Segmentasi Prospek *Business-to-Business*(B2B) Alat Berat pada Kawasan Industri di Jawa Tengah Menggunakan Metode *Fuzzy K-Modes Clustering*

Moch Yahya[1)*], M. Al Haris[2)], Ali Imron[3)], Josua Aditya Manuel[4)], Muhammad Novan Budi Prasetya[5)]
[1,2,3)] Program Studi Statistika, Fakultas Sains dan Teknologi Pertanian, Universitas Muhammadiyah Semarang, Indonesia
[4,5)] PT. United Tractors, Tbk, Semarang, Indonesia

Received: 22 October 2026 Accepted: 13 November 2026 Published: 1 December 2026

*mochyahya002@gmail.com

Kata Kunci: *Fuzzy K-Modes*; segmentasi B2B; *clustering* data kategorik; *Silhouette Coefficient*; alat berat.

DSI: Jurnal Data Science Indonesia is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak: Abstrak: Segmentasi prospek *Business-to-Business* (B2B) membantu mengidentifikasi karakteristik perusahaan sebagai dasar penyusunan strategi penawaran produk alat berat. Penelitian ini menerapkan dan mengevaluasi *Fuzzy K-Modes* untuk segmentasi 364 prospek di KIC, KIK, KIW, dan KITB berdasarkan sektor industri, status operasional, dan skala perusahaan. Parameter dipilih dari 32 kombinasi *fuzzifier* dan jumlah klaster, kemudian dievaluasi melalui pengujian stabilitas dan sensitivitas variabel. Model terpilih menggunakan $\alpha = 1,3$ dan $k = 10$ dengan *Silhouette Coefficient* 0,520 (kategori cukup). Stabilitas model tergolong rendah hingga moderat, dengan rata-rata *Adjusted Rand Index* (ARI) 0,522 antar-*seed* dan 0,473 pada *bootstrap*; sebanyak 27,5% prospek berada di batas antarklaster. Klaster terutama dibentuk oleh sektor industri, dengan skala perusahaan sebagai pembeda tambahan, sedangkan kawasan bukan faktor utama. Pemetaan produk menghasilkan 13 kombinasi paket, dengan enam klaster cukup seragam untuk ditawari satu paket. Hasil penelitian menjadi dasar awal identifikasi prospek dan penyusunan penawaran, dengan mempertimbangkan keterbatasan stabilitas model serta perlunya validasi menggunakan data penawaran dan transaksi aktual.

Pendahuluan

Industri pengolahan merupakan kontributor terbesar terhadap perekonomian Jawa Tengah dengan kontribusi sebesar 33,18% terhadap produk domestik regional bruto pada triwulan II-2026, ketika perekonomian provinsi tumbuh sebesar 5,71%[1]. Perkembangan aktivitas industri tersebut juga tercermin dari pertumbuhan dan aktivitas perusahaan di sejumlah kawasan industri, antara lain Kawasan Industri Wijayakusuma (KIW) dan Kawasan Industri Candi (KIC) di Semarang, Kawasan Industri Kendal (KIK) di Kendal, serta Kawasan Industri Terpadu Batang (KITB) di Batang[2]. Keempat kawasan tersebut berada dalam koridor industri utara Jawa Tengah yang menghubungkan Semarang, Kendal, dan Batang, sehingga relevan untuk dikaji sebagai bagian dari lingkungan pasar industri. Keberadaan perusahaan dari berbagai sektor usaha dan tahapan operasional menjadikan kawasan tersebut sebagai lingkungan prospek *Business-to-Business* (B2B) yang potensial bagi pemasok alat berat. Karakteristik tersebut memberikan dasar untuk melakukan pemetaan prospek berdasarkan kondisi perusahaan dan potensi kebutuhan peralatan industri.

Perusahaan yang beroperasi di kawasan industri memiliki karakteristik organisasi, sektor usaha, dan kondisi operasional yang beragam, sehingga potensi kebutuhan peralatannya tidak dapat diasumsikan seragam[3]. Perbedaan aktivitas perusahaan dapat berkaitan dengan variasi potensi kebutuhan peralatan, seperti *forklift*, peralatan *material handling*, dan alat berat konstruksi, meskipun karakteristik perusahaan tidak secara langsung membuktikan adanya permintaan pembelian[4]. Dalam konteks pemasaran industri, identifikasi kelompok perusahaan dengan karakteristik relatif serupa dapat membantu pemasok memahami heterogenitas pasar dan mengembangkan pendekatan penawaran yang lebih terarah[5]. Segmentasi

prospek menjadi relevan untuk mendukung penentuan prioritas pasar, identifikasi kelompok perusahaan potensial, dan penyusunan rekomendasi produk berdasarkan karakteristik masing-masing kelompok[6]. Dengan demikian, persoalan penelitian tidak hanya berkaitan dengan pengelompokan perusahaan, tetapi juga bagaimana hasil segmentasi dapat diterjemahkan menjadi informasi yang berguna dalam aktivitas pemasaran B2B alat berat.

Penelitian mengenai segmentasi pasar industri telah membahas penggunaan karakteristik organisasi pembeli sebagai dasar pembentukan segmen, sedangkan berbagai penelitian lain telah menerapkan pendekatan *clustering* untuk segmentasi pelanggan dan pengelompokan objek berdasarkan karakteristik tertentu[5]. Namun, karakteristik perusahaan sebagai prospek B2B alat berat dapat didominasi variabel kategorik, sehingga pemilihan metode pengelompokan perlu mempertimbangkan struktur data tersebut[7]. Penelusuran literatur melalui Google Scholar dan Scopus menggunakan kata kunci *B2B market segmentation*, *industrial market segmentation*, *Fuzzy K-Modes*, dan *categorical clustering* pada publikasi periode 2016–2026 belum mengidentifikasi penelitian yang secara spesifik membahas segmentasi prospek B2B pemasok alat berat berdasarkan karakteristik kategorik perusahaan di kawasan industri utara Jawa Tengah. *Fuzzy K-Modes* menjadi salah satu pendekatan yang relevan karena mampu mengolah data kategorik sekaligus memberikan derajat keanggotaan pada setiap klaster, sehingga perusahaan dapat merepresentasikan kedekatan terhadap lebih dari satu segmen[8]. Meskipun demikian, penerapan *Fuzzy K-Modes* tetap memerlukan evaluasi untuk memastikan kualitas dan stabilitas klaster yang dihasilkan dalam merepresentasikan karakteristik prospek B2B alat berat[9].

Penelitian bertujuan mengembangkan segmentasi prospek B2B alat berat menggunakan *Fuzzy K-Modes*, mengevaluasi kualitas dan stabilitas hasil pengelompokan, serta mengidentifikasi karakteristik segmen untuk menyusun rekomendasi produk. Evaluasi dilakukan dengan mempertimbangkan kualitas pemisahan klaster, konsistensi hasil terhadap variasi inisialisasi dan *resampling*, serta sensitivitas terhadap pemilihan variabel. Secara praktis, diharapkan menghasilkan kerangka identifikasi dan prioritas prospek yang dapat mendukung penyesuaian penawaran produk berdasarkan karakteristik perusahaan di KIW, KIK, KIC, dan KITB. Secara metodologis, memberikan evaluasi empiris mengenai penerapan *Fuzzy K-Modes* pada konteks segmentasi prospek industri dengan data kategorik. Cakupan penelitian dibatasi pada empat kawasan industri tersebut, sehingga hasilnya merepresentasikan konteks pemetaan prospek pada lokasi penelitian dan tidak dimaksudkan untuk mewakili seluruh kawasan industri di Jawa Tengah bagian utara.

Tinjauan Literatur

Segmentasi pasar merupakan pendekatan mengelompokkan pelanggan berdasarkan kesamaan karakteristik, kebutuhan, dan perilaku guna menyesuaikan strategi pemasaran[5]. Dalam pasar industri, Mora Cortez *et al.* (2021) membedakan segmentasi menjadi makrosegmentasi berdasarkan karakteristik organisasi pembeli dan mikrosegmentasi berdasarkan proses pengambilan keputusan pembelian. Makrosegmentasi menempatkan sektor industri, lokasi, dan kondisi operasional sebagai karakteristik organisasi yang relevan untuk mengidentifikasi kelompok prospek B2B[5].

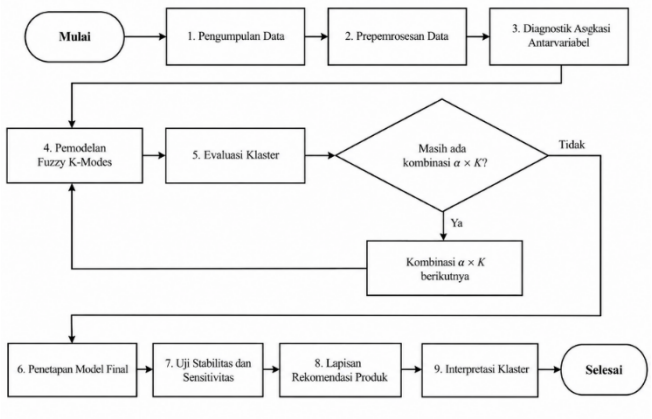
Metode *clustering* memiliki pendekatan berbeda dalam mengelompokkan data berdasarkan tipe variabel, representasi klaster, dan ukuran ketidakmiripan[10]. *K-Modes* menggunakan modus dan *simple matching dissimilarity* untuk mengelompokkan data kategorik[11], *Fuzzy K-Modes* mengembangkan pendekatan tersebut dengan menambahkan derajat keanggotaan sehingga setiap objek dapat memiliki kedekatan terhadap beberapa klaster sekaligus[12]. Fleksibilitas representasi memungkinkan *Fuzzy K-Modes* mengakomodasi ketidakpastian batas antarkelompok yang tidak direpresentasikan secara langsung oleh metode partisi tegas[13], sementara pengembangan *fuzzy* centroid oleh Kim *et al.* (2004) menunjukkan potensi peningkatan kualitas pengelompokan pada data kategorik[14]. Di sisi lain, Gower-PAM dan *hierarchical clustering* mengandalkan matriks ketidakmiripan[15], MCA mengombinasikan reduksi dimensi *clustering*[16], *Latent Class Analysis* menggunakan pendekatan probabilistik mengidentifikasi kelas laten[17]. *Fuzzy K-Modes* relevan dipertimbangkan untuk segmentasi prospek B2B alat berat karena mengakomodasi karakteristik kategorik sekaligus merepresentasikan derajat kedekatan perusahaan terhadap beberapa segmen[18].

Fuzzy clustering memungkinkan suatu objek memiliki derajat keanggotaan pada beberapa klaster,

sedangkan *Fuzzy K-Modes* menggabungkan mekanisme tersebut dengan modus sebagai pusat klaster dan *matching dissimilarity* sebagai ukuran ketidakmiripan[19]. Pengembangan metode mencakup pendekatan multivariat[8], *possibilistic* dan metaheuristik[19], serta berbagai modifikasi algoritma[20], meskipun sensitivitas terhadap inisialisasi dan ukuran ketidakmiripan masih menjadi tantangan[21]. Penelitian terdahulu telah menerapkan *Fuzzy K-Modes* pada segmentasi pasar konsumen[22] dan pengelompokan pemasok[18]. Namun, penelusuran literatur belum mengidentifikasi penerapan spesifik metode tersebut untuk segmentasi prospek B2B pemasok alat berat pada kawasan industri di Indonesia. Penelitian ini merupakan studi penerapan dan evaluasi empiris *Fuzzy K-Modes* berdasarkan kualitas dan stabilitas pengelompokan menghasilkan profil segmen serta rekomendasi produk berdasarkan karakteristik perusahaan.

Metode Penelitian

Penelitian berdesain kuantitatif-deskriptif dengan analisis klaster non-hierarki. Alur penelitian terdiri atas delapan tahap (Gambar 1), dengan Tahap 4 dan 5 diulang untuk setiap kombinasi α × k hingga seluruh kombinasi kandidat selesai dimodelkan dan dievaluasi.



Gambar 1 Diagram Alir Penelitian

1. Pengumpulan Data

   Data sekunder berasal dari pemetaan (*mapping*) 364 prospek B2B alat berat oleh Bagian *Sales* pada Mei–Agustus 2026 di KIC, KIK, KITB, dan KIW. Seluruh prospek hasil pemetaan digunakan sebagai unit analisis, tetapi data tersebut tidak mencakup seluruh perusahaan di keempat kawasan. Penelitian menggunakan tiga variabel pembentuk klaster dan satu variabel *profiling* (Tabel 1). Sektor industri terdiri atas 16 kategori, sedangkan skala perusahaan mengacu pada jumlah tenaga kerja dan investasi berdasarkan Permenperin Nomor 64/M-IND/PER/7/2016.

Tabel 1 Variabel Penelitian dan Distribusi Kategori (n = 364)

Simbol	Variabel	Peran	Kategori (jumlah; proporsi)
X1	Sektor Industri	Pembentuk klaster	16 kategori; terbesar Tekstil, Garmen, Kulit & Alas Kaki (45; 12,4%)
X2	Status Operasional	Pembentuk klaster	Beroperasi (349; 95,9%); Progres Pembangunan (15; 4,1%)
X3	Skala Perusahaan	Pembentuk klaster	Besar (193; 53,0%); Menengah (141; 38,7%); Kecil (30; 8,2%)
–	Kawasan Industri	*Profiling*	KIC (139; 38,2%); KIK (127; 34,9%); KITB (47; 12,9%); KIW (51; 14,0%)

   Kawasan industri digunakan untuk *profiling* dan mengevaluasi keterkaitan klaster dengan lokasi. Kebutuhan alat berat tidak dimasukkan dalam model, tetapi menjadi lapisan rekomendasi setelah klaster terbentuk (Bagian 3.8). Seluruh variabel dipertahankan sebagai kategorik nominal untuk perhitungan *matching dissimilarity*.

2. Prapemrosesan Data

   Prapemrosesan meliputi pemeriksaan tipe data, nilai hilang, jumlah kategori, profil unik, dan duplikasi

nama perusahaan. Seluruh variabel dikonversi ke tipe *string* sebagai kategorik nominal tanpa normalisasi sehingga diperoleh matriks data berukuran 364 × 3, dan variabel dengan kategori dominan di atas 97% dikeluarkan dari model sebagai kriteria yang ditetapkan peneliti.

3. Diagnostik Asosiasi Antarvariabel

   Tanpa pembobotan, *Fuzzy K-Modes* memberi kontribusi sama pada setiap variabel, sehingga dua variabel yang berasosiasi kuat membawa informasi hampir sama dan berpengaruh ganda pada klaster. Diagnostik ini menilai redundansi, bukan menentukan bobot.

   Asosiasi antarvariabel diukur dengan Cramér's V terkoreksi bias [23] untuk enam pasangan dari empat variabel, yaitu tiga variabel pembentuk klaster dan Kawasan (Persamaan 1). Simbol r dan c adalah jumlah baris dan kolom tabel kontingensi.

$$V_{kor} = \sqrt{\frac{\phi^2_{kor}}{\min(r_{kor}-1, c_{kor}-1)}} \quad \text{.....(1)}$$

   Pasangan dengan $V_{kor} \geq 0{,}5$ ditelaah berdasarkan relevansi bisnis, dengan dampak keputusan diuji pada Bagian 3.7. Kecukupan frekuensi sel dievaluasi melalui frekuensi harapan, termasuk proporsi sel kurang dari 5 dan nilai minimum (Tabel 3). Pasangan dengan lebih dari 20% sel berfrekuensi harapan kurang dari 5 diinterpretasikan secara indikatif.

4. Pemodelan *Fuzzy K-Modes*

   Algoritma yang digunakan adalah *Fuzzy K-Modes* tanpa pembobotan atribut [8]; implikasi pilihan ini diperiksa pada Bagian 3.3 dan 3.7. Notasi: n = 364 prospek; p = 3 atribut kategorik; k jumlah klaster; α > 1 *fuzzifier*; u_{li} derajat keanggotaan prospek ke-i pada klaster ke-l; z_l modus klaster ke-l; ε ambang toleransi penurunan J; n_{init} banyaknya inisialisasi acak; I(·) fungsi indikator (bernilai 1 bila kondisi di dalam kurung terpenuhi dan 0 bila tidak).

$$d(x_i, z_l) = \sum_{j=1}^{p} \delta(x_{ij}, z_{lj}) \quad \text{.....(2)}$$

   dengan δ = 0 bila kategori sama dan δ = 1 bila berbeda, sehingga seluruh atribut diperlakukan setara.

$$J(U, Z) = \sum_{l=1}^{k} \sum_{i=1}^{n} (u_{li})^{\alpha}\, d(x_i, z_l) \quad \text{.....(3)}$$

$$u_{li} \in [0,1], \quad \sum_{l=1}^{k} u_{li} = 1, \quad 0 < \sum_{i=1}^{n} u_{li} < n \quad \text{.....(4)}$$

$$u_{li} = \frac{1}{\sum_{h=1}^{k} \left(\frac{d(x_i, z_l)}{d(x_i, z_h)}\right)^{\frac{1}{\alpha-1}}} \quad \text{.....(5)}$$

Bila x_i identik dengan z_l, maka $u_{li} = 1$ dan $u_{hi} = 0$ untuk $h \neq l$.

$$z_{lj} = \arg \max_{c} \sum_{i=1}^{n} (u_{li})^{\alpha} \cdot I(x_{ij} = c) \quad \text{.....(6)}$$

Parameter α (fuzzifier) mengatur tingkat tumpang tindih keanggotaan antarklaster, sedangkan k menentukan jumlah klaster. Nilai α mendekati 1 menghasilkan partisi lebih tegas, sementara nilai lebih besar menyebarkan keanggotaan. Keduanya ditentukan melalui evaluasi model kandidat (Bagian 3.5) dan seleksi (Bagian 3.6), bukan diestimasi oleh algoritma.

Kandidat parameter adalah α ∈ {1,1; 1,3; 1,5; 2,0} dan k ∈ {3, 4, …, 10}, sehingga diperoleh 4 × 8 = 32 model. Setiap kombinasi (α, k) dimodelkan dengan langkah berikut:

a. Inisialisasi modus awal $Z^{(0)}$ dengan inisialisasi Huang [11] [19] .
b. Perbarui U memakai Persamaan 5.
c. Perbarui Z memakai Persamaan 6.
d. Hitung J (Persamaan 3). Ulangi langkah 2 dan 3 hingga penurunan J < ε atau iterasi maksimum tercapai, dengan ε = 10^{-10} dan iterasi maksimum 100 iterasi.
e. Ulangi langkah 1 sampai 4 sebanyak n_{init} = 300 karena *Fuzzy K-Modes* sensitif terhadap pusat awal[21]. Simpan solusi dengan J terendah sebagai model kombinasi tersebut.
f. Defuzzifikasi: $label_i = \arg\max_l u_{li}$.

Inisialisasi Huang dipilih karena inisialisasi Cao yang deterministik tidak dapat diulang dengan *seed* berbeda sehingga tidak memungkinkan penilaian stabilitas antar *seed* [8]. Keluaran setiap model berupa matriks U, modus Z, nilai J, dan label defuzzifikasi; algoritma dijalankan di Python dengan NumPy, scikit-learn, SciPy, dan pandas.

5. Evaluasi Klaster

Setiap model dari Bagian 3.4 dievaluasi pada label defuzzifikasi dengan jarak *matching dissimilarity* tanpa bobot (Persamaan 2). *Silhouette Coefficient* :

$$a_i = \frac{1}{|C_p|-1}\sum_{j\neq i,\, x_j\in C_p} d(x_i, x_j), \quad x_i \in C_p \text{.....(7)}$$

$$b_i = \min_{q\neq p}\left\{\frac{1}{|C_q|}\sum_{x_j\in C_q} d(x_i, x_j)\right\} \text{.....(8)}$$

$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}}, \quad s_i = 0 \text{ bila } |C_p| = 1 \text{.....(9)}$$

$$SC = \frac{1}{n}\sum_{i=1}^{n} s_i \text{.....(10)}$$

Interpretasi nilai mengikuti pedoman umum *Silhouette*: SC > 0,70 struktur kuat; 0,51 sampai 0,70 cukup; 0,26 sampai 0,50 lemah; ≤ 0,25 tidak ada struktur berarti[24]. SC dihitung pada hasil defuzzifikasi sehingga tidak mengukur keanggotaan parsial, dan pada data kategorik nilainya bergantung pada ukuran jarak antarkategori [25]. Pada data berkardinalitas rendah banyak objek identik, sehingga nilai SC dibaca hati-hati. Indeks keanggotaan *fuzzy* :

$$\text{PC} = \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{k} u_{li}^2 \text{.....(11)}$$

$$\text{MPC} = 1 - \frac{k}{k-1}(1 - \text{PC}) \text{.....(12)}$$

$$\text{PE} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{k} u_{li} \ln u_{li} \text{.....(13)}$$

PC mengukur ketegasan keanggotaan (1/k sampai 1), MPC menormalkan PC ke rentang 0 sampai 1, dan PE mengukur ketidakpastian keanggotaan (0 sampai ln k). *Fuzzy Silhouette* membobot *Silhouette* setiap prospek dengan selisih keanggotaan terbesar dan terbesar kedua (Persamaan 14), dengan s_i dari Persamaan 9.

$$\text{FSC} = \frac{\sum_{i=1}^{n}(u_{(1)i} - u_{(2)i})s_i}{\sum_{i=1}^{n}(u_{(1)i} - u_{(2)i})} \text{.....(14)}$$

Prospek batas ditandai bila selisih keanggotaan terbesar dan kedua kurang dari τ (Persamaan 15). P_b adalah proporsi prospek batas. Nilai τ = 0,2 ditetapkan sebelum hasil dilihat, dan sensitivitas terhadap τ dilaporkan.

$$P_b = \frac{1}{n}\sum_{i=1}^{n} I\left(u_{(1)i} - u_{(2)i} < \tau\right), \quad \tau = 0{,}2 \text{.....(15)}$$

Hasil evaluasi seluruh model dirangkum dalam tabel model kandidat. Tidak satu pun ukuran ini membuktikan bahwa segmen berguna secara bisnis; kegunaan bisnis dinilai pada Bagian 3.8 dan 3.9.

6. Penetapan Model Final

Pemilihan model dilakukan dalam dua tahap, yaitu pemilihan parameter (α^*, k^*) dari tabel model kandidat (Bagian 3.5) dan penetapan solusi final, dengan aturan berikut:

a. Seluruh model kandidat dinilai tanpa batas proporsi klaster minimum, dan proporsi klaster terkecil dilaporkan sebagai informasi pendukung. Rentang k yang dimulai dari 3 adalah keputusan peneliti berbasis kebutuhan bisnis, bukan sifat statistik.
b. Dari seluruh model kandidat dibentuk daftar pendek, yaitu model dengan selisih SC terhadap SC tertinggi tidak lebih dari 0,02. Pasangan terpilih (α^*, k^*) ditetapkan dari daftar pendek tersebut sebagai model dengan rata-rata ARI berpasangan antar *seed* tertinggi (Bagian 3.7).
c. Solusi final dipilih dari seluruh percobaan pada (α^*, k^*), yaitu solusi hasil pemilihan model, solusi pada 10 *seed* stabilitas, dan 30 percobaan tambahan, dengan memilih solusi yang memiliki J terendah. Seluruh metrik evaluasi dihitung ulang pada solusi final dan *random seed* dicatat. Model final memuat α^*, k^*, matriks U^*, modus Z^*, dan label final.

7. Uji Stabilitas dan Sensitivitas

Uji stabilitas mempertahankan α* dan k* untuk menilai konsistensi struktur klaster melalui 10 random seed dan 30 sampel bootstrap, menggunakan ARI dan sebaran J. Uji sensitivitas membandingkan model utama (S1) dengan empat skenario perubahan variabel (S2–S5), menggunakan ARI dan NMI pada parameter optimal masing-masing maupun parameter yang disamakan dengan S1. ARI=1 menunjukkan partisi identik, sedangkan nilai mendekati 0 menunjukkan kesesuaian setara acak. Hasil stabilitas

diinterpretasikan secara deskriptif tanpa ambang baku dan ketidakstabilan dilaporkan sebagai keterbatasan.

8. Lapisan Rekomendasi Produk

   Rekomendasi produk bukan variabel pembentuk klaster. Aturan pada Tabel 2 memetakan sektor, status, dan skala setiap prospek ke paket produk setelah klaster terbentuk. Aturan berasal dari pengalaman tenaga penjualan dan belum divalidasi dengan data penawaran atau transaksi; Fase Bangun adalah status Progres Pembangunan, dan ukuran genset mengikuti katalog (50, 160, dan 500 kVA), bukan beban aktual.

Tabel 2. Aturan Rekomendasi Produk

Produk	Aturan
Genset	50 kVA untuk skala Kecil, 160 kVA untuk Menengah, 500 kVA untuk Besar; hanya prospek Beroperasi
Forklift	3 ton baterai untuk prospek Beroperasi di luar sektor Konstruksi & Bahan Bangunan, Real Estat & Pengelola Kawasan, Energi, dan Jasa Penunjang; 5 ton diesel untuk Konstruksi & Bahan Bangunan yang Beroperasi
Excavator dan *bulldozer*	Prospek Fase Bangun, atau sektor Konstruksi & Bahan Bangunan, Real Estat & Pengelola Kawasan, dan Energi
Truck	Prospek Fase Bangun, atau sektor Konstruksi & Bahan Bangunan, Logistik, dan Makanan

Kesesuaian paket(KP) mengukur persentase prospek dalam klaster yang menerima paket sama dengan paket modus klaster (Persamaan 16), dengan g_i paket prospek ke-i. KP tinggi berarti klaster cukup seragam untuk ditawari satu paket. KP naik otomatis bila jumlah kelompok bertambah, sehingga selalu dilaporkan bersama jumlah kelompok.

$$\mathrm{KP} = \frac{1}{n}\sum_{l=1}^{k} \max_{g} |\{i \in C_l : g_i = g\}| \quad \text{.....(16)}$$

9. Interpretasi Klaster

   Interpretasi memakai model final (Bagian 3.6) serta hasil Bagian 3.7 dan 3.8. Profil setiap segmen disusun dari modus z_{lj} dan persentase dominasi kategori tersebut pada klaster ke-l (Persamaan 17).

$$P_{lj} = \frac{|\{i \in C_l : x_{ij} = z_{lj}\}|}{|C_l|} \quad \text{.....(17)}$$

   Prospek perbatasan, yaitu $u_{(1)i} - u_{(2)i} < \tau$, dibaca sebagai prospek yang relevan bagi lebih dari satu segmen. Nama segmen diberikan setelah profil dibaca. Rekomendasi adalah implikasi hasil segmentasi dan tidak membuktikan peningkatan penjualan, efisiensi distribusi, atau penghematan biaya.

Hasil Penelitian

1. Karakteristik Data dan Hasil Prapemrosesan

   Karakteristik tiga variabel pembentuk klaster dan variabel Kawasan pada 364 prospek disajikan pada Tabel 1. Tidak ada nilai hilang, dan data memuat 48 profil unik dari 364 prospek pada tiga variabel pembentuk klaster.

   Sektor Industri (X1) tidak memiliki kategori yang menonjol: terbesar Tekstil, Garmen, Kulit & Alas Kaki (45 prospek; 12,4%), disusul Makanan, Minuman & Agroindustri (42), Kayu, Furnitur & Barang Kayu (40), dan Logistik, Pergudangan & Transportasi (36), sedangkan Mainan, Olahraga & Dekorasi memuat 5 prospek dan Kerajinan & Perhiasan 1 prospek. Tidak ada variabel dengan kategori dominan melebihi ambang 97%, sehingga ketiganya dipertahankan; Status (95,9% Beroperasi) mendekati ambang dan hampir tidak membedakan prospek, sehingga dampak pengeluarannya diuji pada Bagian 4.8.

2. Diagnostik Asosiasi Antarvariabel

   Keterkaitan antarvariabel diuji pada enam pasangan dari empat variabel dengan Cramér's V terkoreksi bias, disertai kecukupan frekuensi harapan (E = frekuensi harapan sel; Tabel 3).

Tabel 3. Cramér's V Terkoreksi Bias dan Kecukupan Frekuensi Harapan Antarvariabel

Variabel pertama	Variabel kedua	V	Sel E < 5, n (%)	Sel E < 1	E min.	Keterangan
X1 Sektor Industri	X3 Skala Perusahaan	0,556	23 (47,9)	7	0,08	Asosiasi kuat (≥ 0,5)

Variabel pertama	Variabel kedua	V	Sel E < 5, n (%)	Sel E < 1	E min.	Keterangan
X2 Status Operasional	Kawasan Industri	0,449	2 (25,0)	0	1,94	Sedang
X1 Sektor Industri	Kawasan Industri	0,219	35 (54,7)	6	0,13	Sedang
X1 Sektor Industri	X2 Status Operasional	0,218	18 (56,2)	10	0,04	Sedang
X3 Skala Perusahaan	Kawasan Industri	0,207	2 (16,7)	0	3,87	Sedang
X2 Status Operasional	X3 Skala Perusahaan	0,027	1 (16,7)	0	1,24	Hampir tidak berasosiasi

Hanya satu dari enam pasangan mencapai V terkoreksi ≥ 0,5, yaitu Sektor Industri dan Skala Perusahaan (0,556), sedangkan lima pasangan lainnya berkisar 0,027–0,449. Tingginya proporsi sel berfrekuensi harapan kurang dari 5 pada empat pasangan menyebabkan nilai V dibaca indikatif. Ketiga variabel tetap dipertahankan karena Skala menentukan ukuran genset pada rekomendasi produk, sementara dampak asosiasi Sektor dan Skala diuji pada Bagian 4.8. Frekuensi teramati dan harapan lengkap disajikan dalam materi suplementer.

3. Pemodelan *Fuzzy K-Modes*

Empat nilai α (1,1; 1,3; 1,5; 2,0) dan k=3 sampai 10 menghasilkan 32 model kandidat, masing-masing dijalankan dengan 300 inisialisasi acak (Tabel 4).

Tabel 4 *Silhouette Coefficient* Model Kandidat menurut α dan k

α \ k	3	4	5	6	7	8	9	10
1,1	0,297	0,243	0,252	0,334	0,381	0,367	0,476	0,447
1,3	0,296	0,243	0,267	0,321	0,364	0,437	0,442	0,471
1,5	0,319	0,210	0,272	0,302	0,322	0,443	0,471	0,434
2,0	0,318	0,196	0,243	0,312	0,390	0,430	0,395	0,390

Proporsi klaster terkecil pada seluruh kandidat berkisar 0,82% sampai 11,54%, dan 13 dari 32 model memiliki klaster terkecil di bawah 5% data. *Silhouette* tertinggi terdapat pada α=1,1 dan k=9 (0,476), disusul α=1,3 dan k=10 serta α=1,5 dan k=9 (masing-masing 0,471). Ketiganya berselisih tidak lebih dari 0,02 sehingga masuk daftar pendek (Tabel 5), sedangkan model terdekat berikutnya (α=1,1; k=10) hanya 0,447; ketiganya diterima karena tidak ada batas proporsi klaster minimum.

Tabel 5 Daftar Pendek Model Kandidat

α	k	*Silhouette*	FSC	PC	MPC	Klaster terkecil (%)	ARI antar *seed* (rata-rata)	ARI antar *seed* (min.)
1,3	10	0,471	0,751	0,676	0,640	1,65	0,522	0,369
1,1	9	0,476	0,764	0,716	0,681	5,22	0,492	0,251
1,5	9	0,471	0,720	0,600	0,550	2,20	0,448	0,244

Selisih *Silhouette* antarmodel hanya 0,005, sehingga pemilihan akhir didasarkan pada *stabilitas seed*. Model α=1,3 dan k=10 memiliki rata-rata ARI tertinggi (0,522; minimum 0,369), dibandingkan dua model alternatif. Namun, stabilitas tersebut masih terbatas dan bukan optimum global. Nilai k=10 juga berada pada batas atas rentang pengujian, sehingga rentang k lebih luas berpotensi menghasilkan pilihan berbeda. Perilaku indeks terhadap k pada α=1,3 disajikan pada Tabel 6.

Tabel 6. Indeks Validitas menurut k pada α=1,3

k	*Silhouette*	FSC	PC	MPC	PE	Klaster terkecil (%)
3	0,296	0,506	0,666	0,500	0,561	10,16
4	0,243	0,511	0,605	0,474	0,690	8,24
5	0,267	0,587	0,597	0,496	0,761	6,59
6	0,321	0,653	0,605	0,526	0,788	6,59
7	0,364	0,654	0,624	0,562	0,773	6,04
8	0,437	0,739	0,648	0,598	0,755	4,95
9	0,442	0,717	0,688	0,650	0,717	4,67
10	0,471	0,751	0,676	0,640	0,739	1,65

Silhouette meningkat dari 0,243 (k=4) menjadi 0,471 (k=10), tetapi kenaikan belum tentu mencerminkan struktur klaster lebih baik karena hanya terdapat 48 profil unik. Proporsi klaster terkecil juga menurun hingga 1,65% pada k=10, sedangkan PC tertinggi (0,688) terdapat pada k=9 dan PE terendah (0,561) pada k = 3. Peningkatan α menurunkan PC dan meningkatkan PE, menunjukkan keanggotaan semakin merata. Fungsi objektif J menurun seiring bertambahnya k tanpa titik siku yang jelas, sehingga tidak

menjadi dasar tunggal pemilihan model.

4. Penyesuaian Ulang Model Akhir
   Solusi final dipilih dari 41 percobaan pada α = 1,3 dan k = 10, dengan fungsi objektif terendah 89,407 yang dicapai dua percobaan ber-ARI 0,622, menunjukkan adanya partisi berbeda dengan biaya sama. Model menghasilkan *Silhouette* 0,520 (cukup), PC 0,722, MPC 0,691, dan rata-rata keanggotaan tertinggi 0,740. Klaster terkecil mencakup 20 prospek (5,49%), menunjukkan keanggotaan relatif tegas meski masih tumpang tindih. Seluruh pemeriksaan implementasi terpenuhi, termasuk normalisasi keanggotaan, kesesuaian fungsi objektif, dan 180 pengujian penggantian modus tanpa penurunan nilai J.
5. Evaluasi Stabilitas
   Stabilitas model tergolong rendah hingga moderat, dengan rata-rata ARI antar-*seed* 0,522 (minimum 0,369) dan terhadap solusi final 0,529 (minimum 0,374). Pada 264 prospek non-batas, ARI meningkat menjadi 0,713. Sementara itu, 30 sampel *bootstrap* menghasilkan rata-rata ARI 0,473 (minimum 0,349), menunjukkan sensitivitas model terhadap inisialisasi dan komposisi sampel.
6. Profil dan Interpretasi Klaster
   Profil sepuluh klaster model final disusun dari modus dan dominasi setiap variabel sebagai dasar penamaan segmen (Tabel 7); n adalah jumlah prospek menurut penugasan tegas dan Σu jumlah derajat keanggotaan seluruh prospek pada klaster.

Tabel 7. Profil Klaster Model Final

C	Nama segmen	n	Σu	Sektor dominan (%)	Skala dominan (%)	Status dominan (%)	Kawasan dominan (%)
C1	Skala besar lintas sektor	73	25,9	Elektronik, Listrik & Otomotif (26)	Besar (79)	Beroperasi (95)	KIK (55)
C2	Kimia dan plastik	26	29,7	Kimia, Plastik & Karet (100)	Besar (73)	Beroperasi (100)	KIC (46)
C3	Skala menengah lintas sektor	67	39,7	Logistik, Pergudangan & Transportasi (45)	Menengah (91)	Beroperasi (93)	KIC (63)
C4	Logam dan mesin	24	32,6	Logam, Mesin & Fabrikasi (100)	Besar (92)	Beroperasi (96)	KIC (46)
C5	Makanan dan agroindustri	37	44,9	Makanan, Minuman & Agroindustri (100)	Besar (89)	Beroperasi (97)	KIC (43)
C6	Perdagangan dan distribusi	23	32,6	Perdagangan & Distribusi (100)	Menengah (78)	Beroperasi (100)	KIC (65)
C7	Konstruksi dan bahan bangunan	24	36,0	Konstruksi & Bahan Bangunan (100)	Menengah (100)	Beroperasi (92)	KITB (38)
C8	Farmasi dan alat kesehatan	20	29,7	Farmasi, Kosmetik & Alat Kesehatan (100)	Besar (100)	Beroperasi (100)	KIC (35)
C9	Kayu dan furnitur	29	42,1	Kayu, Furnitur & Barang Kayu (100)	Menengah (100)	Beroperasi (97)	KIC (31)
C10	Tekstil dan alas kaki	41	50,7	Tekstil, Garmen, Kulit & Alas Kaki (100)	Besar (100)	Beroperasi (98)	KIK (41)

Delapan klaster memuat satu sektor secara utuh, sedangkan C1 (73 prospek; 20,1%) dan C3 (67 prospek; 18,4%) memiliki sektor beragam, masing-masing didominasi skala Besar (79%) dan Menengah (91%). Status Beroperasi mendominasi seluruh klaster (92–100%). Selisih jumlah anggota dan Σu pada C1 (25,9) serta C3 (39,7) menunjukkan penyebaran keanggotaan lebih besar. Prospek skala Kecil hanya terdapat pada lima klaster, sementara 15 prospek berstatus Progres Pembangunan tersebar di tujuh klaster. Sebaran berdasarkan kawasan disajikan pada Tabel 8.

Tabel 8 Sebaran Klaster menurut Kawasan (Jumlah Prospek)

Kawasan	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	Total
KIC	14	12	42	11	16	15	6	7	9	7	139
KIK	40	10	13	9	12	4	7	7	8	17	127
KITB	9	2	7	2	4	1	9	3	7	3	47

KIW	10	2	5	2	5	3	2	3	5	14	51
Total	73	26	67	24	37	23	24	20	29	41	364

Tidak ada klaster yang terkonsentrasi penuh pada satu kawasan. Konsentrasi tertinggi terdapat pada C6 di KIC (65%), C3 di KIC (63%), dan C1 di KIK (55%), sementara keempat kawasan tersebar pada seluruh klaster, sejalan dengan NMI Kawasan yang rendah (0,058). Rata-rata keanggotaan fuzzy berkisar 0,78–1,00 pada C2 dan C4–C10, sedangkan C1 (0,33) dan C3 (0,52) lebih rendah akibat penyebaran keanggotaan yang mengikuti kesamaan skala industri, yaitu Besar dan Menengah.

7. Prospek Batas

Prospek batas didefinisikan sebagai prospek dengan selisih keanggotaan tertinggi dan kedua (margin) kurang dari 0,20. Sebanyak 100 dari 364 prospek (27,5%) memenuhi kriteria tersebut, seluruhnya berasal dari C1 (56), C2 (7), dan C3 (37), dengan dominasi C1 dan C3 (93 prospek). Jumlah ini relatif stabil pada variasi ambang margin 0,10–0,30 dan keanggotaan maksimum 0,50, tetapi meningkat hingga 125 prospek (34,3%) pada ambang keanggotaan maksimum 0,70–0,80, menunjukkan sensitivitas hasil terhadap pemilihan ambang.

8. Analisis Sensitivitas Set Variabel

Tabel 9 membandingkan model utama (S1) dengan empat set variabel alternatif. Setiap skenario memilih α dan k dengan prosedur yang sama, sehingga ARI terhadap S1 bercampur dengan perbedaan α dan k; kolom ARI dengan α dan k disamakan memisahkan pengaruh set variabel dari pemilihan parameter. *Silhouette* tidak dibandingkan antarskenario karena jumlah variabel berbeda.

Tabel 9. Hasil Analisis Sensitivitas Set Variabel

Skenario	p	α	k	*Silh.*	ARI vs S1	ARI vs S1 (α, k sama)	NMI Kaw.	NMI Sek.	NMI Stat.	NMI Skala
S1 Model utama (3 variabel)	3	1,3	10	0,520	1,000	1,000	0,058	0,771	0,011	0,391
S2 Tanpa Skala	2	1,3	10	0,733	0,557	0,557	0,080	0,853	0,125	0,154
S3 Tanpa Status	2	1,1	10	0,550	0,523	0,648	0,051	0,757	0,009	0,450
S4 Sektor langka digabung	3	1,3	9	0,437	0,517	0,630	0,053	0,651	0,007	0,558
S5 Ditambah Kawasan (4 variabel)	4	1,1	6	0,289	0,226	0,310	0,550	0,279	0,044	0,413

Penambahan Kawasan (S5) menghasilkan perubahan partisi terbesar (ARI 0,226) dan meningkatkan NMI Kawasan menjadi 0,550, sehingga Kawasan menjadi penentu klaster. Penghapusan Skala (S2) juga mengubah partisi (ARI 0,557) dan memperkuat dominasi sektor (NMI 0,853), sedangkan penghapusan Status (S3) menunjukkan pengaruh relatif kecil (NMI 0,009). Penggabungan sektor langka (S4) menghasilkan ARI 0,517. Secara keseluruhan, partisi dipengaruhi oleh sektor dan skala serta tidak sepenuhnya independen terhadap pemilihan variabel.

9. Paket Produk Rekomendasi

Aturan Tabel 2 diterapkan pada setiap prospek, lalu paket dirangkum menurut klaster model final. Tabel 10 menyajikan sebaran paket seluruh prospek.

Tabel 10. Sebaran Paket Produk Hasil Aturan

Paket	Jumlah	Persen (%)
Genset 500 kVA + *Forklift* 3 ton baterai	127	34,9
Genset 160 kVA + *Forklift* 3 ton baterai	73	20,1
Genset 500 kVA + *Forklift* 3 ton baterai + *Truck*	38	10,4
Genset 160 kVA + *Forklift* 3 ton baterai + *Truck*	29	8,0
Genset 160 kVA + *Forklift* 5 ton diesel + *Excavator* + *Bulldozer* + *Truck*	22	6,0
Genset 50 kVA + *Forklift* 3 ton baterai	20	5,5
Genset 500 kVA + *Excavator* + *Bulldozer*	16	4,4
Excavator + *Bulldozer* + *Truck*	15	4,1
Genset 50 kVA + *Forklift* 3 ton baterai + *Truck*	10	2,7
Genset 160 kVA	6	1,6
Genset 500 kVA + *Forklift* 5 ton diesel + *Excavator* + *Bulldozer* + *Truck*	4	1,1
Genset 160 kVA + *Excavator* + *Bulldozer*	3	0,8
Genset 500 kVA	1	0,3

Aturan menghasilkan 13 paket berbeda. Dua paket terbesar menampung 200 prospek (54,9%), yaitu genset 500 kVA dengan *forklift* 3 ton baterai (127; 34,9%) dan genset 160 kVA dengan *forklift* 3 ton baterai (73; 20,1%), sedangkan tujuh paket masing-masing kurang dari 5% dan bersama-sama menampung 55 prospek (15,1%). Tabel 11 merangkum paket modus pada tiap klaster.

Tabel 11. Paket Modus dan Kesesuaian Paket per Klaster

C	Nama segmen	n	Paket modus	KP (%)	Jumlah paket berbeda
C1	Skala besar lintas sektor	73	Genset 500 kVA+*Forklift* 3 ton baterai	37,0	8
C2	Kimia dan plastik	26	Genset 500 kVA+*Forklift* 3 ton baterai	73,1	2
C3	Skala menengah lintas sektor	67	Genset 160 kVA+*Forklift* 3 ton baterai + *Truck*	43,3	6
C4	Logam dan mesin	24	Genset 500 kVA + *Forklift* 3 ton baterai	87,5	3
C5	Makanan dan agroindustri	37	Genset 500 kVA + *Forklift* 3 ton baterai + *Truck*	86,5	3
C6	Perdagangan dan distribusi	23	Genset 160 kVA + *Forklift* 3 ton baterai	78,3	2
C7	Konstruksi dan bahan bangunan	24	Genset 160 kVA + *Forklift* 5 ton diesel + *Excavator* + *Bulldozer* + *Truck*	91,7	2
C8	Farmasi dan alat kesehatan	20	Genset 500 kVA + *Forklift* 3 ton baterai	100,0	1
C9	Kayu dan furnitur	29	Genset 160 kVA + *Forklift* 3 ton baterai	96,6	2
C10	Tekstil dan alas kaki	41	Genset 500 kVA + *Forklift* 3 ton baterai	97,6	2

Enam klaster (C4, C5, C7, C8, C9, dan C10) cukup seragam (KP 86,5% sampai 100%; 1 sampai 3 paket berbeda), sehingga masing-masing dapat ditawari satu paket. Klaster C2 (73,1%) dan C6 (78,3%) sebagian seragam, sedangkan C1 (37,0%; 8 paket berbeda) dan C3 (43,3%; 6 paket berbeda) tidak seragam karena menggabungkan sektor dengan aturan berbeda. Keseluruhan KP model *Fuzzy K-Modes* adalah 70,3%. Untuk C1 dan C3, paket ditentukan per prospek memakai aturan Tabel 2.

PEMBAHASAN

Fuzzy K-Modes pada sektor, status, dan skala menghasilkan sepuluh klaster dengan α=1,3. Delapan klaster (C2 dan C4–C10) memuat satu sektor secara utuh, sedangkan C1 dan C3 menampung 140 prospek (38,5%) dari berbagai sektor dengan skala yang sama, yaitu Besar dan Menengah. Pola ini tercermin pada NMI terhadap sektor (0,771) dan skala (0,391), sementara hubungan dengan kawasan (0,058) dan status (0,011) hampir tidak ada. Struktur klaster memiliki kualitas cukup berdasarkan *Silhouette* (0,520), tetapi stabilitas antar-*seed* (0,522) dan *bootstrap* (0,473) masih rendah hingga moderat. Kondisi tersebut dipengaruhi oleh keterbatasan 48 profil unik dari 364 prospek dan jarak Hamming yang hanya bernilai 0–3, sehingga 27,5% prospek tergolong prospek batas dan ARI meningkat dari 0,522 menjadi 0,713 ketika prospek batas dikeluarkan.

Konfigurasi terpilih menghasilkan dua partisi dengan fungsi objektif terendah yang memiliki ARI 0,622, menunjukkan bahwa solusi optimal belum sepenuhnya konsisten antar-inisialisasi. Keanggotaan *fuzzy* terutama menunjukkan prospek yang memiliki kemiripan dengan beberapa modus berskala sama, sehingga nilai tambahnya terbatas pada identifikasi prospek batas. Secara praktis, enam klaster cukup seragam untuk ditawarkan satu paket, sedangkan C1 dan C3 memerlukan pendekatan per prospek. Dengan demikian, hasil segmentasi dapat digunakan sebagai dasar penawaran awal dan prioritas kunjungan, bukan bukti peningkatan penjualan atau efisiensi distribusi. Temuan mengenai dominasi karakteristik organisasi pembeli sejalan dengan makrosegmentasi B2B oleh Mora Cortez *et al.* (2021), tetapi berbeda dalam peran lokasi yang tidak membentuk klaster pada penelitian ini dan baru menjadi penentu ketika dimasukkan sebagai variabel pembentuk (S5). Sejalan dengan Kuo *et al.* (2018), penelitian ini juga menemukan sensitivitas terhadap inisialisasi. Sementara itu, pendekatan Kuo *et al.* (2026) menggunakan jarak probabilistik sebagai alternatif jarak Hamming yang berpotensi mengurangi kesamaan jarak antarmodus [22][8], meskipun dampaknya belum diuji pada data penelitian ini.

Keterbatasan pada penggunaan tiga variabel, Status yang hampir konstan, serta sektor yang berasosiasi

kuat dengan skala (V=0,556) sehingga informasinya saling tumpang tindih. Aturan rekomendasi produk masih berdasarkan pengalaman tenaga penjualan dan belum divalidasi menggunakan data penawaran atau transaksi, sedangkan ukuran genset mengacu pada katalog, bukan kebutuhan beban aktual. Penelitian juga belum membandingkan metode klaster lain, membatasi proporsi minimum klaster, maupun memperluas rentang nilai k. Selain itu, data berasal dari satu periode pemetaan pada empat kawasan dan merupakan hasil pemetaan Bagian *Sales*, bukan sensus atau sampel acak perusahaan, sehingga peluang setiap perusahaan untuk masuk ke dalam data tidak diketahui. Dengan demikian, hasil penelitian hanya menggambarkan 364 prospek yang dianalisis dan tidak dapat digeneralisasikan sebagai struktur pasar seluruh perusahaan di kawasan tersebut, kawasan industri lain, atau periode berbeda.

KESIMPULAN

Penelitian berhasil menerapkan *Fuzzy K-Modes* pada 364 prospek di KIC, KIK, KITB, dan KIW dengan konfigurasi $\alpha = 1,3$ dan $k = 10$. Nilai k perlu ditafsirkan hati-hati karena berada pada batas atas rentang pengujian. Model menghasilkan *Silhouette* 0,520, dengan stabilitas rendah hingga moderat (ARI sekitar 0,5) dan 27,5% prospek berada di batas antarklaster. Segmentasi terutama mengikuti sektor dan skala, menghasilkan delapan klaster homogen sektor dan dua klaster campuran. Enam klaster memiliki rekomendasi paket produk relatif seragam, sedangkan dua lainnya memerlukan rekomendasi berdasarkan karakteristik prospek. Hasil bersifat deskriptif untuk data pemetaan satu periode dan tidak dapat digeneralisasikan ke seluruh perusahaan maupun membuktikan peningkatan penjualan atau efisiensi distribusi. Penelitian selanjutnya perlu memperluas rentang k, membandingkan metode klaster kategorik, memvalidasi rekomendasi dengan data penawaran atau transaksi, serta mempertimbangkan penambahan variabel dan pengembangan model melalui pembobotan atribut atau jarak probabilistik.

PERNYATAAN ETIKA DAN KETERSEDIAAN DATA

Data penelitian bersumber dari pemetaan prospek oleh Bagian *Sales* PT United Tractors Tbk Cabang Semarang dengan persetujuan resmi perusahaan. Sebagian penulis berafiliasi dengan perusahaan dan terlibat dalam penyediaan data serta penyusunan aturan rekomendasi produk. Identitas perusahaan disamarkan dan hasil disajikan secara agregat untuk menjaga kerahasiaan. Data bersifat privat dan tidak tersedia untuk publik. Tabel pendukung dan indikator evaluasi seluruh model kandidat disediakan sebagai materi suplementer.

REFERENCES

[1] BPS Provinsi Jawa Tengah, "Perekonomian Jawa Tengah triwulan II-2026 tumbuh 5,71 persen," Badan Pusat Statistik Provinsi Jawa Tengah. Accessed: Oct. 14, 2026. [Online]. Available: https://jateng.bps.go.id/id/pressrelease/2026/08/05/1683/perekonomian-jawa-tengah-triwulan-ii-2026-tumbuh-5-71-persen.html

[2] S. Sunarti, W. Tyas, M. Helmi, L. A. Dewi, and N. C. Medina, "Impact of Industrial Estate Development on Infrastructure Sustainability in the Wijayakusuma Industrial Area, Semarang," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1353, 2024, doi: 10.1088/1755-1315/1353/1/012003.

[3] M. H. Nguyen, H. Vu, T. T. M. Tran, and S. A. Pham, "Machinery Usage and Productivity in Manufacturing: Firm-Level Matter in Developing Countries," *Emerging Science Journal*, 2025, doi: 10.28991/esj-2025-09-04-015.

[4] Q.-X. Zhao, X.-D. Zhang, and P. Wang, "Multi-Type Equipment Selection and Quantity Decision Optimization in Intelligent Warehouse," *IEEE Access*, vol. 12, pp. 63515–63527, 2024, doi: 10.1109/access.2024.3395288.

[5] R. Mora Cortez, A. Højbjerg Clarke, and P. V. Freytag, "B2B market segmentation: A systematic review and research agenda," *J. Bus. Res.*, vol. 126, pp. 415–428, 2021, doi: 10.1016/j.jbusres.2020.12.070.

[6] M.-G. Wu, P. Andreev, and M. Benyoucef, "Profiling before scoring: a two-stage predictive model for B2B lead prioritization," *Journal of Personal Selling & Sales Management*, vol. 46, pp. 313–329, 2026,

doi: 10.1080/08853134.2026.2633231.
[7] S. Bouguelia and H. Kheddouci, "A New Method of B2B Customer Segmentation Based on Firmographic Data, and RFM and Graph Models," *2024 IEEE International Conference on e-Business Engineering (ICEBE)*, pp. 81–86, 2024, doi: 10.1109/icebe62490.2024.00021.
[8] R. J. Kuo, M. Cendana, T. P. Q. Nguyen, and F. E. Zulvia, "A multivariate fuzzy weighted K-modes algorithm with probabilistic distance for categorical data," *Journal of ICT Research and Applications*, vol. 18, no. 2, pp. 93–107, 2024, doi: 10.5614/itbj.ict.res.appl.2023.18.2.2.
[9] T. Liu, H. Yu, and R. H. Blair, "Stability estimation for unsupervised clustering: A review," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 14, 2022, doi: 10.1002/wics.1575.
[10] A. Ahmad and S. S. Khan, "Survey of State-of-the-Art Mixed Data Clustering Algorithms," *IEEE Access*, vol. 7, pp. 31883–31902, 2019, doi: 10.1109/ACCESS.2019.2903568.
[11] Z. Huang, "Extensions to the k-means algorithm for clustering large data sets with categorical values," *Data Min. Knowl. Discov.*, vol. 2, no. 3, pp. 283–304, 1998, doi: 10.1023/A:1009769707641.
[12] J. Z. Huang and M. K. Ng, "A fuzzy k-modes algorithm for clustering categorical data," *IEEE Transactions on Fuzzy Systems*, vol. 7, no. 4, pp. 446–452, 1999, doi: 10.1109/91.784206.
[13] D. B. M. Maciel, G. J. A. Amaral, R. M. C. R. de Souza, and B. A. Pimentel, "Multivariate fuzzy k-modes algorithm," *Pattern Analysis and Applications*, vol. 20, pp. 59–71, 2017, doi: 10.1007/s10044-015-0465-3.
[14] D.-W. Kim, K.-H. Lee, and D. Lee, "Fuzzy clustering of categorical data using fuzzy centroids," *Pattern Recognit. Lett.*, vol. 25, no. 11, pp. 1263–1271, 2004, doi: 10.1016/j.patrec.2004.04.004.
[15] G. Preud'homme *et al.*, "Head-to-head comparison of clustering methods for heterogeneous data: a simulation-driven benchmark," *Sci. Rep.*, vol. 11, 2021, doi: 10.1038/s41598-021-83340-8.
[16] M. van de Velden, A. I. D'Enza, and F. Palumbo, "Cluster Correspondence Analysis," *Psychometrika*, vol. 82, no. 1, pp. 158–185, 2017, doi: 10.1007/s11336-016-9514-0.
[17] D. A. Linzer and J. B. Lewis, "poLCA: An R Package for Polytomous Variable Latent Class Analysis," *J. Stat. Softw.*, vol. 42, no. 10, pp. 1–29, 2011, doi: 10.18637/jss.v042.i10.
[18] R. J. Kuo, Y. Potti, and F. E. Zulvia, "Application of metaheuristic based fuzzy K-modes algorithm to supplier clustering," *Comput. Ind. Eng.*, vol. 120, pp. 298–307, 2018, doi: 10.1016/j.cie.2018.04.050.
[19] T. Dinh *et al.*, "Categorical data clustering: 25 years beyond K-modes," *Expert Syst. Appl.*, vol. 272, p. 126608, 2025, doi: 10.1016/j.eswa.2025.126608.
[20] H. Malikhatin, A. Rusgiyono, and D. A. Maruddani, "Penerapan k-Modes clustering dengan validasi Dunn Index pada pengelompokan karakteristik calon TKI menggunakan R-GUI," *Jurnal Gaussian*, vol. 10, no. 3, pp. 359–366, 2021, doi: 10.14710/j.gauss.10.3.359-366.
[21] J. Wang, L. Yu, P. Si, G. Wu, and Z. Yang, "Research on improvement of fuzzy K-modes clustering algorithm for categorical data," in *Proceedings of SPIE 13213, ICIPAI 2024*, 2024. doi: 10.1117/12.3035108.
[22] R. J. Kuo, M. Cendana, T. P. Q. Nguyen, and F. E. Zulvia, "Market segmentation using a PSO-based multivariate fuzzy weighted fuzzy K-modes algorithm with probabilistic distance," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 34, no. 3, pp. 351–373, 2026, doi: 10.1142/S0218488526500157.
[23] K. Kim, "A weighted k-modes clustering using new weighting method based on within-cluster and between-cluster impurity measures," *Journal of Intelligent & Fuzzy Systems*, vol. 32, no. 1, pp. 979–990, 2017, doi: 10.3233/JIFS-16157.
[24] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. New York, NY, USA: Wiley, 1990.
[25] X. Gao and S. Wu, "CUBOS: An internal cluster validity index for categorical data," *Tehnički vjesnik*, vol. 26, no. 2, pp. 486–494, 2019, doi: 10.17559/TV-20190109015453.