



Implementasi Algoritma K-Means Clustering untuk Identifikasi Lokasi Strategis Coffee Shop

Lahuri Gofarana Rohman^{*1}, Cecep Nurul Alam¹, Beki Subaeki²

¹Fakultas Sains Dan Teknologi, Program Studi Teknik Informatika, Universitas Islam Negeri Sunan Gunung Djati, Bandung, Indonesia

²Fakultas Teknik, Program Studi Sistem Informasi, Universitas Sangga Buana YPKP, Bandung, Indonesia

Email: ^{1,*}lahurighofarana@gmail.com, ²cecep@uinsgd.ac.id, ³beki.subaeki@usbykp.ac.id

Email Penulis Korespondensi: lahurighofarana@gmail.com

Abstrak—Maraknya pertumbuhan coffee shop di Kota Bandung menyebabkan persaingan antar pelaku usaha semakin ketat, terutama dalam memilih lokasi yang strategis. Pemilihan lokasi yang kurang tepat dapat berdampak pada rendahnya daya tarik pelanggan dan menurunnya keberlanjutan bisnis. Penelitian ini bertujuan untuk mengidentifikasi wilayah strategis untuk pengembangan coffee shop di Kota Bandung menggunakan algoritma K-Means Clustering berbasis spasial. Data yang digunakan merupakan data lokasi rumah makan aktif yang diperoleh dari portal Open Data Kota Bandung, yang mencakup informasi latitude dan longitude. Algoritma K-Means dengan inisialisasi K-Means++ digunakan untuk mengelompokkan lokasi rumah makan ke dalam tiga kluster, berdasarkan kedekatan geografis. Proses klasterisasi dilakukan dalam dua iterasi, dimulai dengan penentuan centroid awal, perhitungan jarak menggunakan rumus Euclidean, dan pembaruan centroid hingga konvergen. Hasil akhir menunjukkan bahwa wilayah Jl. Aceh Kluster 0 pada koordinat (-6.911431, 107.622713), Jl. Setiabudi Kluster 1 pada koordinat (-6.879891, 107.600774), dan Jl. Kebon Jati Kluster 2 pada koordinat (-6.917228, 107.598990) memiliki potensi strategis berbeda yang sesuai untuk konsep coffee shop tertentu. Evaluasi dilakukan melalui visualisasi sebar spasial, analisis jarak minimum, dan stabilitas kluster. Penelitian ini membuktikan bahwa metode K-Means efektif dalam mendukung pengambilan keputusan berbasis spasial untuk pengembangan bisnis.

Kata Kunci: K-Means Clustering; Lokasi Strategis; Spasial; Coffee Shop; Bandung

Abstract—The rapid growth of coffee shops in Bandung City has led to increasingly fierce competition among business owners, particularly in choosing strategic locations. Inappropriate location selection can negatively impact customer attraction and business sustainability. This study aims to identify strategic areas for coffee shop development in Bandung City using the spatial-based K-Means Clustering algorithm. The data used consists of active food establishment locations obtained from the Open Data Kota Bandung portal, which includes latitude and longitude information. The K-Means algorithm with K-Means++ initialization was used to group the restaurant locations into three clusters based on geographical proximity. The clustering process was carried out in two iterations, beginning with the initial centroid determination, distance calculation using the Euclidean formula, and centroid updates until convergence. Final results show that the areas of Jl. Aceh Cluster 0 at coordinates (-6.911431, 107.622713), Jl. Setiabudi Cluster 1 at coordinates (-6.879891, 107.600774), and Jl. Kebon Jati Cluster 2 at coordinates (-6.917228, 107.598990) have different strategic potentials suited to specific coffee shop concepts. Evaluation was conducted through spatial distribution visualization, minimum distance analysis, and cluster stability. This study confirms that the K-Means method is effective in supporting spatial-based decision-making for business development.

Keywords: K-Means Clustering; Strategic Location; Spasial; Coffee Shop

1. PENDAHULUAN

Lokasi strategis adalah salah satu kunci keberhasilan dalam menjalankan bisnis, termasuk coffee shop, yang kian menjamur di perkotaan seperti Bandung. Pemilihan lokasi yang tepat dapat memengaruhi daya tarik pelanggan, tingkat penjualan, hingga keberlanjutan bisnis secara keseluruhan [1],[2]. Bandung, dengan jumlah penduduk yang terus bertambah dan budaya konsumsi kopi yang tinggi, menjadi salah satu kota potensial untuk pengembangan bisnis coffee shop. Dalam konteks ini, identifikasi lokasi strategis memerlukan pendekatan yang berbasis data, salah satunya adalah dengan menerapkan algoritma K-Means Clustering [3]. K-Means Clustering adalah metode pengelompokan data yang telah digunakan secara luas dalam berbagai bidang, termasuk ritel dan jasa, untuk memetakan pola distribusi demografis dan spasial [4]. Algoritma ini bekerja dengan membagi data ke dalam beberapa kluster berdasarkan kesamaan karakteristik seperti kepadatan populasi, aksesibilitas, hingga jarak ke pusat keramaian [3]. Dalam konteks pemilihan lokasi coffee shop, penelitian menunjukkan bahwa K-Means mampu memberikan rekomendasi berbasis data yang dapat membantu pemilik usaha dalam menentukan lokasi yang memiliki potensi pasar tinggi [5].

Namun, implementasi metode ini memerlukan pemahaman mendalam terhadap data yang digunakan. Dalam hal ini, data dari Open Data Bandung yang mencakup atribut seperti koordinat geografis (latitude, longitude), nama rumah makan, dan alamat dapat dimanfaatkan. Data ini memungkinkan identifikasi pola distribusi geografis coffee shop di Bandung, serta analisis kedekatan lokasi dengan fasilitas umum, pusat keramaian, atau pesaing. Analisis berbasis data ini dapat membantu pemilik usaha menghindari wilayah yang terlalu jenuh dengan kompetitor [6],[7]. Teknologi pemrosesan data spasial kini semakin berkembang dan menjadi alat penting dalam mendukung pengambilan keputusan strategis di berbagai bidang, termasuk dalam bisnis kuliner seperti coffee shop. Salah satu metode populer yang digunakan untuk menganalisis dan mengelompokkan data lokasi adalah algoritma K-Means Clustering. Beberapa penelitian telah membuktikan efektivitas algoritma ini dalam klasifikasi wilayah strategis berdasarkan parameter spasial seperti latitude dan longitude. Dalam penelitian oleh Fathurrohman, metode K-Means digunakan untuk mengelompokkan lokasi coffee shop berdasarkan preferensi dan karakteristik konsumen, menghasilkan visualisasi kluster yang menunjukkan konsentrasi tinggi di area kampus dan pusat perbelanjaan [8]. Penelitian lain yang dilakukan oleh Prasetyo et al. menerapkan K-



Means pada studi kasus Tetra Coffeeshop untuk menganalisis preferensi menu berdasarkan lokasi konsumen. Studi ini berhasil membentuk klaster berdasarkan popularitas menu di area tertentu yang berdekatan dengan pusat bisnis dan perkantoran [5]. Alexander et al. mengembangkan sistem rekomendasi berbasis web untuk pemilihan lokasi café di Kota Pekanbaru menggunakan metode SAW dan Google Maps, yang ditujukan untuk membantu mahasiswa dalam menentukan tempat yang sesuai untuk kegiatan sosial dan akademik, dan menemukan bahwa lokasi menjadi faktor dominan dalam klasifikasi [9]. Di sisi lain, integrasi K-Means dengan algoritma FP-Growth dalam penelitian oleh Nuraeni et al. berhasil mengoptimalkan strategi penjualan dengan akurasi segmentasi pelanggan yang tinggi [10]. Qur'ani dan Nindha menerapkan kombinasi metode clustering dengan association rules untuk segmentasi pelanggan layanan sewa kendaraan, dan temuan mereka mendukung relevansi K-Means untuk klasifikasi perilaku berbasis lokasi dalam bisnis retail dan jasa [11]. Penelitian oleh Fadholi mengimplementasikan weighted K-Means pada data UMKM berdasarkan kesiapan sertifikasi halal, dan menemukan bahwa teknik pengelompokan ini efektif untuk menyusun strategi lokasi usaha mikro yang lebih terfokus [12]. Sibarani dan Akbar juga membuktikan bahwa metode ini efektif untuk pemetaan perilaku konsumen di Bonjour Café and Resto dengan akurasi tinggi melalui integrasi RFM dan Apriori [13].

Dalam pengembangan strategi pemasaran, Putri dan Purnomo menggabungkan algoritma apriori dan K-Means untuk memahami pola pembelian di sektor kuliner, yang membantu dalam penyusunan strategi lokasi promosi yang lebih efektif [14]. Fauzi et al. menerapkan pemetaan persepsi menggunakan K-Means untuk menilai positioning UMKM di Bandung, dan hasilnya membantu dalam penyusunan strategi peningkatan daya saing lokasi usaha [15]. Hazelia dan Belgiawan dalam studi kasus Café Sembilan Bintaro merumuskan strategi pemasaran berbasis positioning dengan memanfaatkan cluster analysis terhadap data lokasi dan perilaku konsumen, membentuk segmen target potensial untuk peningkatan penjualan [16]. Maulida dan Agustina melakukan segmentasi terhadap wisatawan kuliner di Madura menggunakan metode K-Means, dan menghasilkan klaster pengunjung yang mendukung strategi penyusunan produk dan lokasi promosi untuk sektor coffee shop. Sementara itu, Azzahra et al. memanfaatkan K-Means untuk segmentasi pasar kopi literan di Bandung dengan pendekatan visualisasi dan menghasilkan segmentasi lokasi strategis yang mendukung peluncuran produk baru UMKM [17].

Meskipun berbagai penelitian sebelumnya telah menunjukkan efektivitas algoritma K-Means dalam konteks pemetaan spasial dan segmentasi konsumen, kebanyakan studi tersebut tidak secara spesifik menggabungkan analisis spasial berbasis koordinat geografis aktual (latitude-longitude) dari dataset publik resmi seperti Open Data Kota Bandung, dan tidak mengevaluasi klaster secara langsung terhadap konteks aktual wilayah perkotaan dan potensi konsep bisnis yang berbeda (seperti grab & go, mahasiswa, atau komuter). Oleh karena itu, penelitian ini menawarkan kontribusi baru dengan mengimplementasikan algoritma K-Means++ berbasis spasial pada data aktual rumah makan dari Open Data Bandung, dan melakukan analisis spasial kontekstual terhadap hasil klasterisasi. Klaster yang terbentuk dievaluasi berdasarkan sebaran lokasi, jarak ke pusat keramaian, serta kecocokannya untuk konsep coffee shop tertentu. Pendekatan ini belum banyak digunakan dalam penelitian sebelumnya dan diharapkan mampu memberikan rekomendasi lokasi bisnis yang lebih aplikatif dan berbasis kondisi geografis nyata.

Algoritma K-Means Clustering dipilih karena telah terbukti efektif dalam berbagai penelitian terdahulu untuk mengelompokkan data spasial berdasarkan kedekatan lokasi dan karakteristik geografis. Keunggulan algoritma ini terletak pada kemampuannya dalam mengidentifikasi pola distribusi usaha dan konsentrasi wilayah potensial yang belum terjamah oleh kompetitor. Penelitian ini bertujuan untuk menerapkan dan membandingkan hasil klasterisasi menggunakan K-Means dalam pemetaan lokasi strategis coffee shop di Kota Bandung. Fitur utama yang digunakan meliputi titik koordinat (latitude dan longitude) serta informasi lokasi rumah makan dari data Open Data Bandung. Hasil dari klasterisasi ini diharapkan dapat memberikan rekomendasi wilayah dengan peluang pasar tinggi dan tingkat persaingan yang terkendali, sehingga dapat menjadi dasar dalam penentuan lokasi coffee shop yang optimal.

2. METODOLOGI PENELITIAN

2.1 Tahapan penelitian

Penelitian ini dilakukan melalui tahapan-tahapan sistematis untuk mengidentifikasi lokasi strategis coffee shop di Kota Bandung menggunakan algoritma K-Means Clustering. Setiap tahapan dirancang untuk fokus pada proses pengumpulan, pemrosesan, analisis, dan visualisasi data. Dengan pendekatan ini, diharapkan metode yang digunakan mampu menghasilkan model segmentasi wilayah yang efektif dan dapat digunakan untuk rekomendasi lokasi strategis coffee shop. Gambar 1 merupakan ilustrasi alur dari metode penelitian ini.



Gambar 1. Alur Metode Penelitian

Tahapan awal dimulai dengan pengumpulan dan peninjauan data spasial dari sumber terbuka yaitu Open Data Bandung. Data yang dikumpulkan terdiri dari nama rumah makan, alamat, serta titik koordinat geografis (latitude dan longitude). Data tersebut disimpan dalam file yang kemudian diproses untuk kepentingan klasterisasi spasial. Setelah proses akuisisi data selesai, dilakukan tahapan pre-processing data yang berfokus pada pemilihan atribut relevan dan



pengecekan kualitas data. Gambar 1 di atas ini memperlihatkan alur metode penelitian yang terdiri dari empat tahapan utama, yaitu, pengumpulan data spasial, praproses data, proses klusterisasi menggunakan algoritma K-Means, dan visualisasi hasil klusterisasi dalam bentuk peta.

2.2 Pengumpulan Data

Berisi. Data yang digunakan dalam penelitian ini berasal dari portal Open Data Kota Bandung, yang mencakup informasi lokasi usaha kuliner berupa:

- Nama rumah makan
- Alamat
- Koordinat geografis (latitude dan longitude)
- Tahun pencatatan

Data ini dianggap relevan karena menggambarkan sebaran spasial usaha kuliner yang dapat dianalisis untuk menentukan zona dengan potensi tinggi terhadap perkembangan coffee shop. Dataset kemudian difilter dan disimpan dalam file data_filtered.csv untuk digunakan pada tahapan analisis berikutnya.

2.3 Pre-prosesing data

Langkah pre-processing dilakukan untuk menyiapkan data agar siap digunakan dalam proses klusterisasi. Proses ini meliputi:

- Pemilihan fitur latitude dan longitude sebagai basis pengelompokan spasial.
- Pembersihan data dari nilai duplikat atau kosong.
- Penyimpanan data dalam struktur DataFrame untuk pengolahan lebih lanjut.

Selanjutnya, dilakukan eksplorasi awal untuk menentukan jumlah kluster yang optimal menggunakan metode Elbow. Metode elbow merupakan sebuah metode yang diterapkan pada k-means clustering untuk menentukan jumlah optimal dari cluster yang akan dibentuk. Metode elbow membantu dalam menentukan jumlah cluster yang tepat untuk data tertentu[18]. Dalam tahapan ini, model diuji dengan jumlah kluster antara 1 hingga 10 dan nilai inertia (total jarak antar titik dalam kluster) dianalisis untuk melihat titik optimal (elbow point).

2.4 Pemodelan

Model klusterisasi yang digunakan dalam penelitian ini adalah algoritma K-Means dari pustaka sklearn.cluster. Setelah jumlah kluster ditentukan (dalam hal ini dipilih 3 kluster berdasarkan hasil Elbow Method), dilakukan pelatihan model menggunakan koordinat lokasi (latitude, longitude). Model membagi lokasi rumah makan menjadi tiga kluster berdasarkan kedekatan spasial. Dalam penelitian ini, algoritma K-Means++ digunakan sebagai metode inisialisasi centroid awal. Pemilihan centroid dengan metode K-Means++ bertujuan untuk meningkatkan kualitas hasil klusterisasi dan mengurangi kemungkinan terjebak dalam lokal optimum, dibandingkan dengan inisialisasi centroid secara acak pada algoritma K-Means standar. k-means++ adalah teknik peningkatan dari algoritma k-means yang melakukan pemilihan awal centroid secara bertahap dengan probabilitas berbobot terhadap jarak (D^2) dari titik data ke centroid yang sudah dipilih, sehingga mampu meningkatkan akurasi dan kecepatan konvergensi dalam proses clustering[19].

Untuk memastikan bahwa hasil klusterisasi yang diperoleh bersifat reproducible, maka digunakan parameter random_state=42. Dalam Scikit-learn, random_state mengontrol pengacakan yang diterapkan pada data sebelum melakukan pemisahan. Kami menggunakannya saat menggunakan train_test_split untuk membagi data menjadi kumpulan data pelatihan dan pengujian[20]. Penetapan nilai ini akan menjamin bahwa setiap kali kode dijalankan, proses inisialisasi centroid dan hasil klusterisasi akan selalu sama. Hal ini sangat penting dalam penelitian ilmiah agar hasil dapat diuji ulang dan dibandingkan dengan konsisten. Selanjutnya, dilakukan proses pelatihan dan prediksi kluster dengan fungsi fit_predict() pada model K-Means. Fungsi ini akan mengelompokkan data berdasarkan fitur yang diberikan[21]. dalam kasus ini adalah latitude dan longitude. Hasil pengelompokan berupa label kluster disimpan ke dalam kolom baru bernama cluster pada dataset. Setiap data lokasi akan memiliki label 0, 1, atau 2 sesuai dengan kluster yang ditentukan oleh algoritma.

Setelah inisialisasi centroid dilakukan, algoritma K-Means akan memasuki tahap iteratif. Pada setiap iterasi, jarak antara masing-masing titik data terhadap seluruh centroid (dalam hal ini tiga centroid awal yang telah ditentukan) dihitung menggunakan rumus Euclidean distance. Euclidean Distance adalah ukuran jarak lurus atau garis terpendek antara dua titik dalam ruang[22]. Rumus Euclidean digunakan karena merupakan metode paling umum dalam mengukur jarak geometris antar dua titik dalam ruang dua dimensi, yang sesuai untuk atribut lokasi seperti latitude dan longitude.

Rumus umum dari Euclidean distance antara titik data $x=(x_1, x_2)$ dan centroid $c=(c_1, c_2)$ adalah sebagai berikut:

$$d(x, c) = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2} \quad (1)$$

Dalam konteks ini, x_1 dan x_2 mewakili nilai latitude dan longitude dari masing-masing titik data, sedangkan c_1 dan c_2 adalah nilai centroid pada dimensi latitude dan longitude. Setiap titik data akan dihitung jaraknya terhadap ketiga centroid, kemudian akan dikelompokkan ke dalam kluster dengan jarak terdekat. Proses ini diulang secara iteratif: centroid akan diperbarui sebagai rata-rata dari semua titik yang termasuk dalam masing-masing kluster, lalu jarak dihitung ulang hingga perubahan posisi centroid menjadi sangat kecil (konvergen), atau hingga jumlah iterasi maksimum tercapai.

2.5 Evaluasi

Evaluasi model klusterisasi dalam penelitian ini dilakukan dengan pendekatan analisis hasil spasial, khususnya menggunakan perhitungan jarak dan visualisasi. Karena klusterisasi K-Means merupakan metode unsupervised, maka tidak digunakan metrik akurasi seperti pada model klasifikasi[23]. Sebagai gantinya, jarak minimum antar titik ke centroid, konsistensi klaster, dan visualisasi distribusi spasial digunakan sebagai tolok ukur evaluasi.

Tahap pertama evaluasi dilakukan dengan menghitung jarak setiap titik data terhadap masing-masing centroid menggunakan Euclidean distance. Kolom hasil jarak ini dinamai C0, C1, dan C2 untuk masing-masing klaster. Dari perhitungan ini, titik data diberi label ke klaster dengan jarak terpendek yang kemudian disimpan dalam kolom Hasil Kluster. Hasil evaluasi teknis model K-Means meliputi:

- Label klaster untuk setiap titik data (0, 1, atau 2)
- Jarak masing-masing titik ke seluruh centroid
- Penentuan centroid baru pada setiap iterasi
- Visualisasi hasil akhir klusterisasi dalam bentuk grafik sebar spasial

Hasil ini akan digunakan sebagai dasar dalam memberikan rekomendasi lokasi strategis coffee shop berdasarkan kedekatan lokasi dengan pusat aktivitas dan sebaran kompetitor

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pengumpulan data dilakukan melalui portal Open Data Kota Bandung yang menyediakan informasi mengenai rumah makan di wilayah Kota Bandung, dengan atribut seperti id, kode provinsi, nama provinsi, kode kabupaten kota, nama kabupaten kota, nama rumah makan, alamat, latitude, longitude, dan tahun. Seleksi awal diterapkan untuk memastikan hanya data dengan nilai latitude dan longitude yang valid digunakan dalam penelitian ini, kemudian hasilnya akan ditampilkan dalam bentuk tabel untuk keperluan analisis spasial menggunakan algoritma K-Means clustering.

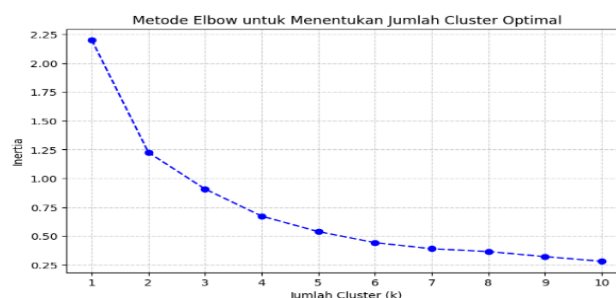
Tabel 1. Data Dari Open Data Bandung

No	Nama Provinsi	Nama Rumah Makan	Latitude	Longitude	Tahun
1	Jawa Barat	Brew & Chew	-6.91537	107.6342868	2020
2	Jawa Barat	Cafe Bali	-6.91345	107.6294417	2020
3	Jawa Barat	Anjis Coffee	-6.90892	107.6277107	2020
4	Jawa Barat	Headklin Barber & Coffee House	-6.91595	107.632708	2020
5	Jawa Barat	garden coffee bdg	-6.91278	107.6326828	2020
...
3556	Jawa Barat	Inj Coffee	-6.90909	107.62109	2020
3557	Jawa Barat	Kukumama	-6.91441	107.62975	2020

Namun, untuk keperluan penelitian ini, data yang digunakan hanya difokuskan pada tiga atribut utama, yaitu nama rumah makan, latitude (garis lintang), dan longitude (garis bujur). Ketiga atribut tersebut dipilih karena menjadi dasar dalam proses analisis spasial menggunakan algoritma K-Means Clustering, khususnya untuk pemetaan dan segmentasi lokasi. Data dengan nilai kosong atau tidak valid pada atribut tersebut akan dihapus dalam proses pre-processing.

3.2 Pre-Processing Data

Setelah data dikumpulkan, dilakukan tahap pre-processing dengan memilih dua fitur spasial utama yaitu latitude dan longitude. Data dibersihkan dari nilai yang kosong dan duplikat agar hasil klusterisasi menjadi lebih akurat. Untuk menentukan jumlah klaster optimal yang akan digunakan dalam algoritma K-Means, dilakukan proses uji menggunakan Metode Elbow. Nilai inertia (jumlah kuadrat jarak dari setiap titik ke centroid klasternya) dihitung untuk jumlah klaster dari 1 hingga 10. Grafik yang dihasilkan menunjukkan penurunan tajam hingga titik $k = 3$, dan setelah itu perubahan nilai inertia cenderung melandai. Oleh karena itu, $k = 3$ dipilih sebagai jumlah klaster yang optimal.



Gambar 2. Hasil dari Metode Elbow

Gambar 2 memperlihatkan grafik Elbow yang digunakan untuk menentukan jumlah kluster optimal. Sumbu horizontal menunjukkan jumlah kluster (k) dari 1 hingga 10, sementara sumbu vertikal menunjukkan nilai inertia, yaitu jumlah kuadrat jarak antara titik data dan centroid masing-masing kluster. Terlihat bahwa nilai inertia mengalami penurunan tajam hingga $k = 3$, setelah itu penurunannya mulai melandai. Titik ini disebut sebagai "elbow point", yang mengindikasikan jumlah kluster optimal karena penambahan kluster setelah titik ini tidak memberikan peningkatan signifikan dalam pemisahan data. Oleh karena itu, nilai $k = 3$ dipilih sebagai jumlah kluster optimal dalam penelitian ini.

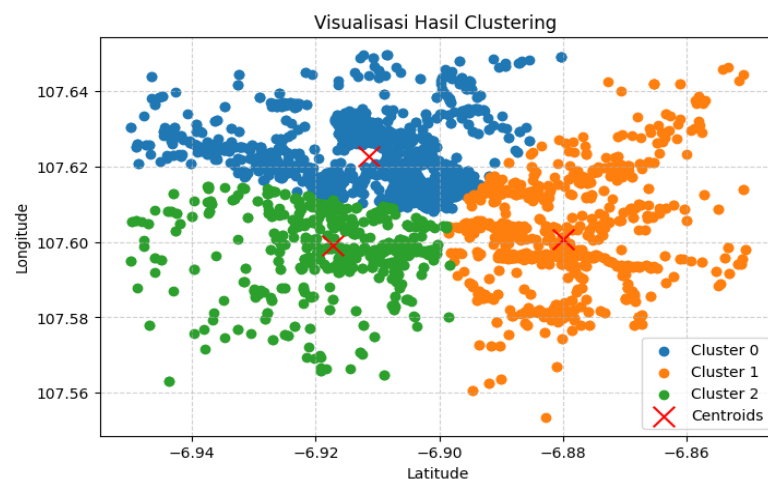
3.3 Pemodelan

3.3.1 Tahap Awal dan Penjelasan Algoritma

Pemodelan dalam penelitian ini menggunakan algoritma K-Means, yang merupakan salah satu metode unsupervised learning paling populer dalam pengelompokan data (clustering). Tujuan utama dari algoritma ini adalah untuk membagi data ke dalam sejumlah kluster berdasarkan kemiripan atau kedekatan nilai antar data. Dalam kasus ini, data yang digunakan berupa koordinat geografis lokasi rumah makan di Kota Bandung, yang ditinjau berdasarkan dua fitur spasial utama yaitu latitude dan longitude.

Pemilihan centroid awal yang akurat sangat memengaruhi hasil akhir klusterisasi. Oleh karena itu, digunakan metode inisialisasi K-Means++, yang merupakan peningkatan dari metode K-Means klasik. Metode ini bekerja dengan memilih titik centroid awal secara bertahap dan mempertimbangkan jarak probabilistik, dengan tujuan mengurangi kemungkinan hasil yang buruk akibat pemilihan awal centroid yang tidak representatif.

Hasil dari inisialisasi centroid menggunakan metode K-Means++ divisualisasikan pada Gambar 3 berikut ini, yang menampilkan penyebaran titik lokasi rumah makan dan posisi centroid awal yang dipilih oleh algoritma. Visualisasi ini memberikan gambaran awal bagaimana klusterisasi dimulai berdasarkan inisialisasi centroid yang cerdas.



Gambar 3 Visualisai Hasil Dari K-means++

Gambar 4 memperlihatkan hasil visualisasi dari proses inisialisasi centroid menggunakan algoritma K-Means++. Dalam grafik ini, titik-titik lokasi rumah makan di Kota Bandung diwakili oleh tiga warna berbeda yang masing-masing menunjukkan hasil awal pengelompokan ke dalam Kluster 0 biru, Kluster 1 oranye, dan Kluster 2 hijau. Titik centroid awal yang dipilih oleh algoritma ditandai dengan simbol silang merah X di tengah masing-masing sebaran kluster. Visualisasi ini memberikan gambaran awal bahwa proses inisialisasi centroid telah berhasil memisahkan titik-titik berdasarkan kedekatan geografis..

Tabel 2. Hasil Centroid Dari K-means++

No	Cluster	Latitude	Longitude
1	0	-6.91143222	107.62273517
2	1	-6.87989136	107.60077369
3	2	-6.91720299	107.59903125

Tabel 2 menyajikan koordinat hasil dari tiga centroid awal yang diperoleh melalui metode K-Means++. Nilai latitude dan longitude pada tabel ini menunjukkan posisi pusat masing-masing kluster yang digunakan sebagai titik acuan dalam proses klusterisasi selanjutnya. Ketiga centroid ini akan menjadi dasar dalam perhitungan jarak Euclidean selama iterasi K-Means untuk menentukan keanggotaan kluster dari masing-masing data lokasi.

3.3.2 Iterasi pertama

Setelah proses inisialisasi centroid awal dilakukan sebagaimana dijelaskan sebelumnya, algoritma K-Means memasuki tahap iteratif. Pada iterasi pertama ini, sistem mulai menghitung jarak antara setiap titik data (dalam hal ini, lokasi rumah



makan) dengan ketiga centroid awal yang telah dipilih secara cermat oleh metode K-Means++. Perhitungan dilakukan menggunakan rumus jarak Euclidean, dengan mempertimbangkan koordinat spasial berupa latitude dan longitude.

Sebelum proses perhitungan jarak dimulai, dilakukan terlebih dahulu konversi nilai latitude dan longitude ke dalam tipe data numerik (float), agar dapat diproses secara matematis oleh sistem. Proses ini dilakukan dengan memanfaatkan fungsi `pd.to_numeric()` dari pustaka `pandas`, yang secara otomatis mengubah string menjadi angka desimal. Konversi ini sangat krusial karena algoritma K-Means dan metode perhitungan jarak Euclidean tidak dapat dijalankan pada data bertipe string. Dengan kata lain, langkah ini merupakan bagian dari tahap pra-pemrosesan untuk memastikan bahwa setiap titik lokasi dapat dihitung jaraknya secara akurat ke centroid.

Jarak yang dihitung dari masing-masing titik ke seluruh centroid akan menjadi dasar dalam menentukan kluster yang paling sesuai. Setiap titik data akan dikelompokkan ke dalam salah satu dari tiga kluster, yaitu kluster 0, kluster 1, atau kluster 2 — berdasarkan jarak terdekat yang dimilikinya terhadap salah satu centroid. Hasil dari proses ini direkam dalam dataset, dengan menambahkan beberapa kolom baru, yaitu:

- C0, C1, dan C2: Menyatakan jarak titik ke masing-masing dari ketiga centroid (kluster 0, 1, dan 2).
- Jarak Terdekat: Nilai jarak minimum dari ketiga jarak yang dihitung, menandakan centroid terdekat.
- Hasil Kluster: Menunjukkan kluster akhir yang dipilih berdasarkan jarak minimum.

Tabel 2 menyajikan hasil dari iterasi pertama, yang menggambarkan bagaimana data mulai terklasifikasi berdasarkan jarak spasial

Tabel 3. Hasil Iterasi Pertama

No	C1	C2	C3	Jarak Terdekat	Hasil Kluster
1	0.012204	0.048804	0.012204	0.012204	C0
2	0.007003	0.044135	0.007003	0.007003	C0
3	0.005574	0.039601	0.005574	0.005574	C0
4	0.010926	0.048126	0.010926	0.010926	C0
5	0.010039	0.045829	0.010039	0.010039	C0
3556	0.002862	0.023503	0.002862	0.002862	C0
3557	0.007621	0.030845	0.007621	0.007621	C0

Tabel 3 tersebut memberikan gambaran awal mengenai proses pembentukan kluster pada iterasi pertama, yang merupakan dasar dari proses pembaruan centroid dan penentuan keanggotaan kluster pada tahap selanjutnya.

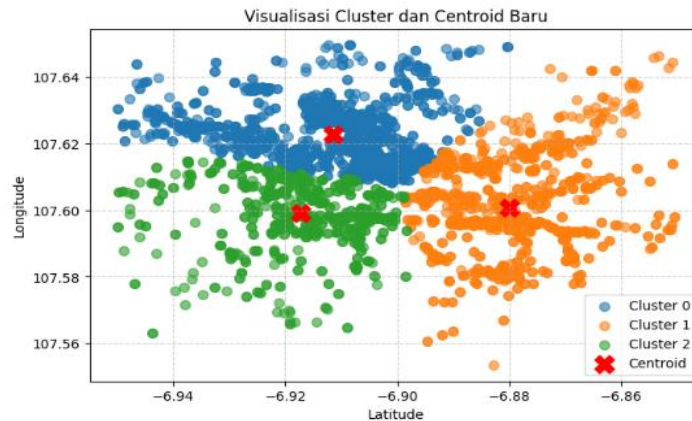
3.3.3 Iterasi Kedua

Pada iterasi kedua, algoritma melakukan pembaruan posisi centroid berdasarkan rata-rata koordinat dari titik-titik yang tergabung dalam masing-masing kluster pada iterasi pertama. Dengan kata lain, setiap centroid baru dihitung ulang berdasarkan pusat gravitasi dari kluster yang terbentuk sebelumnya. Setelah posisi centroid diperbarui, algoritma kembali menghitung jarak semua titik ke centroid yang baru, dan proses pengelompokan diulang kembali. Setiap titik kembali diberi label kluster berdasarkan jarak terdekat. Jika label kluster tidak berubah dibandingkan dengan iterasi sebelumnya, maka algoritma dinyatakan telah mencapai kondisi konvergen. Tabel hasil iterasi kedua juga akan ditampilkan dan memperlihatkan perbandingan terhadap iterasi pertama,

Tabel 4. Hasil Iterasi Kedua

No	C1	C2	C3	Jarak Terdekat	Hasil Kluster
1	0.012226	0.048804	0.035345	0.012226	C0
2	0.007025	0.044135	0.030685	0.007025	C0
3	0.005593	0.039601	0.029898	0.005593	C0
4	0.010947	0.048126	0.033744	0.010947	C0
5	0.010062	0.045829	0.033984	0.010062	C0
3556	0.002848	0.035571	0.023551	0.002848	C0
3557	0.007642	0.045068	0.030889	0.007642	C0

Tabel 4 menampilkan hasil dari iterasi kedua dalam proses klusterisasi menggunakan algoritma K-Means. Kolom C1, C2, dan C3 merepresentasikan jarak Euclidean dari setiap titik data terhadap centroid kluster 0, 1, dan 2. Kolom “Jarak Terdekat” menunjukkan nilai terkecil dari ketiga jarak tersebut, yang menjadi dasar penentuan keanggotaan kluster pada kolom “Hasil Kluster”. Dari tabel ini dapat dilihat bahwa tidak terjadi perubahan label kluster dari iterasi sebelumnya, yang menandakan bahwa proses klusterisasi telah mencapai kondisi konvergen. Data dalam tabel menunjukkan bahwa seluruh titik sampel masih tergolong ke dalam Kluster 0 (C0) berdasarkan jarak minimum ke centroid, menunjukkan konsistensi hasil klusterisasi pada iterasi ini.



Gambar 4. Visualisasi Hasil Dari iterasi kedua

Gambar 4 menyajikan visualisasi hasil akhir dari proses klusterisasi setelah iterasi kedua. Titik-titik data mewakili lokasi rumah makan yang telah dikelompokkan ke dalam tiga kluster yang ditandai dengan warna berbeda: biru untuk Kluster 0, hijau untuk Kluster 1, dan oranye untuk Kluster 2. Posisi centroid akhir ditampilkan dengan simbol "X" berwarna merah. Visualisasi ini memberikan pemahaman spasial terhadap distribusi data dan menunjukkan bahwa setiap kluster terbentuk dalam wilayah yang relatif homogen secara geografis, memperkuat keakuratan hasil dari proses klusterisasi yang telah dilakukan.

3.3.4 Hasil Akhir Klusterisasi

Setelah iterasi kedua selesai dan tidak terjadi lagi perubahan pada label kluster data, maka proses klusterisasi dinyatakan berakhir atau konvergen. Hasil akhir berupa posisi centroid dari setiap kluster menunjukkan lokasi pusat dari kelompok rumah makan yang terbentuk. Berikut adalah posisi akhir centroid dari ketiga kluster, beserta interpretasi wilayah dominan dari masing-masing kluster:

Tabel 5. Nilai Centroid Akhir

No	Cluster	Latitude	Longitude
1	0	-6.911431	107.622713
2	1	-6.879891	107.600774
3	2	-6.917228	107.598990

Tabel 5 menunjukkan koordinat centroid akhir yang diperoleh dari hasil klusterisasi setelah proses iteratif selesai. Kolom "Latitude" dan "Longitude" menampilkan posisi pusat dari masing-masing kluster yang mewakili lokasi konsentrasi rumah makan tertinggi dalam kelompok tersebut. Kluster 0 memiliki pusat di koordinat (-6.911431, 107.622713), Kluster 1 di (-6.879891, 107.600774), dan Kluster 2 di (-6.917228, 107.598990). Informasi ini menjadi dasar untuk evaluasi strategis terhadap potensi wilayah untuk pengembangan coffee shop sesuai karakteristik geografis masing-masing kluster.

3.3.4 Hasil Akhir Klusterisasi

Evaluasi strategis wilayah dilakukan berdasarkan hasil akhir proses klusterisasi serta analisis spasial terhadap titik-titik pusat (centroid) dari masing-masing kluster. Evaluasi ini tidak hanya melihat distribusi geografis tetapi juga mempertimbangkan berbagai faktor kontekstual seperti karakteristik wilayah, kepadatan usaha kuliner, aksesibilitas transportasi, hingga potensi pasar berdasarkan kondisi aktual di lapangan.

Secara teknis, evaluasi dilakukan dengan memanfaatkan sejumlah indikator hasil klusterisasi, antara lain:

- Label kluster untuk setiap titik data: Setiap lokasi rumah makan dalam dataset diklasifikasikan ke dalam salah satu dari tiga kluster, yang dilabeli dengan nilai 0, 1, atau 2.
- Jarak masing-masing titik ke seluruh centroid: Untuk setiap data, dihitung jarak Euclidean terhadap ketiga pusat kluster (C0, C1, C2), dan hasilnya dicatat pada masing-masing kolom.
- Penentuan centroid baru pada setiap iterasi: Proses klusterisasi dijalankan secara iteratif. Centroid diperbarui dengan menghitung rata-rata koordinat dari titik-titik yang termasuk dalam masing-masing kluster pada setiap iterasi. Algoritma dinyatakan konvergen saat tidak terjadi lagi perubahan label kluster antar iterasi.
- Visualisasi hasil akhir klusterisasi: Sebaran spasial lokasi rumah makan berdasarkan hasil klusterisasi divisualisasikan dalam grafik dua dimensi. Tiap titik diberi warna sesuai label kluster, sedangkan posisi centroid ditandai dengan simbol silang merah (X). Visualisasi ini memberikan gambaran spasial yang jelas terhadap wilayah konsentrasi aktivitas kuliner dan potensi lokasi strategis.

Hasil evaluasi berdasarkan wilayah kluster yang terbentuk dijabarkan sebagai berikut:

- Kluster 0 – Wilayah Jl. Aceh, GOR Saparua, Stadion Siliwangi



Wilayah ini berada di pusat keramaian Kota Bandung dan sering digunakan untuk berbagai kegiatan publik seperti olahraga, konser, dan event-event besar. Keberadaan GOR Saperua dan Stadion Siliwangi menjadikan area ini memiliki lalu lintas pejalan kaki yang tinggi, terutama pada akhir pekan dan sore hari. Di sekitarnya juga terdapat berbagai sekolah dan kantor pemerintahan. Berdasarkan hasil klasterisasi, wilayah ini sangat cocok untuk konsep coffee shop grab & go atau kedai santai yang menyasar pengunjung aktif dan pengguna fasilitas publik. Tingkat persaingan cukup tinggi, namun potensi pasar juga besar.

b. Klaster 1 – Jl. Setiabudi, Area Antara ITB dan UPI

Klaster ini merupakan area dengan dominasi aktivitas akademik. Diapit oleh dua perguruan tinggi besar (ITB dan UPI), wilayah ini dihuni oleh mahasiswa, dosen, dan civitas akademika. Selain itu, Jl. Setiabudi juga merupakan jalur wisata ke Lembang, menjadikannya kawasan yang sibuk pada akhir pekan. Coffee shop dengan konsep belajar dan kerja sangat cocok dikembangkan di sini. Meskipun tingkat kompetisi usaha kuliner cukup tinggi, potensi pertumbuhan tetap menjanjikan bagi model bisnis yang unik dan spesifik.

c. Klaster 2 – Jl. Kebon Jati, Sekitar Stasiun Bandung dan Mall Paskal

Wilayah ini merupakan pusat transit dan komersial. Lokasinya berdekatan dengan Stasiun Kereta Api Bandung, Pasar Baru, dan Mall Paskal 23. Aktivitas masyarakat di kawasan ini berlangsung sepanjang hari, menjadikannya sangat ideal untuk coffee shop dengan konsep cepat saji, buka pagi, atau coffee-to-go yang menyasar kalangan komuter dan pelancong. Potensi pasar sangat besar, namun dibutuhkan strategi operasional yang efisien untuk tetap kompetitif di tengah dinamika konsumen yang cepat

4. KESIMPULAN

Penelitian ini menggunakan data lokasi rumah makan di Kota Bandung yang diperoleh dari portal Open Data Kota Bandung dan memanfaatkan algoritma K-Means Clustering dengan metode inisialisasi K-Means++ untuk mengelompokkan data berdasarkan fitur spasial latitude dan longitude. Hasil klasterisasi menghasilkan tiga klaster wilayah strategis, yaitu: Klaster 0 (Jl. Aceh, GOR Saperua, Stadion Siliwangi) yang cocok untuk konsep coffee shop grab & go; Klaster 1 (Jl. Setiabudi, antara ITB dan UPI) yang sesuai untuk tempat belajar dan berkumpul mahasiswa; dan Klaster 2 (Jl. Kebon Jati, sekitar Stasiun Bandung dan Mall Paskal) yang berpotensi untuk coffee shop cepat saji atau to-go. Proses klasterisasi melalui dua iterasi dengan evaluasi berdasarkan konsistensi label klaster, jarak minimum terhadap centroid, serta visualisasi spasial, menunjukkan bahwa model mencapai konvergensi dan hasil klaster yang stabil. Penelitian ini menunjukkan bahwa pendekatan berbasis K-Means Clustering efektif untuk mendukung pengambilan keputusan lokasi bisnis secara spasial dan strategis.

REFERENCES

- [1] D. Hermansyah, A. R. Natasya, I. R. Mukhlis, S. A. Laga, and G. Suprianto, "Sistem Pendukung Keputusan Dalam Menentukan Pemilihan Lokasi Perumahan Strategis Di Sidoarjo Dengan Metode Weighted Product," *INTEGER J. Inf. Technol.*, vol. 8, no. 2, 2023.
- [2] R. R. Oprasto, "Penerapan Metode TOPSIS Dalam Pemilihan Lokasi Usaha Strategis," *J. Data Sci. Inf. Syst.*, vol. 1, no. 3, pp. 109–116, 2023.
- [3] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf. Sci. (Ny)*, vol. 622, pp. 178–210, 2023.
- [4] E. F. L. Avarlina and W. I. Rahayu, "Optimalisasi strategi pemasaran dengan segmentasi pelanggan menggunakan penerapan K-means clustering pada transaksi online retail," *J. Teknol. Dan Inf.*, vol. 13, no. 2, pp. 122–137, 2023.
- [5] D. Prasetyo, W. Lestari, and V. Atima, "Penerapan Clustering Dengan K-Means Untuk Pemilihan Menu Favorit Di Tetra Coffeeshop," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 11, no. 3, 2024.
- [6] D. K. dan P. K. Bandung, "Data Rumah Makan, Restoran, Cafe di Kota Bandung Tahun 2017-2020," Dinas Kebudayaan dan Pariwisata Kota Bandung. [Online]. Available: http://data.bandung.go.id/beta/index.php/portal/detail_data/33e36752-e5f1-429f-829b-745f4dfe3d17
- [7] R. S. Hamid *et al.*, *MANAJEMEN PEMASARAN MODERN: Strategi dan Taktik Untuk Kesuksesan Bisnis*. PT. Sonpedia Publishing Indonesia, 2023.
- [8] M. A. Fathurrohman, "Penentuan Strategi Pengelolaan Coffee Shop di Yogyakarta dengan Mengidentifikasi Perilaku dan Karakteristik Konsumen Menggunakan Metode Association Rules dan Clustering (Studi Kasus Pada Mahasiswa Yogyakarta)," *Repositori Universitas Islam Indonesia*, 2022.
- [9] S. Alexander, M. Mukhsin, and W. Susanti, "Sistem Rekomendasi Cafe di Kota Pekanbaru Menggunakan Metode SAW Terintegrasi Google Maps Berbasis Website," *J. Mhs. Apl. Teknol. Komput. dan Inf.*, vol. 6, no. 3, pp. 104–114, 2025.
- [10] F. Nuraeni, D. Tresnawati, Y. H. Agustin, and G. Fauzi, "Optimization of market basket analysis using centroid-based clustering algorithm and fp-growth algorithm," *J. Tek. Inform.*, vol. 3, no. 6, pp. 1581–1590, 2022.
- [11] A. Qur'ani and F. Nindha, "Perencanaan Strategi Pengelolaan Sewa Kendaraan Share Car Berdasarkan Karakteristik Konsumen menggunakan Metode Clustering dan Association Rules." *Universitas Islam Indonesia*, 2024.
- [12] H. Fadholi, "Pengelompokan Usaha Mikro, Kecil, dan Menengah (UMKM) Berdasarkan Kesiapan Sertifikasi Halal menggunakan Metode Clustering Terbobot." *Universitas Islam Indonesia*, 2024.
- [13] J. M. P. Sibarani, Y. Akbar, and K. Setiawan, "Implementation of RFM Analysis to Enhance Sales Patterns of Food and Beverages at Bonjour Café and Resto Using the Apriori Algorithm," *Int. J. Softw. Eng. Comput. Sci.*, vol. 4, no. 3, pp. 1261–1279, 2024.
- [14] V. E. Putri and H. D. Purnomo, "Integrasi Algoritma Apriori Dan K-Means Dalam Analisis Pola Pembelian Untuk Meningkatkan



- Strategi Pemasaran,” *JIPi (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 10, no. 1, pp. 409–423, 2025.
- [15] F. M. Fauzi, M. E. H. Rosad, and I. Y. Arini, “Perancangan Strategi Positioning pada UMKM AD_Barber Berdasarkan Perseptual Mapping,” *eProceedings Eng.*, vol. 10, no. 3, 2023.
- [16] A. M. Hazelia and P. F. Belgiawan, “Proposed Marketing Strategy to Increase Café Sales (Case Study: Café Sembilan Bintaro),” *Int. Res. J. Econ. Manag. Stud. IRJEMS*, vol. 3, no. 7, 2024.
- [17] P. R. Maulida, F. Agustina, and B. K. Khotimah, “Penentuan Segmentation, Targeting, dan Positioning pada wisatawan desa wisata Lon Malang Madura menggunakan K-means Clustering,” in *Prosiding Seminar Nasional Waluyo Jatmiko*, 2023, pp. 111–120.
- [18] N. A. Maori and E. Evanita, “Metode elbow dalam optimasi jumlah cluster pada k-means clustering,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 14, no. 2, pp. 277–288, 2023.
- [19] D. Arthur and S. Vassilvitskii, “k-means++: The advantages of careful seeding,” Stanford, 2006.
- [20] Kharisma Dharma Putra, “Random State Di Machine Learning,” Bengkel TI. Accessed: Jul. 05, 2025. [Online]. Available: <https://www.bengkelti.com/blog/apa-itu-random-state-di-machine-learning-dan-cara-kerjanya/>
- [21] I. A. Sulasiyah, “Analisis clustering e-learning readiness di pulau jawa menggunakan k-means dan principal component analysis (pca) dengan visualisasi gi.” Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta.
- [22] E. Allen *et al.*, “Spatial quantification of microstructural degradation during fast charge in 18650 lithium-ion batteries through operando X-ray microtomography and Euclidean distance mapping,” *ACS Appl. Energy Mater.*, vol. 5, no. 10, pp. 12798–12808, 2022.
- [23] Y. Gunawan, A. S. Firmansyah, A. Mubarak, B. Subaeki, and K. Manaf, “Implementasi Algoritma Decision Tree untuk Klasifikasi Kemampuan Membaca Al-Qur’an dengan Metode Tahsin,” *Journal of Informatics*, vol. 10, no. 1, 2025.