

Implementasi Algoritma K-Means Clustering dalam Pengelompokan Potensi Lokasi Usaha Waralaba Berdasarkan Karakteristik Wilayah

Nining Suryani¹, Evy Priyanti², Girman Sihombing³

^{1,2,3}Universitas Bina Sarana Informatika

e-mail: ¹nining.nns@bsi.ac.id, ²evy.evp@bsi.ac.id, ³girman.gsh@bsi.ac.id

Diterima	Direvisi	Disetujui
06-03-2026	01-07-2026	15-07-2026

Abstrak - Penelitian ini bertujuan untuk menerapkan algoritma *K-Means Clustering* guna mengelompokkan potensi lokasi bisnis *franchise* Waralaba berdasarkan karakteristik wilayah di DKI Jakarta. Metode yang digunakan dalam penelitian ini adalah pendekatan kuantitatif dengan teknik data mining yang melibatkan 50 data cabang Waralaba. Variabel yang diambil dalam penelitian ini mencakup keramaian, renovasi, lalu lintas, keamanan, tenaga kerja, bahan baku, transportasi, daya beli, luas daerah, usia penduduk, dan prospek usaha. Proses penelitian dibagi menjadi beberapa fase, yaitu pengambilan data, preprocessing data, mengubah data kategorikal menjadi numerik, normalisasi menggunakan metode Min-Max, menentukan jumlah kluster dengan metode *elbow*, menerapkan algoritma K-Means, dan mengevaluasi hasil pengelompokan. Melalui metode *elbow* didapatkan jumlah kluster optimal yang terdiri dari tiga kluster. Hasil pengelompokan menunjukkan bahwa kluster 1 adalah kelompok dengan potensi usaha tertinggi karena memiliki rata-rata indikator yang paling stabil dan tinggi. Kluster 0 dikategorikan sebagai kelompok dengan potensi menengah karena memiliki keramaian dan prospek yang tinggi, tetapi terhambat oleh faktor lalu lintas dan renovasi. Sedangkan, kluster 2 termasuk kelompok dengan potensi rendah karena tingkat keramaian dan prospek yang lebih rendah meskipun kondisi renovasinya cukup baik. Penelitian ini diharapkan dapat membantu pelaku bisnis dalam menentukan lokasi usaha secara lebih objektif, efektif, dan berdasarkan data.

Kata Kunci: klaster, waralaba, k-means

Abstracts - This study aims to apply the *K-Means Clustering* algorithm to group potential franchise business locations based on regional characteristics in DKI Jakarta. The method used in this study is a quantitative approach with data mining techniques involving 50 Franchise branch data. The variables taken in this study include crowds, renovations, traffic, security, labor, raw materials, transportation, purchasing power, area, population age, and business prospects. The research process is divided into several phases, namely data collection, data preprocessing, converting categorical data into numeric, normalization using the Min-Max method, determining the number of clusters with the elbow method, applying the K-Means algorithm, and evaluating the clustering results. Through the elbow method, the optimal number of clusters consisting of three clusters was obtained. The clustering results show that cluster 1 is the group with the highest business potential because it has the most stable and high average indicators. Cluster 0 is categorized as a group with medium potential because it has high crowds and prospects, but is hampered by traffic and renovation factors. Meanwhile, cluster 2 is considered a low-potential group due to lower traffic levels and prospects, despite its relatively good renovation condition. This research is expected to help business owners determine business locations more objectively, effectively, and based on data.

PENDAHULUAN

Perkembangan teknologi informasi dan persaingan bisnis yang semakin ketat menuntut pelaku usaha untuk mampu mengambil keputusan secara tepat dan strategis, khususnya dalam menentukan lokasi usaha. Lokasi usaha merupakan salah satu faktor penting yang memengaruhi keberhasilan suatu bisnis karena berkaitan langsung dengan akses konsumen, potensi pasar, tingkat keramaian, daya beli masyarakat, serta kemudahan distribusi. Pemilihan lokasi yang

kurang tepat dapat menyebabkan rendahnya minat konsumen dan menurunkan potensi keuntungan usaha. (Rahma, I., Arhandi, P. P., & Firdausi, 2020)

Sebelas variabel tersebut dipilih karena mampu memberikan gambaran yang komprehensif mengenai kelayakan dan potensi keberhasilan suatu usaha atau waralaba pada suatu lokasi. Variabel keramaian, lalu lintas, daya beli, luas daerah, dan usia penduduk digunakan untuk mengukur potensi pasar dan karakteristik konsumen. Sementara itu, variabel keamanan, tenaga kerja, bahan baku, dan transportasi berperan dalam menilai kelancaran operasional usaha. Variabel renovasi digunakan untuk memperkirakan

kebutuhan investasi awal yang harus dikeluarkan, sedangkan variabel prospek memberikan gambaran mengenai peluang pertumbuhan usaha di masa depan. Dengan mempertimbangkan seluruh variabel tersebut secara bersamaan, analisis yang dilakukan menjadi lebih objektif, menyeluruh, dan dapat membantu pengambilan keputusan bisnis yang lebih tepat serta meminimalkan risiko kegagalan usaha.

Pada bisnis waralaba makanan dan minuman seperti Waralaba, pemilihan lokasi menjadi faktor yang sangat menentukan keberhasilan usaha. Gerai yang berada pada lokasi strategis cenderung memiliki tingkat kunjungan konsumen yang lebih tinggi dibandingkan dengan lokasi yang kurang mendukung. Oleh karena itu, diperlukan analisis yang mampu membantu pelaku usaha dalam menentukan lokasi yang memiliki potensi terbaik secara objektif dan berbasis data.

Dalam praktiknya, proses penentuan lokasi usaha masih banyak dilakukan secara manual berdasarkan intuisi atau pengalaman pemilik usaha. Pendekatan tersebut dinilai kurang efektif karena tidak mempertimbangkan seluruh karakteristik wilayah secara menyeluruh. Beberapa variabel yang dapat digunakan dalam menentukan potensi lokasi usaha antara lain tingkat keramaian, kondisi lalu lintas, keamanan, daya beli masyarakat, transportasi, serta kondisi lingkungan sekitar. (Gari, E. & Sari, 2025)

Perkembangan bidang data mining memberikan solusi dalam pengolahan data untuk menghasilkan informasi yang dapat digunakan sebagai dasar pengambilan keputusan. Salah satu teknik dalam data mining yang banyak digunakan adalah clustering. Teknik clustering digunakan untuk mengelompokkan data berdasarkan tingkat kemiripan karakteristik tertentu sehingga dapat menghasilkan segmentasi data yang lebih terstruktur. (Nugroho et al., 2022) salah satunya pada pengelompokan provinsi berdasarkan pengguna internet menjadi tiga *cluster* untuk mengetahui provinsi-provinsi mana yang memiliki infrastruktur yang baik, sehingga dapat mendukung penggunaan internet dapat menggunakan metode k-means. (Pramudita & Sumargo, 2019)

Data dapat dibagi menjadi kategori item dengan properti unik menggunakan pengelompokan dengan metode k-means untuk mengidentifikasi perilaku jangka pendek (interval pengamatan per jam) dan perilaku jangka panjang (durasi pengamatan 15 hari) dari unit nirkabel dengan mengekstrak bentuk sel. (Kunchakuri et al., 2025)

Pengelompokan pola pembelian e-commerce di wilayah kecil dengan segmentasi kohort dari data perilaku e-commerce dari berbagai kategori toko dan riwayat pembelian dapat menggunakan metode k-means. (Tabianan et al., 2022)

Peningkatan kualitas pengelompokan sampai dengan 41,85% dengan Penggunaan k-means pada big data pada sampel sintesis memperlihatkan bahwa k-means sangat bermanfaat untuk big data. (Moodi & Saadatfar, 2021)

Metode k-means bekerja dengan cara membagi data ke dalam beberapa cluster berdasarkan kedekatan jarak antar data. Dalam penelitian terkait lokasi usaha,

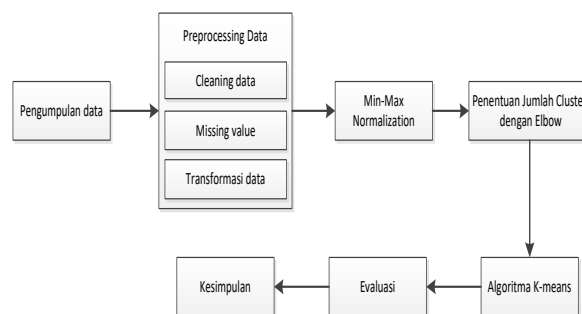
metode K-Means telah banyak digunakan untuk mengidentifikasi wilayah potensial, segmentasi pasar, hingga pengelompokan wilayah usaha berdasarkan karakteristik tertentu. (Ade Wikarta & Defiyanti, 2024) dan dapat digunakan untuk mengelompokkan pendampingan UMKM secara tepat berdasarkan omset dan aset usaha. (W. Sudrajat et al., 2022).

Selain itu, penelitian internasional yang dilakukan oleh menggunakan pendekatan K-Means dalam pengambilan keputusan lokasi internasional membuktikan bahwa clustering mampu menghasilkan keputusan lokasi yang lebih objektif dan dinamis dibandingkan metode konvensional. (Marcelina et al., 2023)

Berdasarkan permasalahan tersebut, penelitian ini dilakukan untuk menerapkan algoritma K-Means Clustering dalam pengelompokan potensi lokasi usaha berdasarkan karakteristik wilayah. Dengan adanya penelitian ini diharapkan dapat membantu proses pengambilan keputusan dalam menentukan lokasi usaha yang strategis serta memberikan rekomendasi wilayah yang memiliki potensi usaha tinggi.

METODE PENELITIAN

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan *data mining*. Metode yang digunakan adalah teknik *clustering* untuk mengelompokkan data lokasi usaha berdasarkan karakteristik wilayah. Penelitian ini bertujuan untuk menemukan pola pengelompokan lokasi usaha tanpa menggunakan label. Adapun tahapan dari penelitian ini sebagai berikut:



Gambar 1. Tahapan penelitian

1. Pengumpulan data

Data yang digunakan adalah studi kasus data waralaba yang memiliki 50 cabang di wilayah DKI Jakarta

Jumlah 50 data pada penelitian ini dianggap memadai karena tujuan penelitian adalah melakukan eksplorasi pola dan pengujian awal model klastering, bukan menghasilkan estimasi populasi dengan tingkat presisi statistik yang tinggi. Literatur terkini pada bidang machine learning dan decision support system menunjukkan bahwa ukuran sampel yang memadai bergantung pada kompleksitas masalah, jumlah variabel, dan tujuan analisis. Untuk studi eksploratif dan pengembangan model awal, dataset berukuran puluhan

observasi masih dapat digunakan untuk mengidentifikasi pola awal, mengevaluasi kelayakan metode, dan menghasilkan temuan pendahuluan yang akan diverifikasi pada penelitian lanjutan dengan data yang lebih besar

2. Processing data
Setelah data terkumpul, maka dilakukan pemrosesan data dengan cara cleaning data, missing value dan transformasi data
3. Tahap berikutnya adalah melakukan normalisasi data menggunakan metode Min-Max agar semua variabel berada dalam skala yang sama.
4. Tahap berikutnya adalah menentukan jumlah cluster dengan metode elbow
5. Penerapan algoritma k-means
6. Tahap evaluasi
7. Kesimpulan

HASIL DAN PEMBAHASAN

1. Pengumpulan data

Menurut (Rajput et al., 2023) ukuran sampel yang layak dalam penelitian berkisar antara 30 hingga 50 responden atau unit analisis. Sementara itu, beberapa penelitian pada bidang sistem pendukung keputusan dan data mining menunjukkan bahwa jumlah sampel minimal 30–50 data sudah dapat digunakan untuk membangun model dan melakukan pengujian awal secara memadai. Oleh karena itu, penggunaan 50 data dalam penelitian ini dinilai telah memenuhi kriteria kecukupan sampel untuk menghasilkan analisis yang representatif terhadap kondisi wilayah DKI Jakarta. Penggunaan 50 data dalam penelitian ini dianggap cukup representatif karena data tersebut telah mencakup berbagai karakteristik wilayah dan kondisi

usaha yang terdapat di DKI Jakarta. Dalam penelitian eksploratif maupun pengembangan model pengambilan keputusan, jumlah sampel sebanyak 30–50 umumnya telah memenuhi batas minimum untuk menghasilkan analisis statistik yang memadai dan menggambarkan variasi data yang ada. Selain itu, DKI Jakarta terdiri dari beragam kawasan dengan tingkat keramaian, daya beli, akses transportasi, kepadatan penduduk, dan aktivitas ekonomi yang berbeda-beda. Dengan pemilihan 50 lokasi atau unit observasi yang tersebar pada berbagai wilayah administratif dan karakteristik lingkungan yang beragam, data yang diperoleh mampu merepresentasikan kondisi umum wilayah DKI Jakarta. Jumlah tersebut juga telah mencukupi untuk mengidentifikasi pola hubungan antarvariabel, mengurangi pengaruh data ekstrem, serta menghasilkan rekomendasi yang lebih stabil dan dapat diandalkan dalam proses analisis maupun pengambilan Keputusan.

Label prospek digunakan hanya sebagai acuan interpretasi dan evaluasi hasil pengelompokan. Pembentukan kelompok tetap didasarkan pada kesamaan karakteristik dari sebelas variabel lainnya, sehingga informasi prospek tidak menjadi faktor yang menentukan terbentuknya klaster. Dengan cara ini, interpretasi terhadap hasil klaster dapat dilakukan tanpa menimbulkan bias logis atau circular reasoning.

Varibel yang digunakan berdasarkan dari dataset lokasi usaha dengan karakteristik wilayah yaitu Keramaian, Renovasi, Lalu lintas, Keamanan, Tenaga kerja, Bahan baku, Transportasi, Daya beli, Luas daerah, Usia penduduk, Prospek usaha, adapun untuk 10 baris data pertama dataset diluat pada tabel 1 dengan penjelasan sebagai berikut:

Tabel 1. Data Waralaba

Waralaba	Keramaian	Reno vasi	Lalu Lintas	Keamanan	Tenaga Kerja	Bahan Baku	Transpor tasi	Daya Beli	Luas Daerah	Usia Penduduk	Prospek
Waralaba Danau Sunter	Tinggi	Tinggi	Tidak macet	Aman	Mudah	Mudah	Mudah	Tinggi	Besar	Tinggi	Sangat Prospek
Waralaba Kelapa Gading Boulevard	Tinggi	Tinggi	Tidak macet	Aman	Mudah	Mudah	Mudah	Tinggi	Besar	Tinggi	Sangat Prospek
Waralaba Kelapa Gading Hybrida	Tinggi	Tinggi	Tidak macet	Aman	Mudah	Mudah	Mudah	Tinggi	Besar	Tinggi	Sangat Prospek
Mxue Mall of Indonesia	Tinggi	Tinggi	Macet	Aman	Mudah	Mudah	Mudah	Tinggi	Besar	Tinggi	Prospek
Waralaba Muara Karang	Tinggi	Sedang	Macet	Aman	Mudah	Mudah	Mudah	Tinggi	Besar	Tinggi	Sangat Prospek
Waralaba Pademangan	Tinggi	Tinggi	Tidak macet	Aman	Mudah	Mudah	Mudah	Sedang	Besar	Tinggi	Sangat Prospek
Waralaba Pademangan Timur	Tinggi	Tinggi	Tidak macet	Aman	Mudah	Mudah	Mudah	Tinggi	Besar	Tinggi	Prospek
Waralaba Cilincing	Sedang	Sedang	Macet	Aman	Mudah	Mudah	Mudah	Tinggi	Besar	Tinggi	Prospek
Waralaba Teluk Gong	Sedang	Sedang	Tidak macet	Aman	Mudah	Mudah	Mudah	Tinggi	Sedang	Tinggi	Prospek

2. Preprocessing data

a. Tahapan ini terdiri dari pembersihan data

Pada tahap ini penulis melakukan cleaning data dengan melengkapi data missing value yang ada.

b. Tahap transformasi data

Data pada penelitian ini sebagian besar berbentuk **kategorikal**, seperti kategori tingkat keramaian, keamanan, daya beli, atau prospek. Meskipun kategori-kategori tersebut mengandung informasi penting, sebagian besar metode analisis data, algoritma machine learning, dan teknik klastering memerlukan input dalam bentuk numerik agar dapat melakukan perhitungan matematis, seperti pengukuran jarak, kemiripan, rata-rata, maupun pembentukan kelompok. Oleh karena itu, data kategorikal perlu ditransformasikan ke dalam bentuk numerik melalui proses pengkodean (encoding) sehingga dapat diolah oleh algoritma tanpa menghilangkan makna informasi yang terkandung di dalamnya.

Transformasi ini tidak dimaksudkan untuk mengubah sifat dasar data menjadi data kuantitatif murni, melainkan sebagai representasi numerik yang memungkinkan proses komputasi dilakukan secara efektif. Selain itu, penggunaan skema pengkodean yang konsisten memastikan bahwa informasi asli pada data kategorikal tetap terjaga sehingga hasil analisis dapat diinterpretasikan sesuai dengan konteks bisnis dan karakteristik variabel yang diteliti.

Karena nilai numerik yang diberikan merupakan hasil pengkodean kategori, interpretasi terhadap angka tersebut tidak didasarkan pada besarnya nilai secara absolut, melainkan pada kategori yang diwakilinya. Dengan demikian, angka yang digunakan berfungsi sebagai representasi komputasional, bukan sebagai ukuran kuantitatif yang memiliki satuan pengukuran tertentu.

Pada tahap ini data yang awalnya berbentuk kategorikal kemudian ditransformasikan ke dalam bentuk numerik. Pemberian bobot 3, 2, dan 1 dilakukan berdasarkan skala ordinal, dimana kategori dengan kondisi terbaik diberikan skor tertinggi dan kategori dengan kondisi terendah diberikan skor terendah (Sugiyono, 2019). Pendekatan ini digunakan karena setiap variabel memiliki tingkatan atau urutan

yang jelas, sebagai berikut:

Tabel 2. Daftar Katagori

Kolom	Kategori
Keramaian	Tinggi = 3, Sedang = 2, Rendah = 1
Renovasi	Tinggi = 3, Sedang = 2, Rendah = 1
Lalu Lintas	Tidak macet = 3, Sedang=2, Macet = 1
Keamanan	Aman = 3, Cukup Aman=2, Tidak aman = 1
Tenaga Kerja	Mudah = 3, Sedang=2, Sulit = 1
Bahan Baku	Mudah = 3, Sedang=2, Sulit = 1
Transportasi	Mudah = 3, Sedang=2, Sulit = 1
Daya Beli	Tinggi = 3, Sedang = 2, Rendah = 1
Luas Daerah	Besar = 3, Sedang = 2, Kecil = 1
Usia Penduduk	Tinggi = 3, Sedang = 2, Rendah = 1
Prospek	Sangat Prospek = 3, Prospek = 2, Tidak Prospek = 1

Berdasarkan tabel 2 diatas, masing-masing alternatif atau kolom memiliki katagori, keramaian dengan katagori tinggi, sedang dan rendah. Renovasi dengan katagori biaya tinggi, biaya sedang, biaya rendah. Lalu lintas dengan katagori tingkat kemacetan yaitu macet, sedang dan tidak macet. Keamanan dengan katagori aman, cukup aman dan tidak aman. Bahan baku dengan katagori mudah didapat, sedang dan sulit. Trasnsportasi dengan katagori mudah, sedang, sulit. Daya beli dengan katagori tinggi, sedang dan rendah. Luas daerah dengan katagori besar, sedang dan kecil. Usia penduduk dengan katagori usia produktif yaitu tinggi, sedang dan rendah. Prospek dengan katagori tingkat potensialnya yaitu prospek, tidak prospek dan sangat prospek. Berdasarkan data set yang ada maka data yang berbentuk katagorikal di transformasikan kedalam bentuk numerik.

Adapun hasil transformasi data untuk 10 baris data pertama dataset dilihat pada tabel 3 sebagai berikut:

Tabel 3. Hasil Transformasi

Waralaba	Keramaian	Reno vasi	Lalu Lintas	Keamanan	Tenaga Kerja	Bahan Baku	Transpor tasi	Daya Beli	Luas Daerah	Usia Penduduk	Prospek
Waralaba Danau Sunter	3	3	3	3	3	3	3	3	3	3	3
Waralaba Kelapa Gading Boulevard	3	3	3	3	3	3	3	3	3	3	3
Waralaba Kelapa Gading Hybrida	3	3	3	3	3	3	3	3	3	3	3
Waralaba Mall of Indonesia	3	3	1	3	3	3	3	3	3	3	2
Waralaba Muara Karang	3	2	1	3	3	3	3	3	3	3	3
Waralaba Pademangan	3	3	3	3	3	3	3	2	3	3	3

Waralaba Pademangan Timur	3	3	3	3	3	3	3	3	3	3	2
Waralaba Cilincing	2	2	1	3	3	3	3	3	3	3	2
Waralaba Teluk Gong	2	2	3	3	3	3	3	3	2	3	2

3. Tahap normalisasi

Setelah data ditransformasikan selanjutnya penulis melakukan normalisasi dengan metode min-max normalisasi dengan rumus

$$X^n = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Adapun hasil normalisasi data untuk 10 baris data pertama sebagai berikut

Tabel 4 Hasil Normalisasi

Waralaba	Keramaian	Renovasi	Lalu Lintas	Keamanan	Tenaga Kerja	Bahan Baku	Transportasi	Daya Beli	Luas Daerah	Usia Penduduk	Prospek
Waralaba Danau Sunter	1	1	1	1	1	1	1	1	1	1	1
Waralaba Kelapa Gading Boulevard	1	1	1	1	1	1	1	1	1	1	1
Waralaba Kelapa Gading Hybrida	1	1	1	1	1	1	1	1	1	1	1
Waralaba Mall of Indonesia	1	1	0	1	1	1	1	1	1	1	0.5
Waralaba Muara Karang	1	0.5	0	1	1	1	1	1	1	1	1
Waralaba Pademangan	1	1	1	1	1	1	1	0.5	1	1	1
Waralaba Pademangan Timur	1	1	1	1	1	1	1	1	1	1	0.5
Waralaba Cilincing	0.5	0.5	0	1	1	1	1	1	1	1	0.5
Waralaba Teluk Gong	0.5	0.5	1	1	1	1	1	1	0.5	1	0.5
Waralaba Walang Baru Semper	1	1	1	1	1	1	1	1	0.5	1	0.5

4. Menentukan jumlah kluster

Untuk menentukan jumlah kluster, penulis menggunakan metode elbow dengan rumus

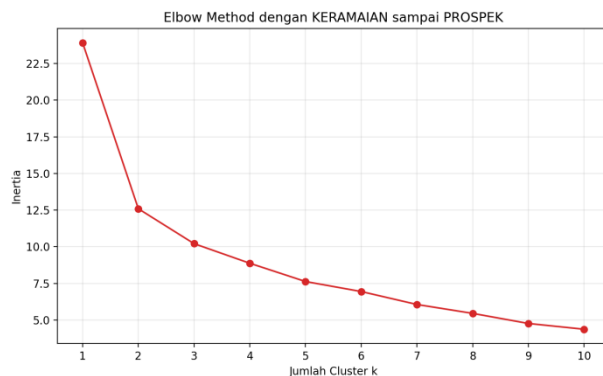
$$SSE = \sum_{k=1}^n x_i \sum_{i \in C_k} (x_i - C_k)^2 \quad (2)$$

Adapun hasilnya pada tabel berikut:

Tabel 5. Hasil Inertia

k	Inertia
1	23.91837
2	12.59007
3	10.20784
4	8.871324
5	7.624573
6	6.941288
7	6.057738
8	5.450595
9	4.761905
10	4.368452

Berdasarkan tabel 5, nilai inertia mengalami penurunan yang cukup signifikan dari k=1 sebesar 23.91837 menjadi 12.59007 pada k=2. Penurunan juga masih terlihat pada k=3 dengan nilai inertia 10.20784. Setelah k=3, penurunan inertia mulai melandai, yaitu menjadi 8.871324 pada k=4 dan terus menurun secara bertahap hingga k=10 sebesar 4.368452, hasil ini dapat divisualisasikan dengan penggambaran grafik dibawah ini.



Gambar 2. Grafik Hasil elbow

Berdasarkan grafik 2 diatas dapat terlihat titik paling siku berada di sekitar k=2 atau k=3 dimana setelahnya inertia mulai terjadi penuruna sehingga lebih landai, karena hal tersebut, maka penulis menentukan jumlah kluster adalah 3 sebagai jumlah kluster yang paling optimal.

Untuk mengevaluasi apakah kluster yang telah terbentuk benar-benar terpisah dengan baik penulis menggunakan Silhouette Coefficient dengan rumus

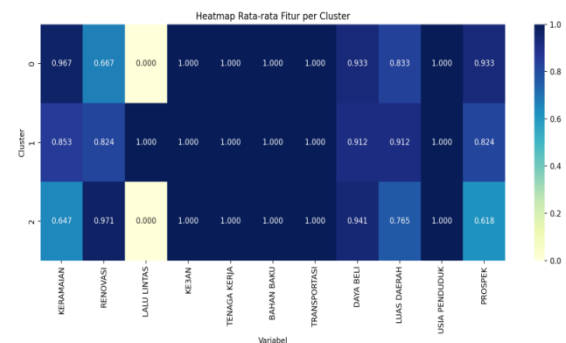
$$S_i = \frac{b_i - s_i}{\max(a_i, b_i)} \quad (3)$$

Silhouette Score semakin mendekati nilai 1 maka kualitas pemisahan cluster semakin baik.(Julianti &

Widiartha, 2024) Berdasarkan hasil perhitungan dengan Silhouette Score dari dataset maka diperoleh 0.329. berdasarkan tabel rentang nilai Silhouette(R. Sudrajat et al.) Nilai 0,329 berada pada rentang 0,26 – 0,50, menunjukkan adanya struktur cluster yang cukup, dimana objek dalam kluster memiliki kemiripan yang cukup tinggi.

5. Penerapan algoritma K-means

Setelah didapatkan jumlah kluster maka diperoleh hasil dari terbagi menjadi tiga kelompok yaitu kluster 0 potensi menengah, kluster 1 potensi tinggi dan kluster 2 potensi rendah. selanjutnya menghitung masing-masing centroid dan dikelompokkan ke dalam kluster berdasarkan jarak terdekat dengan centroid, berikut jumlah anggota tiap kluster yang telah terbentuk dibawah ini:



Gambar 3. Hasil centroid

Hasil centroid menunjukkan adanya perbedaan karakteristik pada masing-masing cluster. **Cluster 0** memiliki rata-rata nilai KERAMAIAN sebesar **0,967**, RENOVASI sebesar **0,667**, LALU LINTAS sebesar **0,000**, dan PROSPEK sebesar **0,933**. Nilai tersebut menunjukkan bahwa cluster ini didominasi oleh outlet yang memiliki tingkat keramaian dan prospek yang tinggi, namun kondisi renovasi relatif lebih rendah serta faktor lalu lintas cenderung rendah. Dengan kata lain, cluster ini dapat diartikan sebagai kelompok outlet yang cukup potensial, tetapi masih memiliki beberapa keterbatasan pada aspek pendukung tertentu.

Lokasi yang mengalami kemacetan parah tetap dapat memiliki prospek usaha yang tinggi karena dampak negatif kemacetan sering kali dikompensasi oleh besarnya potensi pasar yang tersedia. Kemacetan umumnya terjadi pada kawasan dengan kepadatan penduduk tinggi, pusat aktivitas ekonomi, perkantoran, pusat perbelanjaan, atau kawasan transit yang ramai. Kondisi tersebut mencerminkan tingginya jumlah calon konsumen yang berada di sekitar lokasi usaha. Bagi sebagian jenis usaha, seperti makanan dan minuman, ritel, minimarket, apotek, atau layanan sehari-hari, jumlah konsumen potensial sering kali memiliki pengaruh yang lebih besar terhadap keberhasilan bisnis dibandingkan kemudahan akses kendaraan.

Cluster 1 memiliki nilai rata-rata KERAMAIAN sebesar **0,853**, RENOVASI sebesar **0,824**, LALU LINTAS sebesar 1,000, dan PROSPEK sebesar 0,824. Secara umum, cluster ini menunjukkan kondisi yang paling stabil karena hampir seluruh indikator utama berada pada kategori tinggi. Oleh sebab itu, cluster ini dapat ditafsirkan sebagai kelompok outlet dengan performa yang relatif paling baik dan paling unggul dibandingkan cluster lainnya.

Cluster 2 memiliki rata-rata KERAMAIAN sebesar 0,647, RENOVASI sebesar 0,971, LALU LINTAS sebesar 0,000, dan PROSPEK sebesar 0,618. Hasil ini menunjukkan bahwa cluster tersebut memiliki kondisi renovasi yang sangat baik, tetapi tingkat keramaian dan prospeknya relatif lebih rendah. Berdasarkan karakteristik tersebut, cluster ini dapat dipahami sebagai kelompok outlet yang secara fasilitas atau kesiapan internal cukup baik, namun belum didukung oleh potensi pasar yang sama kuatnya seperti cluster lain.

KESIMPULAN

Berdasarkan temuan dari penelitian yang sudah dilakukan, algoritma K-Means Clustering efektif digunakan untuk mengelompokkan potensi lokasi untuk usaha waralaba Waralaba berdasarkan karakteristik area di DKI Jakarta. Proses penelitiannya mencakup beberapa tahap seperti pengumpulan data, pemrosesan awal data, pengubahan data kategorikal menjadi numerik, normalisasi melalui metode Min-Max, analisis untuk menentukan jumlah cluster dengan metode elbow, dan terakhir proses pengelompokan menggunakan algoritma K-Means. Hasil dari metode elbow mengungkapkan bahwa jumlah cluster yang paling sesuai adalah tiga cluster. Setelah proses pengelompokan selesai, data dikelompokkan menjadi cluster 0 yang diwakili oleh potensi menengah, cluster 1 yang menunjukkan potensi tinggi, dan cluster 2 yang memiliki potensi rendah. Cluster 1 menonjol dengan performa yang baik karena hampir semua indikator, seperti kepadatan penduduk, renovasi, lalu lintas, dan peluang usaha, berada dalam kategori tinggi dan stabil. Cluster 0 menggambarkan potensi usaha yang cukup baik namun masih menghadapi beberapa kekurangan di aspek tertentu, sedangkan cluster 2 memiliki fasilitas internal yang baik tetapi belum didukung oleh tingkat kepadatan dan prospek pasar yang memadai.

Secara keseluruhan, metode K-Means Clustering terbukti efektif dalam mendukung pengambilan keputusan untuk menentukan lokasi usaha dengan cara yang lebih objektif dan berdasarkan data, sehingga dapat dijadikan rujukan strategis dan menjanjikan. Berdasarkan temuan dari penelitian yang telah dilaksanakan, berikut adalah beberapa saran yang dapat disampaikan:

Penelitian di masa depan dianjurkan untuk menggunakan data yang lebih banyak dan merentang area yang lebih luas agar hasil pengelompokan

menjadi lebih tepat dan mencerminkan keadaan sebenarnya. Variabel dalam penelitian bisa diperluas dengan menambahkan elemen lain seperti intensitas persaingan bisnis, jumlah penduduk, tingkat penghasilan masyarakat, dan kebiasaan konsumen. Studi selanjutnya bisa melakukan perbandingan antara algoritma K-Means dengan teknik pengelompokan lainnya seperti DBSCAN, Fuzzy C-Means, atau Hierarchical Clustering untuk menemukan metode yang paling efektif.

Hasil dari penelitian ini diharapkan bisa dijadikan pertimbangan bagi para pelaku usaha dan investor dalam memilih tempat usaha yang paling berpotensi. Pengembangan sistem berbasis data mining untuk analisis lokasi usaha perlu dilakukan lebih lanjut agar proses pengambilan keputusan dapat berlangsung secara cepat, objektif, dan efisien.

REFERENSI

- Ade Wikarta, N. C., & Defiyanti, S. (2024). Pengelompokan UMKM berdasarkan kategori usaha dan sebaran wilayah di Jawa Barat dengan clustering K-Means. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13 No. 3, 494–505.
<https://doi.org/https://doi.org/10.23960/jitet.v13i3.6889>
- Gari, E., & Sari, R. (2025). Analisis segmentasi wilayah penjualan menggunakan algoritma K-Means Clustering. *Jurnal Infortech*, 7 No 2. <https://doi.org/https://doi.org/10.31294/infortech.v7i2.11385>
- Julianti, S., & Widiartha, I. M. (2024). Algoritma K-Means untuk Clustering Provinsi di Indonesia Berdasarkan Kasus Stunting. *Jurnal Nasional Teknologi Informasi Dan Aplikasinya*, 2(3), 573–582.
- Kunchakuri, N., Kandari, C., & Patchigolla, V. L. (2025). A Novel Method for Clustering High-Dimensional Data Using K-Means. *First International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT)*, 1197–1202.
<https://doi.org/10.1109/CE2CT64011.2025.10939407>
- Marcelina, D., Kurnia, A., & Terttiaavini. (2023). Analisis klaster kinerja usaha kecil dan menengah menggunakan algoritma K-Means Clustering. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3 No.2.
<https://doi.org/https://doi.org/10.57152/malcom.v3i2.952>
- Moodi, F., & Saadatfar, H. (2021). An improved K-means algorithm for big data. *IET Software*, 16(1), 48–59.
<https://doi.org/https://doi.org/10.1049/sfw2.12032>
- Nugroho, J., Suprpto, Y., & Setijadi, E. (2022). Clustering tingkat risiko klasifikasi lapangan usaha menggunakan metode K-Means. *Jurnal*

- Teknologi Informasi Dan Ilmu Komputer*, 9(3).
<https://doi.org/https://doi.org/10.25126/jtiik.2022915217>)
- Pramudita, D. A., & Sumargo, B. (2019). Pengelompokan Pengguna Internet dengan Metode K-Means Clustering. *Jurnal Statistika Dan Aplikasinya*, 3 No. 1, 1–12.
<https://doi.org/https://doi.org/10.21009/JSA.03101>
- Rahma, I., Arhandi, P. P., & Firdausi, A. T. (2020). Penerapan metode hierarchical clustering dan K-means clustering untuk mengelompokkan potensi lokasi penjualan LinkAja. *Jurnal Informatika Polinema*, 6(1), 15–22.
<https://doi.org/https://doi.org/10.33795/jip.v6i1.287>
- Rajput, D., Chen, C.-C., & Wei-Jen, W. (2023). Evaluation of a decided sample size in machine learning applications. *Journal: BMC Bioinformatics*, Vol. 24, N, 48.
- Sudrajat, R., Hadiana, A. I., & Melina. (n.d.). Evaluasi Kualitas Klaster Wilayah Rawan Bencana Menggunakan K-Means dengan Silhouette dan Elbow Method. *Jurnal Algoritma*, 22(2), 127–139.
- Sudrajat, W., Cholid, I., & Petrus, J. (2022). Penerapan Algoritma K-Means Clustering untuk Pengelompokan UMKM Menggunakan Rapidminer. *JUPITER: Jurnal Penelitian Ilmu Dan Teknologi Komputer*, 14(1), 27–36.
<https://doi.org/https://doi.org/10.5281/4467/5.jupiter.2022.04>
- Sugiyono. (2019). *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Alfabeta.
- Tabianan, K., Velu, S., & Ravi, V. (2022). K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data. *Sustainability (Switzerland)*, 14(12), 1–15.
<https://doi.org/10.3390/su14127243>