

Culinary Location Clustering Using the DBSCAN Algorithm

Anestin Halawa^{1*}, Andre Hasudungan Lubis²

^{1,2}Teknik Informatika, Fakultas Teknik, Universitas Medan Area, Kota Medan, 20223, Indonesia

Informasi Artikel

Diterima : 1 Agustus 2025
Revisi : 20 Agustus 2025
Publikasi : 30 September 2025

Kata Kunci:

DBSCAN
Klasterisasi
Lokasi Kuliner
Evaluasi Kinerja
K-Means

ABSTRAK

Industri kuliner memiliki peran penting dalam mendukung pertumbuhan ekonomi kreatif dan pariwisata lokal, sekaligus menjadi bagian integral dari gaya hidup masyarakat urban. Mengingat tingginya jumlah dan keragaman lokasi kuliner, diperlukan pendekatan analitik seperti klusterisasi untuk mengelompokkan lokasi berdasarkan karakteristik pemasaran, guna membantu konsumen dan pelaku usaha dalam pengambilan keputusan yang lebih efisien. Penelitian ini bertujuan untuk mengelompokkan lokasi kuliner berdasarkan atribut pemasaran menggunakan algoritma DBSCAN (Density-Based Spatial Clustering of Applications with Noise). Data sekunder diperoleh dari platform Kaggle yang memuat informasi restoran di Kota Semarang, dengan atribut seperti rating, jumlah ulasan, dan lama operasional. Setelah melalui proses prapemrosesan dan analisis eksploratif, algoritma DBSCAN diterapkan dengan parameter yang disesuaikan untuk mengidentifikasi kluster yang optimal. Hasil klusterisasi menunjukkan terbentuknya 41 kluster dengan karakteristik yang beragam, serta deteksi sejumlah outlier sebagai noise. Evaluasi performa menggunakan Silhouette Score dan Davies-Bouldin Index menunjukkan bahwa DBSCAN menghasilkan kluster yang lebih kompak dan terpisah dengan baik dibandingkan K-Means. Temuan ini membuktikan bahwa DBSCAN lebih adaptif untuk data kuliner yang tidak seragam dan memiliki variasi kepadatan tinggi, sehingga cocok digunakan dalam segmentasi dan pengambilan keputusan strategis di sektor kuliner.

ABSTRACT

The culinary industry plays a vital role in driving creative economic growth and local tourism while also being an integral part of urban lifestyle. Given the high number and diversity of culinary locations, clustering techniques are needed to group them based on marketing characteristics, enabling more efficient decision-making for both consumers and businesses. This study aims to cluster culinary locations based on marketing-related attributes using the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm. Secondary data was obtained from Kaggle, consisting of restaurant information in Semarang City, with attributes such as rating, number of reviews, and operating hours. After preprocessing and exploratory analysis, DBSCAN was applied with adjusted parameters to generate optimal clusters. The results produced 41 clusters with diverse characteristics, including several outliers detected as noise. Performance evaluation using Silhouette Score and Davies-Bouldin Index showed that DBSCAN achieved more compact and well-separated clusters compared to K-Means. These findings demonstrate that DBSCAN is more adaptive for non-uniform culinary data with varying densities and is suitable for segmentation and strategic decision-making in the culinary industry.

This is an open-access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



*Penulis Koresponden

Email: anestinhawawa872@gmail.com

Cara sitasi IEEE:

A. Halawa & A. H. Lubis, "Culinary Location Clustering Using the DBSCAN Algorithm," *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 5, no. 3, pp. 1133-1140, September 2025, doi: 10.30811/jaise.v5i3.7512

1. PENDAHULUAN

Dalam beberapa tahun terakhir, industri kuliner telah mengalami pertumbuhan signifikan seiring dengan meningkatnya minat masyarakat terhadap eksplorasi makanan, baik sebagai kebutuhan primer maupun gaya hidup [1]. Tren kuliner tidak hanya mencerminkan preferensi konsumen tetapi juga berperan sebagai indikator dinamika sosial-ekonomi suatu wilayah. Secara definisi, kuliner merupakan bagian dari budaya yang mencakup segala aspek terkait penyediaan, penyajian, dan konsumsi makanan, termasuk lokasi usaha seperti restoran, kedai, atau *street food* [2]. Perkembangan bisnis kuliner masa kini telah menjadi salah satu penggerak utama ekonomi kreatif global, dengan kontribusi yang cukup signifikan terhadap dunia dan mampu menyumbangkan lapangan pekerjaan yang cukup luas [3]. Fenomena *food tourism* dan digitalisasi pemasarannya semakin luas dan mampu memperlebar jangkauan pasar, sehingga memungkinkan usaha kecil dan menengah (UKM) untuk bersaing secara global.

Dalam industri kuliner yang berkembang pesat, konsumen kini dihadapkan pada beragam pilihan menu, sehingga kualitas layanan menjadi faktor penentu utama dalam pengambilan keputusan. Artinya, pelayanan tidak hanya berperan sebagai pendorong kepuasan pelanggan, tetapi juga menjadi diferensiasi kompetitif bagi pelaku bisnis kuliner, terutama dalam menarik minat wisatawan yang mencari pengalaman yang unik [4]. Oleh karena itu, integrasi antara layanan berkualitas dan strategi menu yang dinamis menjadi krusial untuk memaksimalkan nilai ekonomi sektor kuliner, sekaligus mendukung keberlanjutan usaha di tengah persaingan yang ketat [5].

Dengan demikian, pelanggan memerlukan informasi tentang lokasi kuliner yang berkelompok secara spasial untuk memaksimalkan efisiensi waktu dan biaya dalam eksplorasi gastronomi. Sehingga membutuhkan teknik klasterisasi guna memberikan kemudahan kepada pelanggan untuk menentukan lokasi kuliner tersebut. Klasterisasi lokasi kuliner ini memungkinkan konsumen mengidentifikasi area dengan konsentrasi usaha makanan tertentu, sekaligus menghindari wilayah dengan opsi terbatas yang berpotensi mengurangi pengalaman kuliner mereka. Klasterisasi adalah teknik pengelompokan data ke dalam beberapa kategori berdasarkan kemiripan karakteristik, dimana objek dalam satu kelompok memiliki sifat yang lebih serupa dibandingkan dengan objek di kelompok lain [6]. Metode ini bekerja tanpa panduan label data, sehingga ideal untuk mengeksplorasi pola tersembunyi atau struktur alami dalam dataset [7].

Telah banyak penelitian sebelumnya yang menggunakan teknik klasterisasi, khususnya dalam konteks lokasi wisata dan kuliner, dengan berbagai pendekatan algoritma atau model. Rahman et al. [8] melakukan penelitian tentang analisis geospasial dan statistik distribusi restoran menggunakan pendekatan K-means clustering. Studi ini memanfaatkan sumber data alternatif untuk mengidentifikasi pola lokasi restoran berdasarkan latitude, longitude, harga, dan jenis hidangan. Kemudian, Đokić et al. [9] juga menggunakan teknik klasterisasi untuk menentukan pengelompokan lokasi restoran. Dimana, penelitian tersebut membandingkan kinerja algoritma *Gaussian Mixture Model* (GMM) dan K-means dengan menggunakan data yang berbasis data sosial berbasis lokasi.

Al' Ayubi et al. [10] melakukan penelitian yang menerapkan algoritma K-Means untuk menganalisis data penjualan pada sebuah rumah makan Padang di kawasan Deltamas selama periode 2022 hingga 2023. Penelitian tersebut bertujuan untuk mengelompokkan produk makanan berdasarkan kinerjanya dalam penjualan menjadi tiga kategori utama, yaitu produk dengan penjualan tinggi, penjualan sedang, dan penjualan rendah. Rochman et al. [11] melakukan penelitian yang berfokus pada pengelompokan objek wisata di wilayah Madura dengan mempertimbangkan jarak terhadap fasilitas publik sebagai kriteria utama. Dalam studi tersebut, pengembangan sistem klasterisasi dilakukan untuk mengidentifikasi objek wisata yang memiliki akses non-strategis terhadap fasilitas umum seperti hotel, SPBU, restoran, dan masjid, yang merupakan elemen penting dalam kenyamanan dan daya tarik wisatawan. Metode yang digunakan merupakan kombinasi antara *Agglomerative Hierarchical Clustering* (AHC) dan K-Means, dengan tujuan meningkatkan akurasi.

Sebagian besar penelitian terdahulu cenderung menggunakan algoritma K-Means Clustering sebagai metode utama dalam proses klasterisasi lokasi kuliner dan wisata. Meskipun K-Means memiliki keunggulan dalam hal kesederhanaan dan efisiensi komputasi, algoritma ini memiliki sejumlah keterbatasan yang signifikan. Salah satu kelemahan utama dari K-Means adalah kebutuhan untuk menentukan jumlah kluster (k) secara eksplisit di awal proses, yang tidak selalu mencerminkan struktur alami dari data. Selain itu, K-Means

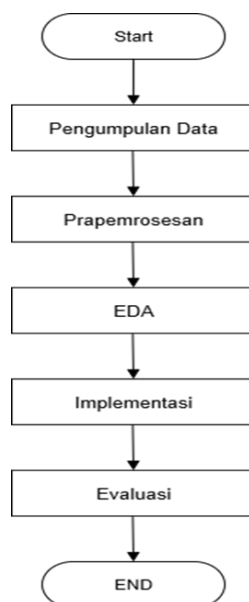
kurang efektif dalam menangani data dengan bentuk klaster yang tidak bulat atau tidak berukuran seragam, serta rentan terhadap outlier yang dapat memengaruhi hasil akhir klasterisasi secara signifikan [12].

Berdasarkan keterbatasan tersebut, penelitian ini mengusulkan penggunaan algoritma *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) sebagai alternatif yang lebih adaptif [13]. DBSCAN adalah algoritma klasterisasi berbasis kepadatan yang mengelompokkan data berdasarkan kedekatan spasial dan jumlah tetangga dalam radius tertentu. DBSCAN mampu mengidentifikasi klaster dengan bentuk yang tidak beraturan serta mendeteksi dan mengeliminasi *outlier* sebagai *noise* tanpa perlu menentukan jumlah klaster di awal proses [14]. Sehingga, hal ini menjadikannya lebih sesuai untuk menganalisis sebaran spasial lokasi kuliner yang kompleks dan heterogen.

Berdasarkan permasalahan yang telah diuraikan, penelitian ini mengusulkan penerapan algoritma DBSCAN untuk melakukan klasterisasi lokasi kuliner berdasarkan karakteristik spasialnya. Dengan pendekatan ini, diharapkan dapat diperoleh pemetaan klaster yang lebih akurat dan representatif tanpa perlu menentukan jumlah klaster secara eksplisit serta tetap mampu mengeliminasi data outlier. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam mendukung perencanaan dan pengembangan sektor kuliner berbasis data, khususnya dalam konteks pemanfaatan analitik spasial di wilayah perkotaan.

2. METODE

Penelitian ini melalui beberapa tahapan sistematis guna menghasilkan klasterisasi yang akurat terhadap data lokasi kuliner berdasarkan atribut yang relevan. Tahapan pertama adalah pengumpulan data sekunder yang diperoleh dari platform *Kaggle*. Selanjutnya, dilakukan proses prapemrosesan data yang mencakup pembersihan data, penanganan nilai hilang, normalisasi, dan transformasi atribut agar sesuai untuk analisis lebih lanjut. Setelah itu, dilakukan *Exploratory Data Analysis* (EDA) untuk memahami pola distribusi data, hubungan antar variabel, serta mengidentifikasi kemungkinan *outlier* yang dapat memengaruhi hasil klasterisasi. Tahap implementasi mencakup penerapan algoritma DBSCAN terhadap data yang telah diproses, dengan penyesuaian parameter *epsilon* dan *minPts* untuk menghasilkan klaster yang optimal. Terakhir, dilakukan evaluasi terhadap hasil klasterisasi menggunakan visualisasi dan interpretasi karakteristik masing-masing klaster guna menilai sejauh mana model mampu memisahkan kelompok produk kuliner secara bermakna berdasarkan atribut pemasaran yang dianalisis. Gambar 1 di bawah ini menunjukkan tahapan dalam penelitian.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh secara sekunder dari platform *Kaggle* yang berjudul “*Restaurant in Semarang, Central Java, Indonesia*”. Adapun tautan untuk data ini adalah sebagai berikut: <https://www.kaggle.com/datasets/ryanalhakim/restaurant-in-semarang-central-java-indonesia>. Dataset ini memuat informasi mengenai berbagai restoran yang beroperasi di wilayah Kota Semarang, dan

dikumpulkan melalui proses *web scraping* terhadap Google Maps. Secara umum, data terdiri atas beberapa atribut yang relevan dengan aspek pemasaran, antara lain nama restoran, kategori kuliner, rating pengguna, jumlah ulasan (*user reviews*), serta estimasi harga. Dataset ini tidak bersifat spasial karena fokus analisis diarahkan pada atribut non-geografis yang bersifat kuantitatif maupun kategorikal.

2.2. Prapemrosesan Data

Tahapan prapemrosesan data dilakukan untuk memastikan kualitas dan kesiapan data sebelum diterapkan ke dalam proses klusterisasi [15]. Langkah pertama yang dilakukan adalah pemeriksaan terhadap nilai *null* atau data yang hilang, yang dapat berpotensi untuk mengganggu performa model apabila tidak ditangani secara tepat. Setelah itu, dilakukan pemilihan atribut dengan mempertimbangkan relevansi terhadap tujuan analisis, di mana atribut yang tidak memberikan kontribusi signifikan terhadap pemodelan akan dihilangkan atau diabaikan. Selanjutnya, atribut kategorikal seperti kategori restoran ditransformasikan ke dalam bentuk numerik menggunakan teknik *label encoding* agar dapat dikenali oleh algoritma DBSCAN yang bekerja pada data numerik. Untuk menyetarakan skala antar atribut, penelitian ini menggunakan metode *Min-Max Scaling*, sehingga tidak terjadi dominasi nilai pada atribut tertentu yang dapat memengaruhi hasil klusterisasi secara tidak proporsional.

2.3. EDA

Proses klusterisasi Tahapan EDA dilakukan untuk memahami karakteristik dasar data dan mengidentifikasi pola distribusi dari masing-masing atribut yang digunakan dalam proses klusterisasi [16]. Analisis dimulai dengan perhitungan statistik deskriptif seperti nilai minimum, maksimum, rata-rata, dan standar deviasi pada atribut numerik, seperti rating, jumlah rating, dan waktu operasional. Kemudian, visualisasi terhadap distribusi data pada penelitian ini menggunakan histogram. Histogram digunakan untuk menggambarkan frekuensi distribusi dari setiap atribut numerik, sehingga pola sebaran data dapat dikenali secara eksplisit. Hasil dari analisis eksploratif ini menjadi dasar dalam penyesuaian parameter algoritma dan validasi kualitas data sebelum tahap klusterisasi dilakukan.

2.4. Implementasi

Tahapan implementasi dalam penelitian ini berfokus pada penerapan algoritma DBSCAN untuk mengelompokkan data lokasi kuliner berdasarkan atribut pemasaran yang telah diproses. DBSCAN merupakan algoritma klusterisasi yang bekerja dengan mengidentifikasi area dengan kepadatan tinggi dan memisahkannya dari area yang jarang memiliki data, sehingga mampu membentuk kluster dengan bentuk arbitrer dan mendeteksi outlier secara otomatis. Terdapat dua parameter utama dalam DBSCAN, yaitu *epsilon* (ϵ) yang menyatakan radius pencarian tetangga, dan *minimum points* (*minPts*) yang menunjukkan jumlah minimal titik dalam radius tersebut untuk membentuk suatu kluster.

Algoritma DBSCAN bekerja melalui beberapa langkah utama. Pertama, setiap titik data diperiksa untuk menentukan apakah terdapat cukup banyak tetangga dalam jarak ϵ . Jika jumlah tetangga $\geq \text{minPts}$, maka titik tersebut dianggap sebagai *core point* dan menjadi pusat kluster. Titik yang berada dalam radius ϵ dari *core point* disebut sebagai *density-reachable* dan akan dimasukkan ke dalam kluster yang sama. Titik yang tidak termasuk ke dalam kluster mana pun dan tidak memiliki cukup tetangga dianggap sebagai *noise* atau outlier. Secara matematis, jarak antar titik dihitung menggunakan rumus *Euclidean Distance* dua dimensi sebagai berikut:

$$d(p, q) = \sqrt{(x_p - x_q)^2 + (y_p - y_q)^2} \quad (1)$$

Dimana $d(p, q)$ adalah jarak antara titik p dan q , serta x dan y merupakan nilai atribut numerik yang telah dinormalisasi. Hasil dari proses ini berupa sejumlah kluster yang terdiri dari restoran dengan karakteristik pemasaran yang serupa, serta beberapa titik yang dikategorikan sebagai outlier karena tidak memiliki kepadatan data yang cukup.

2.5. Evaluasi

Evaluasi terhadap hasil klusterisasi dilakukan untuk menilai kualitas dan validitas pemisahan antar kluster yang dihasilkan oleh algoritma DBSCAN. Dua metrik utama yang digunakan dalam penelitian ini adalah *Silhouette Index* (SI) dan *Davies-Bouldin Index* (DBI), yang merupakan metode evaluasi internal berdasarkan kedekatan antar titik dalam kluster dan keterpisahan antar kluster. SI mengukur seberapa mirip suatu data dengan kluster yang sendiri dibandingkan dengan kluster lain. Nilai SI berada dalam rentang -1 hingga

1, di mana nilai yang mendekati 1 menunjukkan bahwa objek berada dalam kluster yang tepat, sedangkan nilai mendekati -1 menunjukkan bahwa objek mungkin salah kluster. Rumus dari SI untuk satu titik i adalah:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (2)$$

Dimana, nilai $a(i)$ adalah rata-rata jarak antara i dengan semua titik lain dalam kluster yang sama (*intra-cluster distance*) dan $b(i)$ adalah jarak rata-rata terkecil dari titik i ke semua titik dalam kluster lain yang paling dekat (*nearest-cluster distance*).

DBI merupakan metrik evaluasi internal yang digunakan untuk menilai kualitas hasil klusterisasi berdasarkan kompaksi intra-kluster dan pemisahan antar-kluster. Metrik ini mengukur seberapa mirip satu kluster dengan kluster lainnya, di mana nilai yang lebih rendah menunjukkan kluster yang lebih baik. Pada Persamaan (3) dan (4), nilai R_{ij} digunakan untuk mengukur tingkat kemiripan antara kluster i dan j , yang dihitung berdasarkan rasio antara total dispersi dalam kluster (SSW) terhadap jarak antar pusat kluster (SSB). Nilai SSW_i merepresentasikan rata-rata jarak antara setiap titik data dalam kluster ke i terhadap pusat klusternya, sehingga mencerminkan tingkat kompaksi internal. Sementara itu, SSB_{ij} adalah jarak Euclidean antara pusat kluster i dan j , yang menggambarkan tingkat separasi antar kluster [17]. Nilai-nilai tersebut dirumuskan sebagai berikut.

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \quad (3)$$

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} (R_{ij}) \quad (4)$$

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dari proses analisis dan implementasi metode klusterisasi menggunakan algoritma DBSCAN serta evaluasi pembandingan dengan K-Means. Setiap tahapan yang telah dijelaskan dalam metode—mulai dari EDA, proses klusterisasi, hingga evaluasi performa—dihasilkan dalam bentuk visualisasi, statistik deskriptif, dan pengukuran metrik evaluatif. Pembahasan difokuskan pada interpretasi karakteristik kluster yang terbentuk, efektivitas model dalam mengelompokkan data. Analisis EDA berguna untuk memahami karakteristik data. Statistik deskriptif menunjukkan bahwa rata-rata rating restoran berada pada kisaran tinggi, sementara jumlah ulasan memiliki variasi yang lebar, menandakan ketimpangan popularitas antar restoran.

Pada tahap implementasi, algoritma DBSCAN diterapkan pada data yang telah dinormalisasi. Penyesuaian parameter epsilon (ϵ) dan minimum points (minPts) dilakukan secara iteratif untuk memperoleh pemisahan kluster optimal. Hasil klusterisasi DBSCAN membentuk beberapa kluster dengan kepadatan berbeda serta mendeteksi sejumlah titik sebagai noise. Kluster ini mencerminkan variasi karakteristik restoran, seperti segmen dengan rating tinggi namun ulasan rendah, atau sebaliknya. Evaluasi kinerja membandingkan DBSCAN dan K-Means akan menggunakan metrik SI dan DBI.

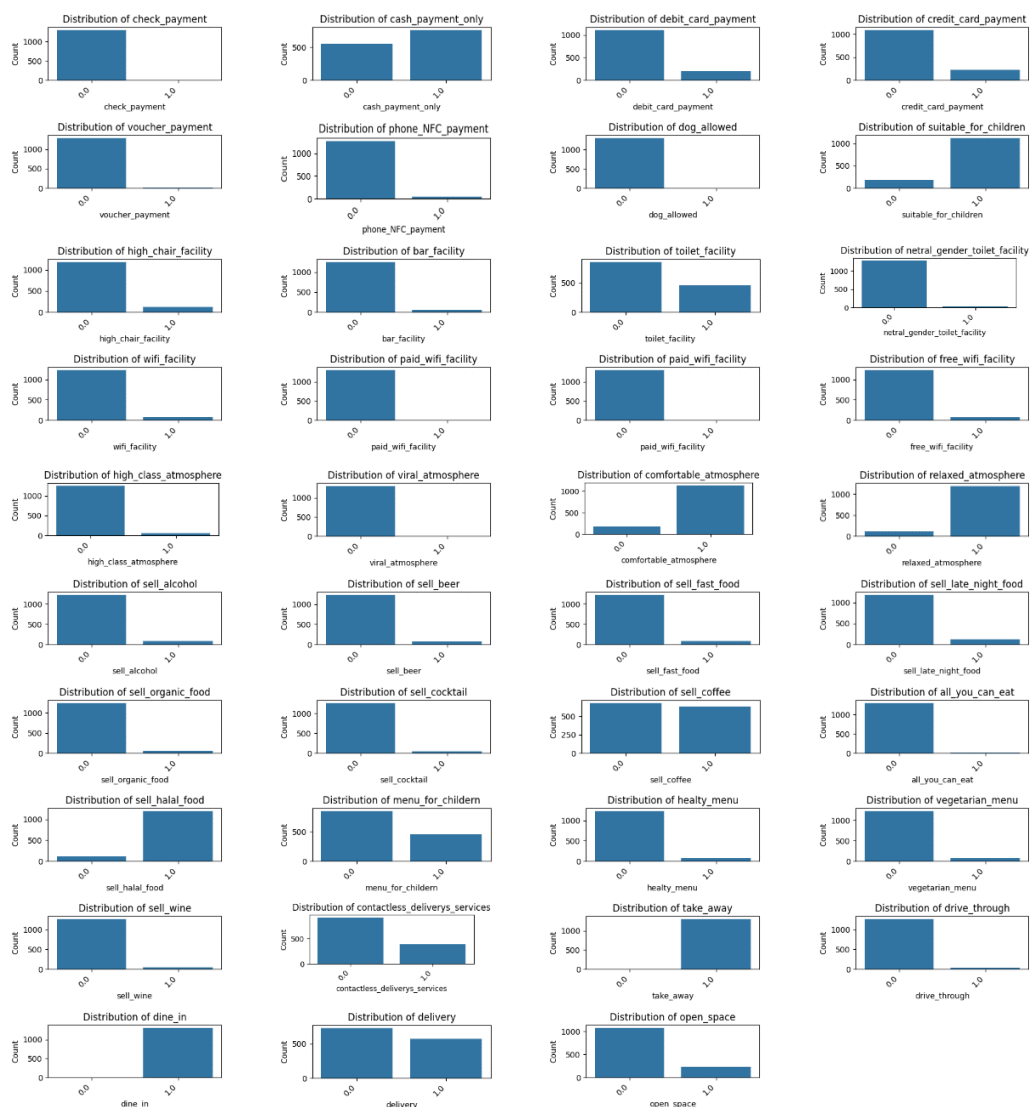
3.1. EDA

Berdasarkan hasil dari statistik deskriptif, rata-rata rating restoran berada pada kisaran tinggi, sedangkan jumlah ulasan menunjukkan variasi yang lebar, menandakan ketimpangan popularitas antar restoran. Visualisasi distribusi menggunakan histogram memperlihatkan pola distribusi yang tidak simetris serta adanya outlier, terutama pada atribut jumlah ulasan. Temuan ini menjadi dasar dalam penentuan parameter klusterisasi yang sesuai.

Tabel 1. Statistik deskriptif

Resto Rating	Rating Numbers	Average Operation Hours
1311.000000	1311.000000	1311.000000
4.527841	552.739130	70.388863
0.496241	1691.891371	2186.193579
1.000000	1.000000	0.400000
4.400000	4.000000	7.700000
4.600000	21.000000	10.900000
4.900000	234.500000	12.000000
5.000000	18208.000000	79167.000000

Berdasarkan hasil analisis statistik deskriptif pada Tabel 1, diketahui bahwa atribut Resto_Rating memiliki nilai rata-rata sebesar 4,52 dengan simpangan baku 0,49, menunjukkan bahwa sebagian besar restoran memiliki penilaian yang relatif tinggi dan tidak terlalu menyebar. Atribut Rating_Numbers, yang merepresentasikan jumlah ulasan dari konsumen, menunjukkan rata-rata sebesar 552,73 namun dengan simpangan baku yang sangat besar, yaitu 1.691,89, yang mengindikasikan adanya variasi yang signifikan antar restoran dalam hal tingkat partisipasi ulasan. Hal ini terlihat dari rentang data yang sangat ekstrem, mulai dari jumlah ulasan minimum sebanyak 1 hingga maksimum mencapai 18.208 ulasan. Sementara itu, atribut Average_Operation_Hours memiliki nilai rata-rata sebesar 70,38 jam per minggu, namun dengan sebaran yang sangat lebar (standar deviasi sebesar 2.186,19), menandakan adanya ketidakwajaran atau kemungkinan pencilaan pada data jam operasional. Temuan ini menjadi pertimbangan penting dalam tahap prapemrosesan, khususnya untuk identifikasi dan perlakuan terhadap outlier yang dapat memengaruhi hasil klusterisasi secara signifikan.



Gambar 2. Histogram dari berbagai atribut

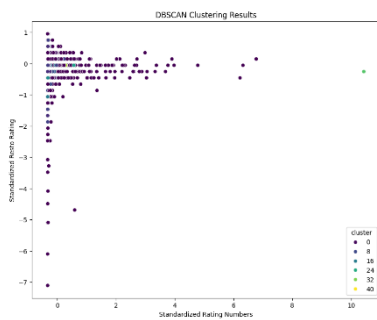
Berdasarkan hasil eksplorasi terhadap variabel kategorikal biner pada Gambar 2, dapat disimpulkan bahwa sebagian besar lokasi kuliner dalam dataset memiliki karakteristik fasilitas dan layanan yang cenderung homogen. Terlihat bahwa fitur seperti *dine-in*, *delivery*, dan *take-away* merupakan tiga layanan dominan yang dimiliki hampir seluruh restoran, sebagaimana ditunjukkan oleh distribusi nilai 1 yang mendekati jumlah maksimum sampel. Selain itu, fitur *free wifi*, *comfortable atmosphere*, serta menu *for children* juga menunjukkan tingkat kemunculan yang tinggi, menandakan preferensi atau standar layanan umum yang

ditawarkan oleh restoran di Semarang. Sebaliknya, beberapa fitur memiliki distribusi yang sangat rendah, seperti *check payment*, *voucher payment*, *high class atmosphere*, dan *sell wine*, yang menunjukkan bahwa layanan atau atmosfer tersebut jarang ditemukan dan cenderung menjadi fitur eksklusif bagi segmen tertentu. Pola-pola distribusi ini penting untuk diperhatikan karena dapat memengaruhi struktur kluster yang terbentuk, khususnya dalam membedakan kelompok restoran berdasarkan kelengkapan fasilitas dan strategi layanan yang mereka terapkan.

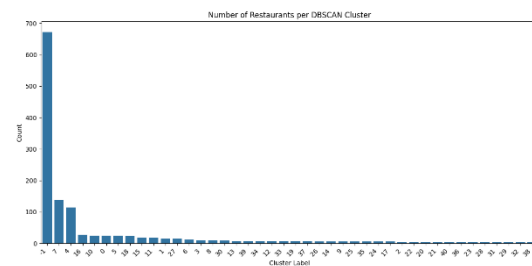
3.2. Implementasi

Pada tahap ini diterapkan algoritma DBSCAN diterapkan dengan parameter ϵ (*epsilon*) = 0,8 dan *minPts* = 5. Gambar 3 menunjukkan hasil visualisasi DBSCAN, di mana sebagian besar titik berada pada cluster dominan (warna ungu) dan beberapa titik lain tersebar sebagai cluster kecil maupun *noise* dan gambar ke-dua menunjukkan distribusi jumlah restoran pada tiap cluster DBSCAN.

Penerapan algoritma DBSCAN pada data menghasilkan pembentukan 41 kluster yang merefleksikan keragaman karakteristik restoran berdasarkan atribut rating dan jumlah ulasan. Sebagian besar data terkonsentrasi pada kluster besar yang mendominasi distribusi, sedangkan sejumlah kluster berukuran kecil menunjukkan keberadaan kelompok dengan karakteristik unik. Hal ini mengindikasikan adanya segmen pasar yang berbeda, seperti restoran dengan rating tinggi tetapi ulasan rendah atau sebaliknya. Selain itu, terdapat titik-titik *noise* yang terdeteksi, yang merepresentasikan entitas dengan karakteristik tidak sesuai pola mayoritas. Pembentukan kluster ini memberikan dasar penting untuk segmentasi pasar yang lebih terarah dalam konteks analisis data kuliner.



Gambar 3. Hasil klasterisasi DBSCAN



Gambar 4. Pola penyebaran kluster

Berdasarkan hasil visualisasi klasterisasi DBSCAN pada Gambar 3, menunjukkan bahwa sebagian besar restoran terkonsentrasi dalam satu kluster dominan dengan kepadatan tinggi, sedangkan beberapa kluster lainnya memiliki jumlah anggota yang relatif kecil, bahkan terdapat titik-titik yang teridentifikasi sebagai *noise*. Distribusi jumlah anggota kluster yang ditunjukkan pada Gambar 4 memperkuat temuan ini, di mana terdapat ketimpangan yang signifikan antara kluster utama dan kluster minoritas. Pola ini mengindikasikan bahwa DBSCAN berhasil memisahkan data ke dalam kelompok yang memiliki karakteristik homogen sekaligus mendeteksi outlier secara efektif. Temuan ini sejalan dengan karakteristik data yang menunjukkan ketimpangan popularitas restoran, sehingga mendukung relevansi DBSCAN sebagai metode klasterisasi yang adaptif untuk data dengan variasi kepadatan tinggi.

3.4 Benchmarking dan Evaluasi

Dalam penelitian ini, tahapan benchmarking dilakukan dengan menggunakan algoritma K-Means sebagai pembandingan terhadap DBSCAN guna mengevaluasi kualitas klasterisasi yang dihasilkan. Penentuan jumlah kluster optimal pada K-Means adalah tiga. Evaluasi kualitas klasterisasi dilakukan menggunakan dua metrik utama, yaitu SI dan DBI. Hasil perhitungan SI dan DBI untuk kedua algoritma dirangkum pada Tabel 1 berikut:

Tabel 2 Algoritma Klasterisasi Menggunakan DBI dan Silhouette Score

Algoritma	Davies Bouldin Index	Silhouette Score
DBSCAN	0.3325524309245344	0.7250663232274625
K-Means	1.1582776067946565	0.24076376306446182

Berdasarkan data yang ada pada tabel diatas menunjukkan bahwa DBSCAN memiliki SI lebih tinggi dan DBI lebih rendah, Hasil evaluasi disajikan pada Tabel 2 dan menunjukkan bahwa algoritma DBSCAN memiliki performa yang lebih baik dibandingkan K-Means. DBSCAN memperoleh nilai SI sebesar 0.725 dan DBI sebesar 0.333, sedangkan K-Means menghasilkan SI sebesar 0.241 dan DBI sebesar 1.158. Nilai-nilai ini mengindikasikan bahwa DBSCAN mampu membentuk kluster yang lebih kompak dan terpisah dengan baik,

sekaligus mendeteksi noise secara efektif. Oleh karena itu, DBSCAN direkomendasikan sebagai metode klusterisasi yang lebih sesuai untuk karakteristik data kuliner yang tidak seragam dan memiliki variasi kepadatan yang tinggi.

4. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma DBSCAN untuk melakukan klusterisasi lokasi kuliner berdasarkan atribut pemasaran seperti rating, jumlah ulasan dan lama operasional restoran. Hasil yang diperoleh menunjukkan bahwa DBSCAN mampu membentuk kluster yang lebih representatif dibandingkan algoritma K-Means, ditunjukkan melalui nilai *Silhouette Score* yang lebih tinggi dan *Davies-Bouldin Index* yang lebih rendah. Kluster yang terbentuk mencerminkan variasi nyata dalam karakteristik restoran, termasuk segmen-segmen dengan kepadatan tinggi dan keberadaan outlier yang secara akurat diidentifikasi sebagai noise oleh DBSCAN. Dengan demikian tujuan penelitian yang telah dirumuskan pada bagian pendahuluan dapat tercapai yaitu memperoleh model klusterisasi yang adaptif terhadap struktur data yang kompleks dan tidak seragam. Ke depan hasil penelitian ini berpotensi dikembangkan lebih lanjut dengan menambahkan variabel spasial dan demografis, serta integrasi sistem rekomendasi berbasis kluster untuk mendukung pengambilan keputusan strategis dalam industri kuliner.

REFERENSI

- [1] K.-S. Lee, "Culinary aesthetics: World-traveling with culinary arts," *Ann. Tour. Res.*, vol. 97, p. 103487, 2022.
- [2] M. J. Stone, S. Migacz, and E. Sthapit, "Connections between culinary tourism experiences and memory," *J. Hosp. Tour. Res.*, vol. 46, no. 4, pp. 797–807, 2022.
- [3] A. R. Putra, E. Ernawati, J. Jahroni, T. S. Anjanarko, and E. Retnowati, "Creative economy development efforts in culinary business," *J. Soc. Sci. Stud.*, vol. 2, no. 1, pp. 21–26, 2022.
- [4] I. C. Dewi *et al.*, *Trend Bisnis Food and Beverages Menuju 2030*. Penerbit Lakeisha, 2022.
- [5] M. Mutmainna, T. Kumalasari, and Y. Yunarti, "Analysis of the Business Environment and Entrepreneurship Strategy in Improving SME Culinary Enterprises," *J. Econ. Resour.*, vol. 7, no. 2, pp. 150–159, 2024.
- [6] S. Pitafi, T. Anwar, and Z. Sharif, "A taxonomy of machine learning clustering algorithms, challenges, and future realms," *Appl. Sci.*, vol. 13, no. 6, p. 3529, 2023.
- [7] A. H. Lubis and E. Ramayana, "A Review on Appropriateness of Partitional Clustering Algorithms in Handling Transactional Data," *Int. J. Research Rev.*, vol. 10, no. 9, pp. 162–169, 2023.
- [8] M. S. Rahman, T. D. Sarkar, and M. Y. Emon, "Comprehensive Geospatial and Statistical Analysis of Restaurant Distribution Using K-means Clustering in Machine Learning Technology," in *2024 4th International Conference on Advancement in Electronics & Communication Engineering (AECE)*, 2024, pp. 461–466.
- [9] K. Đokić, K. Potnik Galić, and K. Štavlić, "Comparison of Clustering Algorithms for Optimal Restaurant Location Selection Using Location-Based Social Networks Data," in *10th International Scientific Symposium Region, Entrepreneurship, Development (RED 2021)*, 2021, pp. 677–690.
- [10] A. J. S. Al' Ayubi, A. S. Sunge, and others, "Application of the K-Means Clustering Algorithm for Sales Analysis in a Padang Restaurant Business," *Int. J. Educ. Life Sci.*, vol. 3, no. 1, pp. 1643–1654, 2025.
- [11] E. M. S. Rochman *et al.*, "A combination of algorithm agglomerative hierarchical cluster (AHC) and K-means for clustering tourism in Madura-Indonesia," *J. Math. Comput. Sci.*, vol. 12, p. Article--ID, 2022.
- [12] 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.
- [13] A. Fawzia Omer, H. A. Mohammed, M. A. Awadallah, Z. Khan, S. U. Abrar, and M. D. Shah, "Big data mining using K-Means and DBSCAN clustering techniques," in *Big Data Analytics and Computational Intelligence for Cybersecurity*, Springer, 2022, pp. 231–246.
- [14] H. V. Singh, A. Girdhar, and S. Dahiya, "A Literature survey based on DBSCAN algorithms," in *2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2022, pp. 751–758.
- [15] A. Amato and V. Di Lecce, "Data preprocessing impact on machine learning algorithm performance," *Open Comput. Sci.*, vol. 13, no. 1, p. 20220278, 2023.
- [16] E. J. M. Carranza, "Exploratory data analysis," *Encycl. Math. Geosci.*, pp. 1–5, 2021.
- [17] A. H. Lubis, W. R. Utami, and J. H. Lubis, "Implementation of k-means clustering for the job provision in urban village," *J. Mat. Dan Ilmu Pengetah. Alam LLDikti Wil. I*, vol. 3, no. 1, pp. 21–31, 2023.