



Department of Digital Business

Journal of Artificial Intelligence and Digital Business (RIGGS)

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 4 No. 2 (2025) pp: 4944-4949

P-ISSN: 2963-9298, e-ISSN: 2963-914X

Segmentasi Wajib Pajak Pribadi Berdasarkan Kode PTKP dan Penghasilan Neto Menggunakan Algoritma K-Means Clustering

Dimastito Prasetyo¹, Muhammad Rifkhan², Yoel Michael Sihombing³, Zurnan Alfian⁴

Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia

titodimas188@gmail.com; muhammadrifkan60@gmail.com; yoel9736@gmail.com; dosen02678@unpam.ac.id

Abstrak

Penerimaan pajak dari Wajib Pajak Orang Pribadi (WPOP) merupakan salah satu sumber pendapatan negara yang sangat penting dalam mendukung pembangunan nasional. Namun, karakteristik WPOP sangat beragam, baik dari segi penghasilan maupun status keluarga yang memengaruhi besaran Penghasilan Tidak Kena Pajak (PTKP). Keberagaman ini menuntut adanya strategi pengelolaan pajak yang berbasis data agar kebijakan perpajakan menjadi lebih tepat sasaran. Penelitian ini bertujuan untuk melakukan segmentasi terhadap WPOP berdasarkan kode PTKP dan penghasilan neto menggunakan algoritma K-Means Clustering. Penelitian menggunakan metode Knowledge Discovery in Database (KDD) yang terdiri dari lima tahapan: seleksi data, praproses data, transformasi data, pemodelan dengan algoritma K-Means, dan evaluasi hasil menggunakan Davies-Bouldin Index (DBI). Hasil pengelompokan menunjukkan adanya beberapa kluster wajib pajak dengan karakteristik serupa yang dapat menjadi dasar dalam merancang kebijakan perpajakan yang lebih adil, personal, dan efisien. Segmentasi ini diharapkan dapat membantu otoritas pajak dalam meningkatkan kepatuhan dan optimalisasi penerimaan pajak

Kata kunci: Wajib Pajak Pribadi, PTKP, Penghasilan Neto, K-Means Clustering, Segmentasi Data, Data Mining

1. Latar Belakang

Pajak merupakan salah satu instrumen vital dalam mendukung keberlanjutan pembangunan nasional. Sebagai sumber utama penerimaan negara, pajak berfungsi untuk membiayai berbagai kebutuhan negara seperti penyediaan infrastruktur, layanan kesehatan, pendidikan, subsidi sosial, serta penguatan sektor-sektor strategis lainnya. Seiring meningkatnya kebutuhan negara dalam membangun fondasi ekonomi yang kuat dan inklusif, maka optimalisasi penerimaan pajak menjadi suatu keniscayaan.

Salah satu komponen terbesar dalam penerimaan pajak adalah Pajak Penghasilan (PPH), khususnya dari kelompok Wajib Pajak Orang Pribadi (WPOP). Kontribusi kelompok ini sangat signifikan karena mencerminkan partisipasi individu dalam mendukung anggaran negara. Namun, dalam praktiknya, WPOP terdiri dari individu dengan karakteristik yang sangat beragam, baik dari segi jenis pekerjaan, tingkat penghasilan, status keluarga, maupun tingkat kepatuhan terhadap kewajiban perpajakan. Keragaman ini menuntut adanya strategi pengelolaan pajak yang lebih cerdas, responsif, dan berbasis data.

Dalam konteks penghitungan PPh, salah satu aspek yang sangat berpengaruh adalah keberadaan Penghasilan Tidak Kena Pajak (PTKP). PTKP merupakan bentuk keringanan pajak yang diberikan kepada wajib pajak berdasarkan status pernikahan dan jumlah tanggungan. Kode PTKP seperti TK/0, K/0, K/1, hingga K/3 merepresentasikan status dan kondisi keluarga yang berbeda. Hal ini penting karena kebijakan yang tidak mempertimbangkan variasi ini secara tepat berpotensi menimbulkan ketidakadilan dan menurunkan tingkat kepatuhan pajak.

Transformasi digital dan ketersediaan big data memberikan peluang besar untuk memanfaatkan pendekatan analitik canggih dalam memahami karakteristik wajib pajak. Salah satu pendekatan yang relevan adalah teknik segmentasi berbasis data mining, khususnya menggunakan algoritma K-Means Clustering. Metode ini dapat mengelompokkan wajib pajak berdasarkan kesamaan atribut seperti penghasilan neto dan kode PTKP, sehingga pola perilaku perpajakan dapat dikenali dengan lebih baik. Dengan demikian, pemerintah dapat menyusun strategi kebijakan pajak yang lebih personal, akurat, dan berkeadilan sosial.

Segmentasi Wajib Pajak Pribadi Berdasarkan Kode PTKP dan Penghasilan Neto Menggunakan Algoritma K-Means Clustering

Pemanfaatan algoritma K-Means dalam konteks segmentasi wajib pajak menjadi penting karena bersifat unsupervised learning, yang tidak memerlukan data berlabel, serta efisien dalam mengolah data berskala besar. Dengan mengidentifikasi kelompok-kelompok wajib pajak yang homogen, maka intervensi kebijakan seperti edukasi, sosialisasi, dan insentif fiskal dapat dilakukan secara lebih terarah. Penelitian ini bertujuan untuk mengeksplorasi penerapan algoritma K-Means dalam segmentasi Wajib Pajak Pribadi berdasarkan penghasilan neto dan kode PTKP, serta menganalisis hasil segmentasi sebagai masukan strategis bagi kebijakan perpajakan di Indonesia.

2. Metode Penelitian

Penelitian ini dilakukan dengan pendekatan kuantitatif menggunakan metode data mining untuk mengelompokkan Wajib Pajak Pribadi berdasarkan atribut kode PTKP dan penghasilan neto. Teknik yang digunakan dalam proses pengelompokan adalah algoritma K-Means Clustering, dengan tahapan yang mengikuti proses Knowledge Discovery in Database (KDD). Proses ini melibatkan lima langkah utama, yaitu seleksi data, praproses data, transformasi data, proses pengelompokan (clustering), dan evaluasi hasil.

2.1. Data Mining dan Knowledge Discovery in Database (KDD)

Data mining adalah proses penggalian informasi yang berguna dari kumpulan data besar yang sebelumnya tidak diketahui. Proses ini melibatkan tahapan-tahapan dalam Knowledge Discovery in Database (KDD), yakni seleksi data, pembersihan data, transformasi data, proses penambangan data, dan evaluasi hasil. KDD menjadi pendekatan yang sistematis untuk memperoleh pola dan wawasan baru dari data yang kompleks.

Pada tahap seleksi data, peneliti mengumpulkan data simulasi Wajib Pajak Pribadi yang terdiri dari informasi kode PTKP dan nilai penghasilan neto per tahun. Data ini disusun mengacu pada ketentuan resmi dari Direktorat Jenderal Pajak mengenai tarif dan klasifikasi PTKP, sebagaimana dijelaskan dalam buku Praktikum Perpajakan 1 oleh Saksono Budi dan Desi Kurniawati seperti Tabel 1.

Table 1. Data PTKP yang digunakan

| Keterangan | Kode PTKP | Besaran PTKP |
|---|-----------|-----------------|
| Tidak Kawin Tanpa Tanggungan | TK/0 | Rp. 54.000.000 |
| Tidak Kawin punya Tanggungan 1 | TK/1 | Rp. 58.500.000 |
| Tidak Kawin punya Tanggungan 2 | TK/2 | Rp. 63.000.000 |
| Tidak Kawin punya Tanggungan 3 | TK/3 | Rp. 67.500.000 |
| Kawin Tanpa Tanggungan | K/0 | Rp. 58.500.000 |
| Kawin punya tanggungan 1 | K/1 | Rp. 63.000.000 |
| Kawin punya tanggungan 2 | K/2 | Rp. 67.500.000 |
| Kawin punya tanggungan 3 | K/3 | Rp. 72.000.000 |
| Penghasilan istri digabung tanpa tanggungan | K/I/0 | Rp. 112.500.000 |
| Penghasilan istri digabung punya tanggungan 1 | K/I/1 | Rp. 117.000.000 |
| Penghasilan istri digabung punya tanggungan 2 | K/I/2 | Rp. 121.500.000 |
| Penghasilan istri digabung punya tanggungan 3 | K/I/3 | Rp. 126.000.000 |

Kode PTKP seperti TK/0, K/1, atau K/3 diubah ke dalam nilai nominal yang sesuai, misalnya TK/0 sebesar Rp 54.000.000 dan K/3 sebesar Rp 126.000.000. Sementara itu, data penghasilan neto mencerminkan jumlah pendapatan setelah dikurangi biaya-biaya dan PTKP.

Setelah data terkumpul, dilakukan praproses data untuk memastikan data bersih dan siap diolah. Langkah ini mencakup penghapusan nilai kosong, duplikat, atau kesalahan penulisan. Atribut yang tidak relevan seperti nama wajib pajak, alamat, atau informasi non-numerik lainnya dihapus untuk menjaga fokus pada dua variabel utama: kode PTKP (yang telah dikonversi menjadi nilai numerik) dan penghasilan neto. Pada tahap ini juga dilakukan pengkodean numerik terhadap data kategorikal agar dapat diproses oleh algoritma K-Means.

2.2. Algoritma K-Means Clustering

K-Means Clustering adalah metode klusterisasi berbasis partisi yang digunakan untuk mengelompokkan data ke dalam k kelompok berdasarkan jarak terdekat ke titik pusat (centroid). Algoritma ini bersifat unsupervised learning, artinya tidak memerlukan data berlabel dalam proses pelatihannya. Kelebihan utama K-Means terletak pada kecepatan dan efisiensi pemrosesan data dalam jumlah besar, serta kemampuan dalam mengidentifikasi kelompok data dengan karakteristik yang mirip. Pada algoritma *K-Means clustering* data mentah diubah menjadi bentuk yang sesuai untuk proses clustering. Misalnya, kode PTKP yang semula berbentuk teks dikonversi menjadi angka yang merepresentasikan nilai PTKP. Proses ini juga memastikan bahwa seluruh variabel yang digunakan berada dalam skala yang sama atau setara, sehingga perhitungan jarak antar data dalam algoritma K-Means tidak bias terhadap variabel yang memiliki skala lebih besar.

2.3. Penghasilan Tidak Kena Pajak (PTKP) dan Penghasilan Neto

PTKP adalah komponen pengurang dalam perhitungan Pajak Penghasilan, yang besarnya disesuaikan dengan status keluarga dan jumlah tanggungan. Status PTKP dikelompokkan ke dalam kode seperti TK/0, TK/1, K/0, K/1, K/3, dan sebagainya. Semakin tinggi jumlah tanggungan, semakin besar pengurangan pajak yang diberikan. Sementara itu, penghasilan neto adalah penghasilan setelah dikurangi biaya dan PTKP, yang menjadi dasar perhitungan pajak terutang. Kombinasi antara PTKP dan penghasilan neto memberikan gambaran menyeluruh tentang profil ekonomis Wajib Pajak.

2.4. Transformasi data

Transformasi data, yaitu mengubah data mentah menjadi bentuk yang sesuai untuk proses clustering. Misalnya, kode PTKP yang semula berbentuk teks dikonversi menjadi angka yang merepresentasikan nilai PTKP. Proses ini juga memastikan bahwa seluruh variabel yang digunakan berada dalam skala yang sama atau setara, sehingga perhitungan jarak antar data dalam algoritma K-Means tidak bias terhadap variabel yang memiliki skala lebih besar.

2.5. Pengelompokan data

Pengelompokan data Proses dimulai dengan menentukan jumlah kluster (k) yang diinginkan, kemudian algoritma secara acak menetapkan pusat awal (centroid) untuk setiap kluster. Setiap data kemudian dihitung jaraknya ke setiap centroid, dan dikelompokkan ke kluster dengan jarak terdekat. Setelah seluruh data diklasifikasikan, posisi centroid diperbarui dengan menghitung rata-rata nilai setiap kluster. Proses ini diulang hingga tidak ada lagi perubahan signifikan dalam pengelompokan atau hingga iterasi maksimal tercapai.

2.6. Pemrosesan data

Pemrosesan data dan implementasi algoritma dilakukan menggunakan perangkat lunak RapidMiner. Tools ini dipilih karena menyediakan antarmuka visual yang mendukung operator-operator seperti Read CSV untuk membaca data, Select Attribute untuk pemilihan atribut, Nominal to Numerical untuk transformasi variabel kategorikal, Clustering (K-Means) untuk pemodelan, dan Performance (Cluster Distance Performance) untuk evaluasi kinerja model.

2.7. Evaluasi

Evaluasi hasil, yang bertujuan untuk menilai kualitas dari hasil clustering. Evaluasi dilakukan menggunakan Davies-Bouldin Index (DBI), yaitu suatu metrik yang mengukur tingkat kemiripan antar kluster. Semakin kecil nilai DBI yang diperoleh, semakin baik hasil segmentasi yang terbentuk. Dalam penelitian ini, dilakukan pengujian terhadap beberapa nilai k (jumlah kluster) untuk mencari nilai DBI paling optimal. Hasil terbaik adalah yang mendekati nol, karena mencerminkan bahwa kluster yang terbentuk memiliki perbedaan yang jelas satu sama lain dan kesamaan yang tinggi di dalam masing-masing kluster.

Dengan menggunakan tahapan KDD secara menyeluruh dan pendekatan algoritma K-Means Clustering, penelitian ini bertujuan untuk memberikan segmentasi yang jelas terhadap Wajib Pajak Pribadi berdasarkan penghasilan neto dan status PTKP. Hasil segmentasi diharapkan dapat menjadi dasar pertimbangan kebijakan yang lebih tepat sasaran dalam bidang perpajakan.

3. Hasil dan Pembahasan

Hasil penelitian didasarkan pada urutan logis untuk membentuk sebuah cerita. Isinya menunjukkan fakta/data. Dapat menggunakan Tabel dan Angka tetapi tidak mengulangi data yang sama dalam gambar, tabel, dan teks. Untuk lebih memperjelas deskripsi, dapat menggunakan subtitle.

Diskusi adalah penjelasan dasar, hubungan, dan generalisasi yang ditunjukkan oleh hasilnya. Deskripsi menjawab pertanyaan penelitian. Jika ada hasil yang meragukan, tunjukkan secara objektif.

3.1. Seleksi dan Karakteristik Data

Dataset yang digunakan dalam penelitian ini merupakan data simulasi yang dibuat berdasarkan acuan buku Praktikum Perpajakan 1 dan ketentuan peraturan Direktorat Jenderal Pajak (DJP). Data ini mewakili kondisi umum Wajib Pajak Orang Pribadi (WPOP) di Indonesia, dengan mempertimbangkan status perkawinan, jumlah tanggungan, dan besaran penghasilan. Variabel utama dalam penelitian ini yaitu: Kode PTKP dimana kode menggambarkan status keluarga dan jumlah tanggungan WPOP, seperti TK/0 (tidak kawin, tanpa tanggungan), K/3 (kawin, 3 tanggungan), dan KI/3 (penghasilan digabung dengan istri, 3 tanggungan), Penghasilan Bruto dimana total pendapatan sebelum pengurangan PTKP, dan Penghasilan Neto dimana penghasilan yang menjadi dasar perhitungan PPh setelah dikurangi PTKP. Data awal terdiri dari 150 entri dengan variasi yang mencerminkan populasi WPOP berdasarkan kategori PTKP. Data ini dirancang untuk memudahkan proses eksplorasi dan analisis menggunakan algoritma K-Means Clustering.

3.2. Preprocessing Data

Dalam data science, ada dua aspek penting yang perlu diperhatikan yaitu pre-processing dan classification. Pre-processing adalah teknik untuk menyiapkan data agar lebih siap untuk dilakukan lebih lanjut dalam rangka ekstraksi pengetahuan. Classification merupakan suatu pendekatan dalam data science yang juga dikenal dengan supervised algorithm. Tentu tipe data pada teknik classification memerlukan data label. Data Preprocessing merupakan salah satu tahapan dalam melakukan mining data. Sebelum menuju ke tahap pemrosesan. Data mentah akan diolah terlebih dahulu. Data Preprocessing atau praproses data biasanya dilakukan melalui cara eliminasi data yang tidak sesuai. Selain itu dalam proses ini data akan diubah dalam bentuk yang akan lebih dipahami oleh sistem. Tahapan ini dilakukan untuk memastikan data bersih, terstruktur, dan siap digunakan dalam pemodelan algoritma. Tahapan yang dilakukan yaitu : Cleaning dimana menghapus data yang tidak lengkap, duplikat, serta nilai ekstrem (outlier) dengan pendekatan Z-score lebih dari 3, melakukan transformasi Kode PTKP dimana kode PTKP dikonversi menjadi nilai numerik menggunakan teknik label encoding agar dapat diproses oleh algoritma K-Means. Contoh konversi: TK/0 = 0, TK/1 = 1, ..., KI/3 = 11 dan melakukan normalisasi Penghasilan Neto dimana dilakukan dengan metode Min-Max Scaling agar semua nilai berada dalam rentang 0-1, sehingga mencegah dominasi fitur berangka besar dalam perhitungan jarak Euclidean.

3.3 Penentuan Jumlah Kluster Optimal

Untuk menentukan nilai k yang optimal, digunakan dua pendekatan:

- Elbow Method: Metode Elbow (atau metode siku) adalah sebuah metode dalam clustering (pengelompokan data) yang digunakan untuk menentukan jumlah kluster optimal, terutama dalam algoritma K-Means. Metode ini bekerja dengan cara memvisualisasikan hubungan antara jumlah kluster (K) dan Within-Cluster Sum of Squares (WCSS), kemudian mencari titik "siku" pada grafik tersebut, yang mengindikasikan jumlah kluster optimal. Tujuan utama dari metode Elbow adalah untuk menemukan titik pada grafik WCSS vs. jumlah kluster (K) di mana penambahan jumlah kluster lebih lanjut tidak lagi memberikan penurunan yang signifikan pada WCSS. Menggunakan grafik WCSS (Within-Cluster Sum of Squares). Grafik menunjukkan penurunan tajam hingga k = 3, kemudian menurun perlahan. Titik ini merupakan elbow point yang menandakan jumlah kluster optimal.
- Silhouette Score: Silhouette Score adalah salah satu alat yang digunakan untuk mengevaluasi seberapa baik setiap objek dikelompokkan ke dalam kluster yang sesuai. Metode ini memberikan ukuran tentang seberapa dekat setiap titik data dengan kluster yang ditempatinya, dibandingkan dengan kluster lainnya.

Silhouette Score mengukur seberapa baik setiap objek dalam sebuah kluster dan seberapa berbedanya kluster tersebut dibandingkan dengan kluster lain. Nilai Silhouette Score berkisar antara -1 hingga 1. Nilai positif menunjukkan bahwa objek berada di kluster yang tepat, sedangkan nilai negatif menandakan bahwa objek mungkin ditempatkan di kluster yang salah.

Mengukur tingkat kepadatan dan pemisahan antar kluster. Pada $k = 3$ diperoleh nilai $> 0,6$, yang berarti kluster sudah cukup kompak dan berbeda satu sama lain.

3.4. Implementasi Algoritma K-Means

K-means diterapkan melalui:

- $k = 3$: Berdasarkan hasil evaluasi sebelumnya.
- `init = 'k-means++'`: Untuk mengoptimalkan posisi awal centroid.
- `max_iter = 300`: Iterasi maksimum.
- `n_init = 10`: Untuk menghindari local minima.

Hasil klusterisasi memperlihatkan bahwa data terbagi menjadi tiga segmen sebagai berikut:

| <i>Klaster</i> | <i>Kode PTKP Dominan</i> | <i>Rata-rata Penghasilan Neto</i> | <i>Jumlah WP</i> | <i>Interpretasi</i> |
|----------------|--------------------------|-----------------------------------|------------------|------------------------|
| 0 | TK/0, TK/1 | < Rp 60 juta | 65 | Low Income / Pemula |
| 1 | K/2, K/3 | Rp 80–120 juta | 43 | Menengah / Stabil |
| 2 | KI/2, KI/3 | > Rp 120 juta | 42 | High Income / Gabungan |

3.5. Analisis dan Interpretasi Tiap Kluster

Klaster 0 (Low Income): WPOP yang tidak menikah dan/atau tidak memiliki tanggungan, penghasilan neto relatif rendah. Mereka cenderung masih dalam tahap awal karier atau bekerja di sektor informal. Sangat cocok menjadi sasaran program penyuluhan perpajakan dan edukasi literasi fiskal.

- Klaster 1 (Menengah): WPOP dengan status kawin dan tanggungan 1–2 orang, penghasilan berada di kategori menengah. Mereka merupakan kelompok stabil secara finansial dan membutuhkan kemudahan dalam layanan pajak digital seperti e-Filing.
- Klaster 2 (High Income): WPOP yang menggabungkan penghasilan dengan istri (KI) dan memiliki tanggungan tinggi. Mereka merupakan kelompok berpenghasilan besar, wajib pajak potensial untuk target intensifikasi pengawasan dan audit pajak.

3.6. Evaluasi dan Validasi Model

Evaluasi dilakukan dengan WCSS (Within-Cluster Sum of Squares) dimana mengukur total variasi dalam satu kluster. Nilai WCSS cenderung kecil saat data dalam satu kluster homogen, dan Silhouette Score: Nilai > 0.6 menandakan bahwa kluster memiliki batas pemisah yang jelas dan jarak antar kluster signifikan.

4. Kesimpulan

Penerapan K-Means efektif dalam mengelompokkan WPOP ke dalam tiga kluster yang merepresentasikan

karakteristik ekonomi yang berbeda. Klaster 0 terdiri dari WP dengan penghasilan rendah dan status belum menikah, Klaster 1 mewakili kelompok menengah dengan status kawin dan tanggungan sedang, serta Klaster 2 mencakup WP berpenghasilan tinggi dengan penggabungan penghasilan bersama pasangan. Pemilihan atribut kode PTKP dan penghasilan neto terbukti mampu menjadi indikator representatif dalam memahami profil fiskal WP. Kode PTKP menunjukkan status keluarga dan jumlah tanggungan, sedangkan penghasilan neto menggambarkan kapasitas ekonomi sesungguhnya. Evaluasi model menunjukkan performa yang baik, dengan nilai Silhouette Score sebesar 0,64 dan hasil Elbow Method mengindikasikan pemisahan klaster yang optimal pada $k = 3$. Segmentasi ini dapat dimanfaatkan untuk menyusun kebijakan pajak yang lebih personal dan adil, seperti edukasi pajak untuk WP pemula, kemudahan layanan digital untuk WP menengah, serta pengawasan intensif terhadap WP berpenghasilan tinggi. Penelitian ini sejalan dengan studi terdahulu seperti yang dilakukan oleh Rini Nursaniah et al. (2024) dalam konteks PBB, yang menegaskan efektivitas K-Means dalam membentuk klaster berbasis data pajak untuk kepentingan strategis fiskal. Direktorat Jenderal Pajak disarankan menggunakan pendekatan data mining seperti K-Means untuk mendukung kebijakan berbasis segmentasi. Penggunaan algoritma ini dapat meningkatkan efektivitas edukasi, pemeriksaan, dan insentif pajak. Penelitian selanjutnya dapat mengembangkan model dengan menambahkan variabel baru, seperti jenis pekerjaan, usia, domisili, dan riwayat pelaporan SPT agar hasil segmentasi lebih komprehensif. Perlu dilakukan validasi model pada data wajib pajak riil dari DJP agar segmentasi benar-benar mencerminkan kondisi wajib pajak yang sesungguhnya dan mendukung pengambilan kebijakan yang tepat. Implementasi hasil segmentasi dapat dilakukan melalui aplikasi digital atau dashboard perpajakan yang memberikan rekomendasi atau layanan berbasis karakter klaster wajib pajak. Sinergi antara akademisi, praktisi, dan instansi pemerintah perlu ditingkatkan untuk mengembangkan model-model prediktif dan analitik lainnya demi memperkuat basis data perpajakan nasional. Dengan demikian, penelitian ini diharapkan menjadi kontribusi awal dalam mendorong transformasi digital perpajakan yang berbasis analisis data dan mampu menjangkau seluruh lapisan wajib pajak secara adil dan efisien.

Referensi

1. F. Primadeni, A. I. Pradana, and E. Purwanto, "Analisis Clustering Untuk Segmentasi Wilayah Berdasarkan Karakteristik PBB di Kabupaten Sragen," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 1, no. 3, pp. 45–56, Juli 2024.
<https://doi.org/10.52436/1.jpti.500>
2. E. Prayitno, I. J. Perdana, E. Iskandar, B. H. Winarno, and A. A. Subagyo, "Optimalisasi Profitabilitas Ritel Melalui Segmentasi Pelanggan dengan K-Means Clustering," *Informasi Interaktif: Jurnal Informatika dan Teknologi Informasi*, vol. 9, no. 3, pp. 107–116, November 2024.
<https://doi.org/10.37159/jii.v9i3.107>
3. E. Worabai, A. H. M. Muhammad, and T. Hidayat, "Implementasi Metode Cluster Analysis K-Means dalam Segmentasi," *Jurnal Ilmiah Komputasi*, vol. 22, no. 3, pp. 441–447, Oktober 2023.
<https://doi.org/10.32409/jikstik.22.3.3493>
4. K. Auliasari and M. Kertaningtyas, "Penerapan Algoritma K-Means untuk Segmentasi Konsumen Menggunakan R," *Jurnal Teknologi dan Manajemen Informatika*, vol. 5, no. 2, pp. 34–41, 2023.
<https://doi.org/10.26905/jtmi.v5i2.3644>
5. D. Irawan, G. Wijaya, and T. T. Warisaji, "Penerapan Algoritma K-Means Clustering untuk Segmentasi Nasabah Bank," *BIOS: Jurnal Teknologi Informasi dan Rekayasa Komputer*, vol. 6, no. 1, pp. 162–170, 2023.
<https://doi.org/10.37148/bios.v6i1.162>
6. A. Hadi, "Segmentasi Pelanggan Internet Service Provider (ISP) Berbasis Pillar K-Means," *Jurnal Ilmiah Teknologi Informasi Asia*, vol. 13, no. 2, pp. 413–422, 2023.
<https://doi.org/10.32815/jitika.v13i2.413>
7. E. Tjatur Puteri, G. Kusnanto, and C. J. Thomas, "Penerapan K-Means Clustering untuk Segmentasi Pelanggan pada CRM di PT. Unichem Candi Indonesia," *Konvergensi*, vol. 15, no. 2, pp. 51–60, 2024.
<https://doi.org/10.30996/konv.v15i2.3651>