakan klaster yang lebih rapat dan kohesif. Semua tahapan ini dilakukan dengan menggunakan bahasa pemrograman *Python,* serta pustaka pendukung seperti *pandas, numpy, matplotlib, seaborn*, dan *scikit-learn.*

HASIL DAN PEMBAHASAN

1. Klasterisasi Partai Politik (PCA) + Tahun

Dataset yang digunakan dalam analisis ini berisi data terkait suara, kursi, dan persentase partai politik dalam pemilu. Data ini mencakup hasil pemilu yang menggambarkan perolehan suara dan kursi masing-masing partai, serta proporsi partai dalam struktur politik suatu negara. Untuk memudahkan analisis dan memastikan data berada pada skala yang sama, dilakukan standarisasi terhadap variabel dalam *dataset,* termasuk suara, kursi, dan persentase partai. Setelah dilakukan standarisasi, algoritma K-Means digunakan untuk mengelompokkan partai-partai politik ke dalam tiga klaster berdasarkan perolehan suara, kursi, dan persentase mereka dalam pemilu. Ketiga klaster ini mencerminkan kategori partai, yaitu partai besar, partai sedang, dan partai kecil, yang dapat memberikan gambaran mengenai dominasi partai dalam peta politik.

Hasil klasterisasi ini divisualisasikan menggunakan teknik reduksi dimensi PCA *(Principal Component Analysis)* dalam bentuk plot 2D. Visualisasi ini menampilkan distribusi partai-partai berdasarkan klaster yang terbentuk, di mana partai besar, sedang, dan kecil dapat dilihat dengan jelas pada sumbu X dan Y. Dengan menggunakan PCA, informasi dari dimensi yang lebih tinggi diubah menjadi dua dimensi sehingga distribusi partai-partai di ruang klaster dapat divisualisasikan dengan mudah. Visualisasi ini memperkuat pembagian klaster yang dilakukan oleh algoritma *K-Means*, memberikan gambaran yang jelas mengenai seberapa besar perbedaan antara klaster-klaster yang terbentuk berdasarkan data pemilu. Dalam evaluasi visual, terlihat bahwa partai-partai dalam klaster besar memiliki posisi yang lebih terpusat di satu area, sementara partai-partai kecil tersebar di area yang luas, menunjukkan perbedaan signifikan dalam jumlah suara dan kursi.



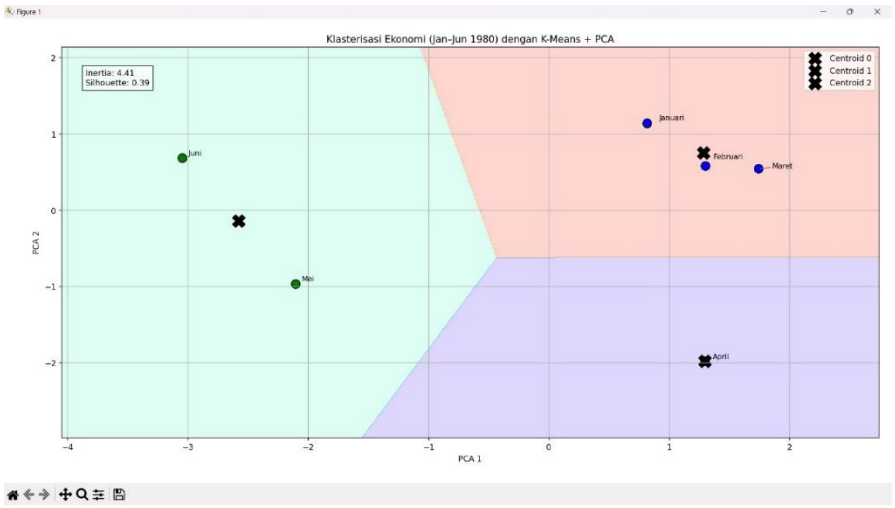
Gambar 1. Data Klasterisasi Setiap Partai Politik Tahun Ke Tahun

Hasil klasterisasi ini memberikan wawasan mendalam mengenai distribusi dan dominasi partai-partai dalam pemilu. Klaster partai besar ini menunjukkan dominasi signifikan dalam hal perolehan suara dan kursi, menggambarkan kekuatan politik yang lebih besar dalam sistem pemilu. Partai-partai ini cenderung memiliki sumber daya dan dukungan yang lebih luas, sehingga mampu mempengaruhi arah kebijakan dan pembentukan pemerintahan. Sebaliknya, partai dalam klaster sedang memiliki posisi yang lebih moderat, dengan perolehan suara dan kursi yang cukup tetapi tidak dominan. Klaster partai kecil, yang tersebar luas dalam visualisasi, mencerminkan partai-partai dengan perolehan suara dan kursi yang lebih rendah, yang sering kali berfungsi sebagai partai alternatif atau dengan pengaruh politik yang terbatas. Evaluasi visual ini memperkuat pemahaman tentang bagaimana struktur partai di pemilu dipetakan berdasar data kuantitatif dan memberikan informasi berharga bagi strategi kampanye dan perencanaan politik di masa depan. Klasterisasi ini juga bisa menjadi alat berguna bagi pemangku kebijakan untuk menilai keseimbangan kekuatan politik dalam sistem demokrasi dan memahami dinamika politik yang ada.

2. Klasterisasi Ekonomi (Jan-Jun 1980) dengan *K-Means* + PCA (Indonesia 1980)

Data yang digunakan dalam analisis ini mencakup data bulanan dari Januari hingga Juni 1980, dengan fitur-fitur yang mencerminkan kondisi ekonomi pada periode tersebut, yaitu inflasi, pengangguran, pertumbuhan uang beredar (M1B, M2), dan suku bunga. Data ini mencerminkan indikator-indikator ekonomi penting yang dapat memberikan gambaran mengenai stabilitas ekonomi suatu negara pada saat itu. Untuk mengidentifikasi pola dan membagi data ke dalam kelompok yang lebih bermakna, dilakukan klasterisasi menggunakan algoritma K-Means dengan jumlah klaster (k) yang ditentukan sebesar 3. Tujuan klasterisasi ini adalah untuk mengelompokkan bulan-bulan dalam *dataset* dalam tiga kondisi ekonomi yang berbeda, yaitu kondisi krisis, kondisi stabil, dan kondisi pemulihan, berdasarkan pola yang muncul pada indikator ekonomi.

Hasil klasterisasi dengan *K-Means* ini menghasilkan tiga kelompok yang menggambarkan dinamika ekonomi Indonesia pada paruh pertama tahun 1980. Klaster pertama, yang mewakili kondisi krisis, menunjukkan bulan-bulan dengan inflasi dan pengangguran yang tinggi, serta suku bunga yang relatif tinggi dan pertumbuhan uang beredar yang mungkin tertekan atau fluktuatif. Klaster kedua, yang menggambarkan kondisi stabil, mencakup bulan-bulan dengan angka inflasi dan pengangguran yang lebih terkendali, serta suku bunga dan pertumbuhan uang beredar yang seimbang, mencerminkan periode di mana ekonomi berjalan relatif normal dan tidak terpengaruh guncangan eksternal besar. Klaster ketiga, yang mencerminkan kondisi pemulihan, menunjukkan bulan-bulan dengan perbaikan pada indikator ekonomi, yang seperti penurunan inflasi, penurunan pengangguran, dan peningkatan pertumbuhan uang beredar. Evaluasi klasterisasi menunjukkan pemisahan antar klaster cukup baik, dengan hasil yang jelas dan dapat dipahami. Berdasarkan hasil evaluasi, klaster-klaster tersebut terpisah dengan baik dan masing-masing mewakili fase-fase yang berbeda dalam siklus ekonomi. Penggunaan *K-Means* dengan k=3 berhasil menciptakan kelompok yang menunjukkan tren berbeda dalam indikator-indikator ekonomi, yang memberikan wawasan mengenai bagaimana kondisi ekonomi Indonesia saat itu berfluktuasi. Evaluasi ini memperkuat pendekatan klasterisasi ini efektif mengidentifikasi dan memetakan perubahan ekonomi.



Gambar 2. Data Klasterisasi Inflasi Ekonomi Bulan Januari-Juni 1980

Hasil klasterisasi ini memberikan gambaran lebih jelas mengenai siklus ekonomi Indonesia pada paruh pertama tahun 1980, dengan membagi bulan-bulan ke dalam tiga kondisi utama: krisis, stabil, dan pemulihan. Klaster pertama, yang mencerminkan kondisi krisis, menunjukkan periode yang penuh tekanan ekonomi, dengan indikator-inflasi dan pengangguran yang tinggi, serta suku bunga yang meningkat. Kondisi ini bisa mencerminkan adanya guncangan ekonomi, seperti krisis moneter atau krisis internasional yang mempengaruhi stabilitas ekonomi domestik. Klaster kedua, mewakili kondisi stabil, menggambarkan periode ketika ekonomi Indonesia dapat mengendalikan inflasi dan pengangguran, dengan tingkat suku bunga yang lebih terkontrol dan pertumbuhan uang beredar yang moderat. Hal ini mencerminkan keberhasilan kebijakan ekonomi dalam menjaga stabilitas. Klaster ketiga menunjukkan kondisi pemulihan, yang mencerminkan perbaikan ekonomi setelah periode krisis. Penurunan inflasi dan pengangguran, serta peningkatan pertumbuhan uang beredar, mengindikasikan langkah-langkah berhasil dalam mendorong pemulihan ekonomi.

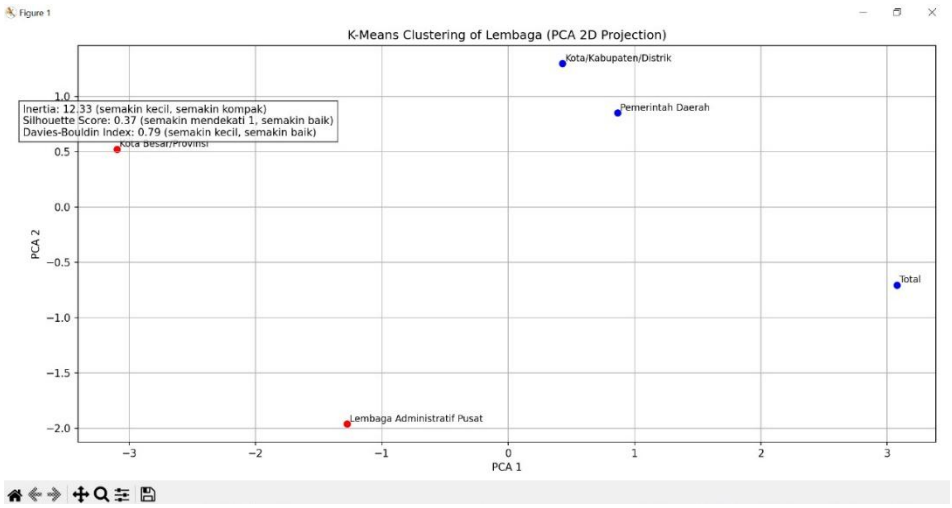
3. *K-Means* Klasterisasi Setiap Lembaga, Administrasi Digital (Korea)

Dataset yang digunakan dalam analisis ini berisi fitur-fitur yang berkaitan dengan tingkat adopsi teknologi dalam lembaga, seperti tingkat pertukaran *e-Document* dan *e-Approval*. Kedua fitur ini mencerminkan sejauh mana lembaga-lembaga tersebut telah mengimplementasikan sistem elektronik mendukung administrasi dan operasional mereka. *e-Document* merujuk penggunaan dokumen digital yang dapat dipertukarkan secara elektronik antara berbagai pihak, sementara *e-Approval* mengacu pada persetujuan dokumen atau keputusan secara elektronik, menggantikan metode manual tradisional. Tujuan dari analisis ini adalah mengelompokkan lembaga-lembaga tersebut berdasarkan tingkat adopsi teknologi yang mereka terapkan, menggunakan algoritma K-Means. Dengan menggunakan *K-Means*, *dataset* dibagi menjadi dua klaster utama yang mewakili lembaga dengan tingkat adopsi teknologi yang rendah dan tinggi.

Hasil klasterisasi menunjukkan dua kelompok besar berdasarkan tingkat penggunaan e-*Document* dan *e-Approval.* Klaster pertama mencakup lembaga-lembaga dengan tingkat adopsi teknologi yang rendah, di mana penggunaan *e-Document* dan *e-Approval* masih terbatas atau baru dimulai. Lembaga dalam klaster ini kemungkinan besar masih bergantung pada metode manual untuk pengelolaan dokumen dan proses persetujuan, dengan sedikit integrasi teknologi dalam operasional mereka. Sementara itu, klaster kedua ini mencakup lembaga-lembaga tingkat adopsi teknologi yang lebih tinggi, di mana sistem *e-Document* dan *e-Approval* sudah diterapkan

secara lebih luas dan menjadi bagian penting dari sistem administrasi dan operasional mereka. Hasil visualisasi menggunakan *Principal Component Analysis* (PCA) dalam dimensi 2D memperlihatkan pemisahan yang cukup baik antara kedua klaster ini. Visualisasi ini memungkinkan pemahaman yang lebih mudah tentang bagaimana lembaga-lembaga ini dikelompokkan berdasarkan tingkat penggunaan teknologi.

Evaluasi terhadap klasterisasi ini menunjukkan hasil yang cukup baik meskipun masih ada ruang untuk peningkatan. Nilai *Silhouette Score* sebesar 0.37 menunjukkan bahwa pemisahan antar klaster sudah cukup jelas, namun masih ada tumpang tindih atau *overlapping* antar klaster yang perlu diperbaiki. Sebuah *Silhouette Score* di atas 0.5 umumnya menunjukkan pemisahan klaster yang sangat baik, sehingga nilai 0.37 ini menandakan bahwa meskipun klasterisasi memberikan hasil yang bermakna, ada beberapa lembaga yang mungkin tidak sepenuhnya sesuai dengan klaster yang mereka masuki. Selain itu, *Davies-Bouldin Index* sebesar 0.79 juga menunjukkan kualitas klasterisasi cukup baik, meskipun masih ada peluang untuk meningkatkan kejelasan pemisahan antar klaster. Semakin kecil nilai *Davies-Bouldin Index*, semakin baik pemisahan antara klaster-klaster yang ada. Dengan kata lain, meskipun klasterisasi ini cukup memadai, ada ruang untuk meningkatkan ketepatan pembagian klaster dengan mempertimbangkan faktor-faktor lain yang mungkin mempengaruhi tingkat adopsi teknologi lembaga.



Gambar 3. Dokumen Antar Lembaga

Hasil analisis ini memberikan gambaran mengenai sejauh mana lembaga-lembaga dalam *dataset* telah mengadopsi teknologi dalam pengelolaan dokumen dan proses persetujuan mereka. Klaster pertama yang mewakili lembaga dengan tingkat adopsi teknologi rendah menunjukkan bahwa masih ada banyak lembaga yang belum sepenuhnya beralih ke sistem digital, yang dapat menjadi indikasi mereka mungkin menghadapi kendala dalam sumber daya atau infrastruktur untuk mengimplementasikan teknologi. Klaster kedua ini, menunjukkan lembaga dengan tingkat adopsi teknologi tinggi, mencerminkan lembaga yang telah berhasil mengintegrasikan teknologi dalam operasional mereka, yang memungkinkan mereka meningkatkan efisiensi, mengurangi biaya, dan mempercepat administrasi. Evaluasi yang menunjukkan nilai *Silhouette Score* 0.37 dan *Davies-Bouldin Index* 0.79 mengindikasikan bahwa meskipun klasterisasi ini memberikan pembagian yang cukup baik antara lembaga dengan adopsi teknologi rendah dan tinggi, terdapat area yang masih bisa diperbaiki. Hal ini menunjukkan faktor lain, seperti kesiapan digital,

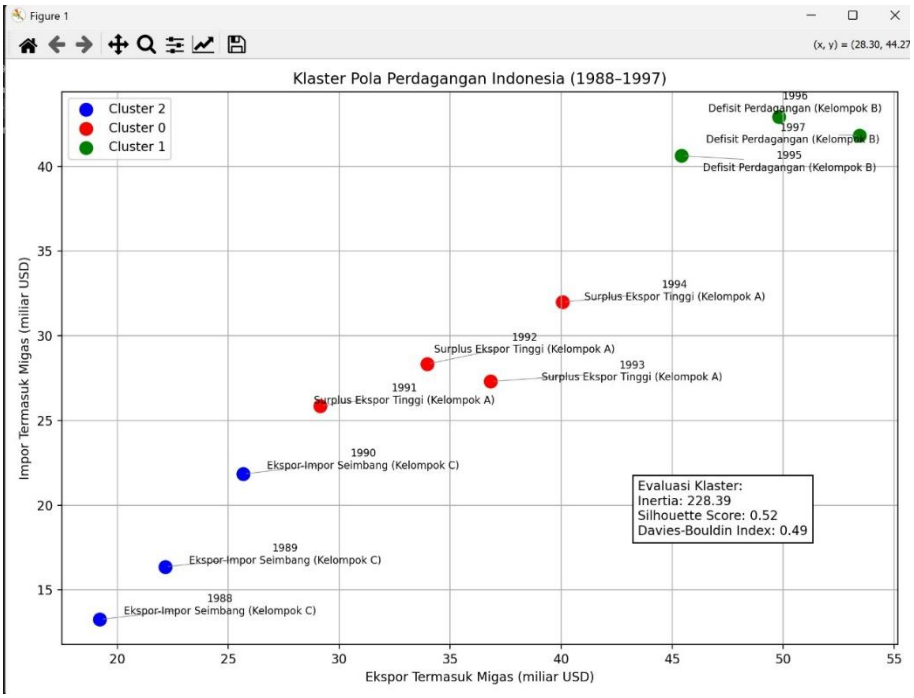
pelatihan sumber daya manusia, atau kebijakan mendukung teknologi, bisa menjadi elemen penting yang mempengaruhi adopsi teknologi dan perlu dipertimbangkan lebih dalam untuk mendapatkan pemisahan klaster yang lebih tajam. Pemangku kebijakan dan pengambil keputusan dapat menggunakan informasi ini untuk merencanakan inisiatif yang efektif dalam mendorong lembaga yang masih tertinggal dalam adopsi teknologi untuk bertransformasi menuju sistem yang lebih modern dan efisien.

4. Klasterisasi Pola Perdagangan Ekspor-Impor (Indonesia 1988–1997)

Data ekspor dan impor Indonesia per tahun digunakan mengidentifikasi pola perdagangan luar negeri Indonesia dengan mengolah data tersebut menjadi fitur-fitur penting, yaitu ekspor, impor, dan selisih (neraca perdagangan). Fitur ekspor mencerminkan total nilai barang dan jasa yang dijual ke luar negeri, sementara impor mencakup total nilai barang dan jasa yang dibeli dari luar negeri. Selisih antara ekspor dan impor, yang mana merupakan indikator neraca perdagangan, menggambarkan apakah suatu tahun mengalami surplus atau defisit di perdagangan internasional. Untuk lebih memahami dinamika ini, digunakan algoritma *K-Means* untuk membagi tahun-tahun tersebut ke dalam tiga klaster utama: surplus tinggi, defisit, dan seimbang, berdasarkan nilai ekspor, impor, dan selisih perdagangan.

Hasil klasterisasi dengan *K-Means* menunjukkan data perdagangan Indonesia dapat dibagi menjadi tiga kelompok yang menggambarkan kondisi perdagangan yang berbeda. Klaster pertama, yang mencerminkan surplus tinggi, berisi tahun-tahun di mana nilai ekspor Indonesia signifikan lebih tinggi dibandingkan dengan impor, menghasilkan neraca perdagangan yang sangat positif. Tahun-tahun dalam klaster yang menggambarkan periode ketika Indonesia berhasil mencatatkan surplus perdagangan yang besar, yang berpotensi memperkuat cadangan devisa dan mendukung kestabilan ekonomi. Klaster kedua, yang menggambarkan defisit, terdiri dari tahun-tahun di mana impor melebihi ekspor, menghasilkan defisit perdagangan yang cukup besar. Tahun-tahun dalam klaster ini yang mencerminkan periode ketika Indonesia mengimpor lebih banyak barang dan jasa daripada yang diekspor, yang dapat menunjukkan ketergantungan yang lebih besar terhadap barang dan jasa dari luar negeri, serta potensi risiko terhadap neraca pembayaran. Klaster ketiga, yang menunjukkan kondisi seimbang, berisi tahun-tahun di mana ekspor dan impor relatif seimbang, dengan neraca perdagangan yang tidak terlalu jauh dari titik nol, mencerminkan keseimbangan antara aliran barang dan jasa yang masuk dan keluar dari Indonesia.

Pola yang ditemukan melalui klasterisasi ini menunjukkan perubahan posisi perdagangan Indonesia dari surplus ke defisit dalam beberapa tahun terakhir. Pada periode tertentu, Indonesia mengalami periode surplus perdagangan yang tinggi, yang didorong faktor-faktor seperti harga komoditas global yang tinggi, peningkatan daya saing produk Indonesia di pasar internasional, atau kebijakan ekspor yang mendukung. Namun, di beberapa tahun berikutnya, terjadi pergeseran ke kondisi defisit, disebabkan faktor, termasuk meningkatnya konsumsi domestik, ketergantungan pada impor barang modal dan konsumsi, serta fluktuasi harga komoditas yang berdampak pada pendapatan ekspor. Visualisasi hasil klasterisasi ini memberikan gambaran yang lebih jelas tentang bagaimana neraca perdagangan Indonesia telah berubah selama periode yang dianalisis, serta membantu untuk mengidentifikasi tren jangka panjang dalam pola ekspor dan impor.



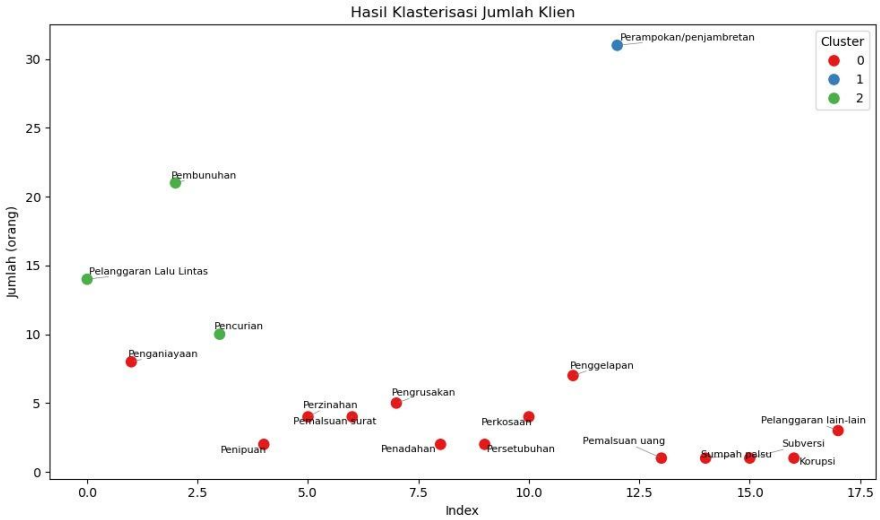
Gambar 4. Klasterisasi Pola Perdagangan Indonesia Tahun 1988 Hingga 1997

Hasil klasterisasi ini yang menggambarkan dinamika perubahan perdagangan Indonesia yang menunjukkan pergeseran dari kondisi surplus perdagangan yang kuat ke kondisi defisit. Klaster surplus tinggi menggambarkan tahun-tahun di mana Indonesia memiliki kinerja ekspor yang sangat baik, yang memberikan kontribusi positif terhadap pertumbuhan ekonomi dan stabilitas neraca pembayaran. Tahun-tahun ini mencerminkan puncak dalam sektor-sektor ekspor Indonesia, seperti komoditas energi, hasil pertanian, atau produk manufaktur mendapat permintaan tinggi di pasar global. Namun, klaster defisit menggambarkan periode di mana ketergantungan Indonesia terhadap impor semakin meningkat, yang bisa menandakan beberapa tantangan perekonomian, yang seperti pertumbuhan yang lebih cepat dari konsumsi domestik atau ketergantungan pada barang-barang modal yang penting untuk pembangunan infrastruktur. Pergeseran dari surplus ke defisit ini juga menunjukkan bahwa Indonesia mungkin perlu lebih fokus pada upaya diversifikasi ekonomi dan peningkatan daya saing produk ekspor, agar dapat kembali mencatatkan surplus perdagangan.

5. Klasterisasi Jumlah Klien, Kriminalitas (BAPAS Purwokerto)

Dataset yang digunakan ini mencakup data jumlah klien pemasyarakatan berdasarkan jenis kejahatan. Setiap baris dalam *dataset* ini mewakili jenis kejahatan tertentu, dengan fitur utama yang digunakan dalam analisis adalah jumlah klien, yang merujuk pada jumlah individu yang terlibat dalam kasus kejahatan tersebut dan menjalani pemasyarakatan. Fitur numerik ini menggambarkan seberapa banyak individu yang terlibat dalam berbagai jenis kejahatan yang tercatat dalam sistem pemasyarakatan. Untuk mengidentifikasi pola atau kelompok yang berbeda dalam data, digunakan algoritma *K-Means* mengelompokkan jenis-jenis kejahatan dalam tiga klaster berdasarkan jumlah kliennya, yang dikategorikan sebagai klaster tinggi, sedang, dan rendah. Klaster ini memberikan gambaran yang lebih jelas tentang distribusi klien pemasyarakatan berdasarkan jenis kejahatan dan skala keterlibatannya. Hasil klasterisasi dengan *K-Means* membagi jenis kejahatan ke dalam tiga klaster utama. Klaster pertama, jenis kejahatan dengan jumlah klien yang tinggi, berisi kejahatan-kejahatan yang sering terjadi dan melibatkan

banyak individu, seperti kejahatan terkait narkotika, pencurian, atau kekerasan. Klaster ini kejahatan yang memiliki dampak besar dalam masyarakat dan cenderung memberikan beban yang lebih tinggi pada sistem pemasyarakatan. Klaster kedua, yang mewakili jenis kejahatan dengan jumlah klien sedang, mencakup kejahatan-kejahatan yang terjadi dengan frekuensi menengah, mungkin melibatkan individu dalam jumlah sedang, seperti penipuan atau pelanggaran peraturan lalu lintas. Klaster ketiga mencerminkan kejahatan-kejahatan yang relatif jarang terjadi atau melibatkan sedikit individu, seperti kejahatan korupsi atau kejahatan yang lebih bersifat khusus dan terisolasi.



Gambar 5. Keadaan Klien Berdasarkan Jenis Kejahatan

Visualisasi hasil klasterisasi yang menggunakan *scatter plot,* yang menunjukkan distribusi jenis-jenis kejahatan dalam ruang dua dimensi. Hasil visualisasi ini yang mengungkapkan bahwa distribusi jenis kejahatan dalam tiga klaster tidak merata, dengan beberapa klaster menunjukkan konsentrasi yang lebih tinggi di area tertentu, sementara yang lainnya tersebar lebih merata. Ketidakseimbangan distribusi ini dapat menunjukkan bahwa beberapa jenis kejahatan memiliki prevalensi yang jauh lebih tinggi dibandingkan yang lainnya, yang bisa menjadi indikasi adanya faktor-faktor tertentu yang mempengaruhi terjadinya kejahatan dalam masyarakat. Evaluasi terhadap hasil klasterisasi menggunakan *Silhouette Score* menghasilkan nilai 0.62, menunjukkan bahwa pemisahan antar klaster cukup baik.

Hasil klasterisasi ini memberikan wawasan yang berguna mengenai distribusi dan prevalensi jenis-jenis kejahatan berdasarkan jumlah klien yang terlibat. Klaster pertama, yang menunjukkan jenis kejahatan dengan jumlah klien tinggi, mencerminkan jenis kejahatan yang paling sering terjadi dan memberi tekanan besar terhadap sistem pemasyarakatan. Kejahatan seperti narkotika, pencurian, dan kekerasan memiliki dampak yang luas, baik dalam hal jumlah kasus maupun jumlah individu yang terlibat, sehingga perlu mendapatkan perhatian dalam kebijakan penanggulangan dan rehabilitasi. Klaster kedua dengan jumlah klien sedang mungkin mencakup kejahatan yang sering terjadi, namun tidak sebanyak jenis kejahatan yang ada pada klaster pertama. Jenis kejahatan ini masih menuntut perhatian, tetapi dengan pendekatan yang lebih terfokus dan strategis, karena tidak sebanyak kejahatan yang termasuk dalam kategori tinggi. Klaster ketiga, yang berisi jenis kejahatan dengan jumlah klien rendah, mungkin mencerminkan jenis kejahatan yang lebih spesifik atau yang melibatkan individu dengan tingkat kejahatan

tertentu, seperti kejahatan ekonomi atau korupsi. Meskipun jumlah kliennya rendah, jenis kejahatan ini seringkali memiliki dampak jangka panjang yang besar terhadap masyarakat dan perekonomian, sehingga tetap penting untuk mendapatkan perhatian pihak berwenang. Evaluasi *Silhouette Score* sebesar 0.62 menunjukkan bahwa klasterisasi ini berhasil membedakan kelompok-kelompok jenis kejahatan dengan cukup baik, meskipun ada sedikit tumpang tindih yang perlu diperhatikan.

Meskipun *K-Means* ini menunjukkan performa baik dalam lima studi kasus yang dianalisis, terdapat beberapa keterbatasan inheren dari algoritma ini. Pertama, *K-Means* sensitif terhadap pemilihan *centroid* awal yang acak, sehingga dapat menghasilkan hasil klasterisasi yang berbeda pada eksekusi yang berbeda. Kedua, algoritma ini mengasumsikan bentuk klaster yang cenderung bulat dan berukuran seragam, sehingga kurang efektif untuk data dengan distribusi non-linear atau klaster dengan ukuran yang sangat berbeda. Ketiga, K-*Means* mensyaratkan penentuan jumlah klaster (k) di awal, yang tidak selalu mudah ditentukan secara optimal. Terakhir, algoritma ini tidak tahan terhadap keberadaan *outlier* yang dapat menggeser posisi *centroid* secara signifikan. Untuk mengatasi keterbatasan, penelitian lanjutan mempertimbangkan penggunaan algoritma alternatif seperti DBSCAN yang tidak memerlukan nilai k dan lebih tahan terhadap *outlier,* atau klasterisasi hierarki yang mampu menangkap struktur data yang kompleks. Selain itu, pendekatan *ensemble* atau kombinasi klasterisasi juga dapat dijajaki meningkatkan stabilitas dan akurasi pengelompokan data multi *domain.* Penelitian ini membuktikan bahwa algoritma *K-Means* mampu digunakan secara efektif dalam proses klasterisasi data pada lima domain berbeda: kriminalitas, administrasi digital, perdagangan internasional, politik, dan ekonomi makro. Hasil klasterisasi menunjukkan kemampuan *K-Means* mengidentifikasi pola-pola tersembunyi yang relevan dengan karakteristik masing-masing domain, serta memberikan informasi yang bermanfaat bagi proses pengambilan keputusan berbasis data. Evaluasi menggunakan metrik *Silhouette Score* dan *Davies-Bouldin Index* mengindikasikan pembagian klaster memiliki tingkat keterpisahan dan kekompakan yang cukup baik. Hal ini menunjukkan bahwa K-Means tetap merupakan salah satu algoritma yang andal dalam eksplorasi data multi *domain*, terutama *dataset* dengan struktur linier dan skala homogen. Namun demikian, perlu dicatat bahwa *K-Means* memiliki beberapa keterbatasan seperti kepekaan terhadap nilai awal *centroid,* kebutuhan menentukan jumlah klaster di awal, dan ketidaksesuaian terhadap bentuk klaster yang kompleks. Oleh karena itu, untuk penelitian lanjutan yang disarankan untuk mengeksplorasi algoritma alternatif seperti DBSCAN atau klasterisasi hierarki guna memperluas perspektif dan meningkatkan kualitas segmentasi data yang lebih kompleks.

KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma *K-Means* dapat diterapkan secara efektif pada berbagai bidang dengan struktur data dan tujuan analisis yang berbeda. Dari kelima studi kasus ini, *K-Means* memiliki fleksibilitas dan efisiensi yang luar biasa dalam menemukan pola tersembunyi dalam data serta memudahkan proses pengambilan keputusan berbasis data yang lebih objektif dan sistematis. Evaluasi menggunakan metrik *Silhouette Score* dan *Davies-Bouldin Index* menegaskan hasil klasterisasi yang dihasilkan memiliki kualitas yang cukup baik, dengan tingkat keterpisahan dan kekompakan klaster yang memadai. *K-Means* memiliki keterbatasan, terutama terkait dengan kepekaan terhadap pemilihan *centroid* awal dan ketergantungan pada jumlah klaster. Oleh karena itu, K-Means tetap menjadi alat yang relevan dan bermanfaat dalam eksplorasi data multi domain, namun untuk penelitian selanjutnya, dipertimbangkan penggunaan algoritma lain seperti DBSCAN atau klasterisasi hierarki untuk membandingkan hasil dan juga

mengatasi keterbatasan K-Means dalam menangani data dengan struktur non-linear atau klaster yang lebih kompleks. Penggunaan algoritma alternatif ini diharapkan dapat memperkaya perspektif dalam proses klasterisasi dan meningkatkan kualitas segmentasi data yang lebih mendalam dan lebih akurat.

DAFTAR REFERENSI

Andriyani, W., Natsir, F., Lubis, H., Yulianing Tyas, S. H., Meidelfi, D., Faizah, S., Nurlaida, N., Kurniawan, H., Wahyuningtyas, I., Hasan, F. N., Afandi, I. R., Sindrawati, S., & Hikmawati, E. (2024). *Perangkat lunak data mining*. Penerbit Widina.

Aulia, B. W., Rizki, M., Prindiyana, P., & Surgana, S. (2023). Peran krusial jaringan komputer dan basis data dalam era digital. *JUSTINFO| Jurnal Sistem Informasi dan Teknologi Informasi*, *1*(1), 9-20. https://doi.org/10.33197/justinfo.vol1.iss1.2023.1253

Badan Pusat Statistik. (2001). *Statistik ekspor-impor Indonesia 1988–1997*. BPS.

data_ekonomi_1980.csv [Data set, dataset internal].

data_klien_kejahatan.csv [Data set, dataset internal].

data_pemilu_gabungan.csv [Data set, dataset internal].

Hackeling, G. (2017). *Mastering Machine Learning with scikit-learn*. Packt Publishing Ltd.

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.

Isra, S., & Fahmi, K. (2019). *Pemilihan umum demokratis*. Rajawali Pers.

Mankiw, N. G. (2010). *Makroekonomi* (Edisi ke-4). Salemba Empat.

Mirna, M., Martanto, M., Dikananda, A. R., & Rifa'i, A. (2025). Algoritma K-Means Untuk Mengelompokan Bantuan Sosial Dalam Penentuan Strategi Distribusi Desa Sigong. Jurnal Informatika Teknologi dan Sains (Jinteks), 7(2), 482-491. https://doi.org/10.51401/jinteks.v7i2.5141

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830. http://scikit-learn.org

Putriana, P., Suarna, N., & Prihartono, W. (2023). Analisis Clustering Prestasi Atlet Pada Berbagai Cabang Olahraga Menggunakan Algoritma K-Means. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(6), 3435-3442. https://doi.org/10.36040/jati.v7i6.8211

Qibtiyah, M. (2024). Pemanfaatan Algoritma K-Means Clustering Dalam Penentuan Prioritas Penerima Program Bantuan Sosial Pendidikan. *JIK: Jurnal Informatika dan Komputer*, *15*(2), 68-76. https://journal.unmaha.ac.id/index.php/jik/article/view/386

Syani, M., & Wahyudi, T. (2024). Klasterisasi Penggunaan Ban dengan Cost Per Kilometer Terendah pada PT. PL menggunakan Metode K-Means. Jurnal Indonesia: Manajemen Informatika dan Komunikasi, 5(3), 2792-2800. https://doi.org/10.35870/jimik.v5i3.1005