

Klasifikasi Profil Wajib Pajak Pribadi Melalui Variabel PTKP dan Penghasilan Neto Menggunakan Pendekatan K-Means Clustering

Muhamad Royhan Adriansyah¹, Rizal Jafar Sidiq², M. Zahavy Al Zenita³, Reva Zaidan Azmi⁴, Fauzan Hanif Ilyasa⁵, Fitri Yanti⁶

^{1,2,3,4,5,6}Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia

Email: ¹royhanadriansyah@gmail.com, ²rizal142@admin.smk.belajar.id, ³zahavyalzenita@gmail.com,
⁴revazaidanazmi@gmail.com, ⁵fauzanhnf0008@gmail.com, ⁶dosen00848@unpam.ac.id

Abstrak—Kontribusi pajak dari Wajib Pajak Orang Pribadi (WPOP) memiliki peran krusial sebagai sumber pendapatan negara untuk membiayai pembangunan. Tantangannya, karakteristik WPOP sangat beragam, baik dari tingkat penghasilan maupun status keluarga yang langsung mempengaruhi besaran Penghasilan Tidak Kena Pajak (PTKP). Keragaman ini mengharuskan adanya pendekatan pengelolaan yang didorong oleh data. Oleh karena itu, penelitian ini dilakukan untuk mengelompokkan WPOP ke dalam segmen-segmen berdasarkan kode PTKP dan penghasilan neto memanfaatkan algoritma K-Means Clustering. Kerangka kerja KDD dijalankan mulai dari pengumpulan data, pembersihan, transformasi, pemodelan, hingga penilaian cluster dengan Davies-Bouldin Index (DBI). Berdasarkan analisis, teridentifikasi sejumlah kluster wajib pajak dengan profil serupa. Temuan ini berpotensi menjadi landasan bagi kebijakan yang lebih tepat sasaran, yang diharapkan dapat meningkatkan kepatuhan wajib pajak dan memaksimalkan pendapatan negara dari sektor pajak..

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

Abstract—Tax contributions from Individual Taxpayers (WPOP) play a crucial role as a source of state revenue to fund development. The challenge lies in the highly diverse characteristics of WPOP, both in terms of income levels and family status, which directly affect the amount of Non-Taxable Income (PTKP). This diversity necessitates a data-driven management approach. Therefore, this research was conducted to segment WPOP into groups based on PTKP codes and net income by utilizing the K-Means Clustering algorithm. The KDD framework was executed, starting from data collection, cleaning, and transformation, to modeling and cluster evaluation using the Davies-Bouldin Index (DBI). The analysis identified a number of taxpayer clusters with similar profiles. This finding has the potential to serve as a foundation for more targeted policies, which are expected to improve taxpayer compliance and maximize state revenue from the tax sector.

Keywords: Individual Taxpayers, PTKP, Net Income, K-Means Clustering, Data Segmentation

1. PENDAHULUAN

Pajak merupakan komponen fundamental yang menjamin pembangunan nasional dapat berlangsung secara berkelanjutan. Sumbangsihnya sebagai sumber penerimaan negara terbesar menjadikannya tulang punggung bagi pembiayaan berbagai program prioritas, seperti pembangunan infrastruktur, penyediaan layanan kesehatan dan pendidikan, penyaluran bantuan sosial, serta penguatan sektor-sektor kunci. Dalam konteks upaya membangun pondasi ekonomi yang kokoh dan inklusif, peningkatan kontribusi dari sektor perpajakan merupakan suatu hal yang tidak terelakkan.

Di antara berbagai jenis penerimaan pajak, kontribusi dari Pajak Penghasilan (PPh) Wajib Pajak Orang Pribadi (WPOP) memegang peranan yang sangat strategis. Signifikansi WPOP terletak pada kemampuannya merepresentasikan partisipasi langsung masyarakat dalam mendanai Anggaran Pendapatan dan Belanja Negara (APBN). Namun, realitasnya, populasi WPOP sangatlah heterogen, dengan variasi yang kompleks mulai dari profesi, besaran penghasilan, kondisi keluarga, hingga tingkat kesadaran dalam memenuhi kewajiban perpajakannya. Kondisi yang beragam ini menuntut pendekatan pengelolaan perpajakan yang lebih cerdas, adaptif, dan berbasis pada analisis data.

Dalam perhitungan PPh, salah satu faktor yang sangat berpengaruh adalah adanya Penghasilan Tidak Kena Pajak (PTKP). PTKP adalah jenis pengurangan pajak yang diberikan kepada pembayar pajak berdasarkan status pernikahan dan jumlah tanggungannya. Kode PTKP seperti TK/0, K/0, K/1, sampai K/3 melambangkan kondisi dan status keluarga yang beragam. Hal ini krusial karena kebijakan yang tidak memperhatikan variasi ini secara akurat dapat menyebabkan ketidakadilan dan mengurangi tingkat kepatuhan pajak.

Digitalisasi dan akses terhadap big data menciptakan peluang signifikan untuk memanfaatkan strategi analisis mendalam dalam mengenali ciri-ciri wajib pajak. Salah satu metode yang relevan ialah teknik segmentation yang didasarkan pada data mining, terutama dengan memanfaatkan algoritma K-Means Clustering. Cara ini membuat kelompok wajib pajak berdasarkan klasifikasi karakteristik seperti penghasilan bersih dan kode PTKP, sehingga perilaku perpajakan dapat dikenali dengan lebih jelas. Sehingga, pemerintah dapat merancang strategi. Untuk kebijakan pajak yang lebih individu, tepat, dan adil secara sosial.

2. METODOLOGI

Penelitian ini mengadopsi pendekatan kuantitatif melalui metode data mining untuk mengelompokkan Wajib Pajak Pribadi berdasarkan karakteristik kode PTKP dan penghasilan bersih. Algoritma K-Means Clustering diterapkan dengan mengikuti kerangka kerja Knowledge Discovery in Databases (KDD), yang meliputi lima tahap utama: seleksi data, pra-pemrosesan data, transformasi data, proses klustering, dan evaluasi hasil.

2.1 Data Mining dan KDD

Data Mining merupakan proses inti dalam Knowledge Discovery in Databases (KDD) untuk mengungkap informasi berharga dari data dalam skala besar. Melalui tahapan KDD yang terstruktur—mulai dari seleksi, pembersihan, transformasi data, hingga proses penambangan dan evaluasi—pola dan wawasan baru dapat diekstraksi dari kumpulan data yang kompleks. Dalam penelitian ini, tahap seleksi data melibatkan pengumpulan data simulasi Wajib Pajak Pribadi yang berisi variabel kode PTKP dan penghasilan bersih tahunan, yang disusun mengacu pada regulasi Direktorat Jenderal Pajak dan buku "Praktikum Perpajakan 1".

Tabel 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. 54.500.000
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 tanpa Tanggungan 1	K/I/1	Rp. 117.000.000
Penghasilan istri digabung tanpa Tanggungan 2	K/I/2	Rp. 121.500.000
Penghasilan istri digabung tanpa 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 pengumpulan data selesai, langkah berikutnya adalah mempersiapkan data agar bersih dan siap untuk dianalisis. Proses ini mencakup penghapusan nilai yang hilang, entri ganda, atau kesalahan penulisan. Elemen-elemen yang tidak berhubungan, seperti nama wajib pajak, alamat, atau data non-numerik lainnya, akan dihilangkan untuk menjaga konsentrasi pada dua variabel inti: kode PTKP (yang telah diubah menjadi angka) dan pendapatan bersih. Pada fase ini, juga dilakukan pengkodean angka pada data kategori agar dapat diolah oleh algoritma K-Means.

2.2 Algoritma K-Means

K-Means Clustering adalah suatu algoritma *unsupervised learning* yang mengelompokkan data ke dalam sejumlah kluster (k) berdasarkan kedekatannya dengan titik pusat (*centroid*). Keunggulan utamanya terletak pada kecepatan dan efisiensinya dalam menangani dataset berskala besar. Sebelum diproses, data mentah seperti kode PTKP yang berbentuk teks harus ditransformasi menjadi numerik dan dinormalisasi. Hal ini memastikan semua variabel memiliki skala yang seragam, sehingga perhitungan jarak antar data dalam algoritma tidak bias oleh perbedaan besaran fitur.

2.3 PTKP dan Penghasilan Neto

PTKP berfungsi sebagai komponen pengurang dalam penghitungan Pajak Penghasilan yang nilainya ditentukan berdasarkan status perkawinan dan jumlah tanggungan keluarga. Kode-kode seperti TK/0, K/1, hingga K/3 merepresentasikan variasi status ini, di mana nilai pengurangan pajak meningkat seiring dengan bertambahnya tanggungan. Di sisi lain, penghasilan neto merupakan dasar pengenaan pajak yang dihitung setelah dikurangi berbagai biaya dan PTKP. Kombinasi antara besaran PTKP dan penghasilan neto ini membentuk profil ekonomi komprehensif dari seorang Wajib Pajak.

2.4 Perubahan pada data



Tahap transformasi data, untuk mengubah data dari mentah hingga menjadi bentuk yang siap untuk pengelompokan. Misalnya, kode pada PTKP yang awalnya berbentuk teks (seperti "K/1") diubah menjadi angka. Proses ini juga menyesuaikan skala semua variabel agar seimbang, sehingga tidak ada variabel yang mendominasi perhitungan jarak dalam algoritma K-Means hanya karena memiliki angka yang lebih besar.

2.5 Pengelompokan

Mekanisme pada pengelompokan data diawali dengan penetapan pada jumlah kluster yang diinginkan (k). Pada tahap inisialisasi, algoritma memilih titik pusat (centroid) awal secara acak. Kemudian, setiap data yang ada akan dihitung jarak ke semua centroid dan digolongkan ke dalam kluster yang centroidnya paling dekat. Begitu semua data telah terdistribusi, posisi setiap centroid dihitung ulang berdasarkan rata-rata dari data yang menjadi anggotanya. Proses ini terus berulang hingga tidak ada lagi pergeseran signifikan dalam komposisi kluster atau hingga batas pada iterasi terpenuhi.

2.6 Pemrosesan Pada data

Seluruh pemrosesan data dan implementasi algoritma dalam penelitian ini dilakukan dengan bantuan perangkat lunak RapidMiner. Pemilihan perangkat lunak ini didasari oleh antarmuka grafisnya yang memudahkan penerapan berbagai fungsi, seperti pembacaan data (*Read CSV*), seleksi atribut, konversi data kategorikal ke numerik (*Nominal to Numerical*), pemodelan klaster dengan K-Means, serta evaluasi kinerja model menggunakan *Cluster Distance Performance*.

2.7 Penilaian

Evaluasi hasil klusterisasi dilakukan untuk menilai efektivitas pengelompokan yang dihasilkan. Pengukuran kualitas menggunakan Davies-Bouldin Index (DBI) sebagai metrik evaluasi, dengan prinsip bahwa nilai DBI yang rendah mengindikasikan performa segmentasi yang lebih unggul. Penelitian ini menguji berbagai variasi nilai k untuk mengidentifikasi konfigurasi optimal, dimana nilai DBI terendah (mendekati nol) mencerminkan pemisahan kluster yang jelas dan homogenitas internal yang tinggi.

3. PEMBAHASAN DAN HASIL

Hasil dari penelitian ini berlandaskan pada alur logis yang dimaksudkan untuk merangkai sebuah narasi. Dalam hal ini, terdapat penyajian fakta atau data yang relevan. Dapat memasukkan tabel dan angka, tetapi tidak diizinkan untuk mengulang data yang identik dalam tabel, dan teks. Agar deskripsi menjadi lebih jelas, subtitle bisa dipakai.

Diskusi mencakup penjelasan mendasar, keterkaitan, serta generalisasi yang dihasilkan dari data. Deskripsi akan memberikan jawaban atas pertanyaan penelitian. Apabila terdapat hasil yang diragukan, hal tersebut perlu ditunjukkan objektif.

3.1 Seleksi dan Karakteristik Data

Sumber data dalam penelitian ini adalah data simulasi yang dikonstruksi dengan berpedoman pada literatur perpajakan (*Praktikum Perpajakan 1*) dan peraturan resmi Direktorat Jenderal Pajak (DJP). Dataset ini dirancang untuk merefleksikan karakteristik umum

Wajib Pajak Orang Pribadi (WPOP) di Indonesia, dengan mempertimbangkan variabel-variabel kunci seperti status pernikahan, jumlah tanggungan, dan tingkat pendapatan. Variabel utama yang digunakan meliputi Kode PTKP (sebagai indikator status keluarga), Penghasilan Bruto, dan Penghasilan Neto. Dengan cakupan 150 entri yang beragam, data ini memadai untuk digunakan dalam analisis eksploratori melalui teknik K-Means Clustering.

3.2 Preprocessing Data Cluster

Proses analisis data diawali dengan pra-pemrosesan, sebuah metode preparasi untuk mentransformasi data mentah menjadi set data yang terstruktur dan siap untuk ekstraksi wawasan. Langkah ini sangat penting untuk memastikan kualitas data sebelum memasuki tahap pemodelan. Beberapa teknik yang diterapkan antara lain:

- a. **Pembersihan Data**, dengan menghilangkan duplikat dan outlier (menggunakan $Z\text{-score} > 3$).
- b. **Transformasi Kode Kategorikal**, di mana variabel seperti Kode PTKP di-encode secara numerik (contoh: TK/0=0, ..., KI/3=11) via Label Encoding.
- c. **Normalisasi**, dengan menerapkan Min-Max Scaling pada variabel Penghasilan Neto untuk menyamakan skalanya ke rentang $[0,1]$, sehingga perhitungan jarak Euclidean pada algoritma K-Means tidak didominasi oleh variabel berskala besar.

Setelah melalui tahap pra-pemrosesan, data kemudian diproses menggunakan teknik klasifikasi. Klasifikasi merupakan bagian dari algoritma *supervised learning* yang mensyaratkan data berlabel untuk pelatihan model. Penerapannya memungkinkan pengelompokan atau prediksi data ke dalam kategori-

kategori yang telah ditentukan.

3.3 Penetapan Jumlah Kluster

Untuk menentukan nilai K yang optimal, digunakan dua pendekatan :

- Metode Elbow, atau yang sering disebut metode siku, merupakan suatu pendekatan dalam analisis kluster yang bertujuan menemukan jumlah kelompok terbaik untuk algoritma K- Means. Cara kerjanya adalah dengan memvisualisasikan hubungan antara jumlah kluster
- (K) dengan variasi dalam kluster (WCSS). Titik "siku" pada grafik, di mana penurunan WCSS mulai melambat secara signifikan, menunjukkan jumlah kluster yang optimal. Pada penelitian ini, grafik menunjukkan penurunan tajam WCSS hingga K=3, yang kemudian menjadi titik siku dan menandakan tiga kluster sebagai pembagian yang paling efektif..
- Silhouette Score: Silhouette Score merupakan salah satu metode untuk menilai sejauh mana setiap entitas tergabung dalam kluster yang tepat. Teknik ini memberikan indikasi mengenai seberapa dekat setiap titik data berhubungan dengan kluster yang ditentukannya, jika dibandingkan dengan kluster lain yang ada. Silhouette Score menilai seberapa optimal posisi tiap objek dalam suatu kluster dan seberapa besar perbedaan kluster tersebut dengan kluster lain yang ada. Rentang nilai Silhouette Score adalah dari -1 hingga 1. Nilai yang positif menunjukkan bahwa objek telah ditempatkan dalam kluster yang benar, sedangkan nilai negatif menunjukkan kemungkinan objek berada di kluster yang tidak tepat. Mengevaluasi tingkat kepadatan serta pemisahan antar kluster. Pada k = 3, diperoleh nilai lebih dari 0,6, yang menunjukkan bahwa kluster sudah cukup padat dan berbeda satu sama lain.

Tabel 2. Penentuan Cluster

No.	Nama Cluster	Karakteristik Utama	Ukuran Cluster (Estimasi)	Persona Wajib Pajak	Implikasi & Rekomendasi Tindakan
1	Cluster 1: Potensi Berkembang	<ul style="list-style-type: none">Pendapatan Neto : Rendah - Menengah (misal, Rp 70 – 150 jt/tahun)Profil PTKP Dominan: TK/0, TK/1 (Lajang, belum/sedikit tanggungan)	Besar (misal, 40% dari WP)	Pegawai baru, profesional muda, fresh graduate, pelaku gig economy	Implikasi: Kelompok terbesar yang butuh edukasi dasar perpajakan. Tindakan: Kampanye digital tentang cara lapor SPT, sosialisasi insentif dasar, e-Filing.
2	Cluster 2: Keluarga Mapan	<ul style="list-style-type: none">Pendapatan Neto: Menengah – Tinggi (misal, Rp 200 – 600 jt/tahun)Profil PTKP Dominan: K/1, K/2, K/3 (Menikah, memiliki tanggungan)	Signifikan (misal, 35% dari WP)	Karyawan senior, manajer, PNS, pemilik UMKM yang sudah mapan.	Implikasi: Tulang punggung penerimaan pajak dengan kepatuhan relatif baik. Tindakan: Layanan pengingat jatuh tempo, sosialisasi perencanaan pajak keluarga (biaya pendidikan, kesehatan), pelayanan prioritas.
3	Cluster 3: Pendapatan Tinggi	<ul style="list-style-type: none">Pendapatan Neto: Sangat Tinggi (misal, Rp 1 Miliar/tahun)Profil PTKP Dominan: Bervariasi	Kecil (misal, 5% dari WP)	Pengusaha besar, direktur, eksekutif C-level, profesional spesialis (dokter, pengacara).	Implikasi: Kontributor pajak terbesar; potensi tax planning yang kompleks dan risiko kepatuhan tinggi. Tindakan: Pengawasan intensif, dedicated account representative, sosialisasi aturan pajak lanjutan.
4	Cluster 4: Berisiko / Anomali	<ul style="list-style-type: none">Pendapatan Neto: Sangat Rendah (di bawah PTKP atau sedikit di atas)Profil PTKP Dominan: Kompleks (misal, K/3)	Sangat Kecil (misal, 2% dari WP)	Pensiunan, pekerja sektor informal, atau potensi under-reporting.	Implikasi: Adanya ketidaksesuaian antara profil tanggungan dan pendapatan yang dilaporkan. Tindakan: Validasi dan verifikasi data, program penyuluhan khusus,

					pemeriksaan acak.
--	--	--	--	--	-------------------

3.4 Implementasi Clustering

K-means akan diterapkan melewati :

- a. $k = 3$: Berdasarkan hasil evaluasi sebelumnya.
- b. $\text{init} = \text{'k-means++'}$: Untuk optimalkan posisi awal centroid.
- c. $\text{max_iter} = 300$: Iterasi maksimum.
- d. $\text{n_init} = 10$: Untuk menghindari local minima.

Hasil klasterisasi memperlihatkan bahwa data terbagi menjadi tiga segmen yaitu sebagai berikut:

Tabel 3. Hasil klasterisasi

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

3.5 Analisis Cluster

Segmentasi ini menghasilkan tiga profil WPOP yang berbeda. Klaster 0 mengidentifikasi WPOP berpenghasilan rendah yang umumnya belum memiliki tanggungan keluarga dan berada pada fase awal karier, sehingga intervensi yang tepat adalah edukasi perpajakan. Klaster 1 merepresentasikan WPOP dengan profil menengah yang stabil, sehingga kebutuhan utamanya adalah fasilitas layanan digital yang efisien. Adapun Klaster 2 merupakan kelompok berpenghasilan tinggi dengan kompleksitas transaksi yang lebih besar, sehingga menjadi subjek yang strategis untuk kegiatan pengawasan dan audit guna menjaga kepatuhan.

3.6 Tahap Evaluasi

Proses evaluasi menggunakan dua pengukuran. Pertama, WCSS (Within-Cluster Sum of Squares) menghitung variasi internal dalam kluster, di mana nilai yang kecil menandakan kluster yang padat dan homogen. Kedua, Silhouette Score digunakan untuk menilai kualitas pemisahan antar kluster. Skor di atas 0,6 menandakan bahwa batas-batas antar kluster telah terdefinisi dengan baik dan memiliki jarak yang signifikan.

4. KESIMPULAN

Penelitian ini berhasil melakukan segmentasi Wajib Pajak (WP) menjadi tiga kluster optimal dengan menggunakan algoritma K-Means. Segmentasi dilakukan berdasarkan kode PTKP, yang mencerminkan status keluarga dan tanggungan, serta pendapatan bersih sebagai indikator kapasitas ekonomi riil. Hasil evaluasi model menunjukkan performa yang sangat memuaskan, yang ditandai dengan Skor Silhouette mencapai 0.64, sementara Elbow Method mengonfirmasi bahwa pemisahan kluster paling efektif pada 3 kelompok. Pada temuan ini sejalan dengan peneliti sebelumnya yang sudah pernah membahasnya, seperti studi Rini Nursaniah et al. (2024), yang juga menegaskan keefektifan K-Means dalam analisis data perpajakan untuk kepentingan yang lebih strategis.

Berdasarkan segmentasi tersebut, Dirjen Pajak dapat merumuskan kebijakan yang lebih personal dan adil. Sebagai contoh, WP pemula dapat diberi pendidikan perpajakan, WP menengah mendapat layanan digital yang lebih baik, sedangkan WP berpenghasilan tinggi dapat menjadi fokus pengawasan yang lebih ketat. Implementasinya dapat juga dilakukan melalui aplikasi digital atau bisa juga dari dashboard perpajakan yang menyediakan rekomendasi dan layanan yang disesuaikan dengan karakteristik setiap kluster.

Untuk pengembangan ke depan, model ini perlu ditingkatkan dengan menambahkan variabel-variabel baru seperti profesi, usia, dan riwayat pelaporan SPT agar segmentasi menjadi lebih komprehensif. Validasi lebih lanjut juga diperlukan menggunakan data riil dari DJP untuk memastikan



akurasi model dalam merepresentasikan kondisi WP yang sebenarnya. Melalui sinergi antara akademisi, praktisi, dan pemerintah, penelitian ini diharapkan dapat menjadi fondasi bagi transformasi digital perpajakan yang berbasis data, sehingga kebijakan yang diambil dapat menjangkau semua lapisan WP secara lebih efisien dan berkeadilan.

REFERENCES

Auliasari, K., & Kertaningtyas, M. (2023). Penerapan algoritma K-means untuk segmentasi konsumen menggunakan R. *Jurnal Teknologi dan Manajemen Informatika*, 5(2), 34–41. <https://doi.org/10.26905/jtmi.v5i2.3644>

Hadi, A. (2023). Segmentasi pelanggan Internet Service Provider (ISP) berbasis pillar K-means. *Jurnal Ilmiah Teknologi Informasi Asia*, 13(2), 413–422. <https://doi.org/10.32815/jitika.v13i2.413>

Irawan, D., Wijaya, G., & Warisaji, T. T. (2023). Penerapan algoritma K-means clustering untuk segmentasi nasabah bank. *BIOS: Jurnal Teknologi Informasi dan Rekayasa Komputer*, 6(1), 162–170. <https://doi.org/10.37148/bios.v6i1.162>

Prayitno, E., Perdana, I. J., Iskandar, E., Winarno, B. H., & Subagyo, A. A. (2024). Optimalisasi profitabilitas ritel melalui segmentasi pelanggan dengan K-means clustering. *Informasi Interaktif: Jurnal Informatika dan Teknologi Informasi*, 9(3), 107–116. <https://doi.org/10.37159/jii.v9i3.107>

Primadeni, F., Pradana, A. I., & Purwanto, E. (2024). Analisis clustering untuk segmentasi wilayah berdasarkan karakteristik PBB di Kabupaten Sragen. *Jurnal Pendidikan dan Teknologi Indonesia*, 1(3), 45–56. <https://doi.org/10.52436/1.jpti.500>

Tjatur Puteri, E., Kusnanto, G., & Thomas, C. J. (2024). Penerapan K-means clustering untuk segmentasi pelanggan pada CRM di PT. Unichem Candi Indonesia. *Konvergensi*, 15(2), 51–60. <https://doi.org/10.30996/konv.v15i2.3651>

Wahyuni, S., & Sriani. (2025). Penerapan algoritma K-means untuk pengelompokan kepatuhan wajib pajak bumi dan bangunan di Kota Medan. *CESS (Journal of Computer Engineering, System and Science)*, 10(1), 325–334. <https://doi.org/10.24114/cess.v10i1.65939>

Worabai, E., Muhammad, A. H. M., & Hidayat, T. (2023). Implementasi metode cluster analysis K-means dalam segmentasi. *Jurnal Ilmiah Komputasi*, 22(3), 441–447. <https://doi.org/10.32409/jikstik.22.3.3493>