Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 124−138
DOI 10.47065/josh.v8i1.11269
https://ejurnal.seminar-id.com/index.php/josh/

Analisis Pengelompokan Spasial dan Profil Digitalisasi Usaha Mikro, Kecil, dan Menengah Menggunakan DBSCAN

Spatial Clustering Analysis and Digitalization Profiling of Micro, Small, and Medium Enterprises Using DBSCAN

Haris Satria Putra*, Yoshua Ardy Putra

Jurusan Administrasi Niaga, Program Studi Administrasi Bisnis Terapan, Politeknik Negeri Jakarta, Depok, Indonesia
Jl. Prof. DR. G.A. Siwabessy, Kukusan, Kecamatan Beji, Kota Depok, Jawa Barat, Indonesia
Email: [1,*]haris.satriaputra@bisnis.pnj.ac.id, [2]yoshua.ardyputra@bisnis.pnj.ac.id
Email Penulis Korespondensi: haris.satriaputra@bisnis.pnj.ac.id
Received: **01/10/2026**, Revised: **18/10/2026**, Accepted: **28/10/2026**, Published**: 29/10/2026**

Abstrak−Penelitian ini dilatarbelakangi oleh belum meratanya pemanfaatan teknologi digital oleh Usaha Mikro, Kecil, dan Menengah (UMKM) di Kota Depok serta terbatasnya informasi yang menghubungkan persebaran geografis UMKM dengan karakteristik digitalisasinya. Informasi tersebut penting untuk mendukung perencanaan pembinaan yang lebih sesuai dengan kondisi wilayah, terutama ketika sumber daya pelatihan dan pendampingan terbatas. Penelitian ini bertujuan mengidentifikasi pola persebaran spasial UMKM di Kota Depok menggunakan algoritma Density-Based Spatial Clustering of Applications with Noise (DBSCAN) dan mendeskripsikan profil digitalisasi pada kelompok yang terbentuk. Penelitian menggunakan pendekatan kuantitatif eksploratif dengan 50 UMKM yang tersebar di 11 kecamatan. Jumlah tersebut digunakan sebagai sampel eksploratif berdasarkan dataset yang tersedia dan tidak dimaksudkan untuk merepresentasikan seluruh populasi UMKM Kota Depok. Pengelompokan dilakukan hanya berdasarkan koordinat latitude dan longitude menggunakan jarak Haversine, sedangkan digitalisasi dianalisis setelah cluster terbentuk (post-hoc profiling) melalui indikator penggunaan media sosial, marketplace, dan situs web. Analisis K-Distance Graph dan pengujian sensitivitas menghasilkan konfigurasi terpilih Eps sebesar 2,1 km dan min_samples sebesar 4. Hasil DBSCAN membentuk empat cluster yang mencakup 45 UMKM, sementara lima data dikategorikan sebagai noise. Evaluasi menghasilkan Silhouette Score sebesar 0,3223 dan Davies-Bouldin Index sebesar 0,6698, yang menunjukkan bahwa pemisahan cluster masih perlu ditafsirkan secara hati-hati. Rata-rata skor digitalisasi tertinggi ditemukan pada Cluster 0 (1,72), sedangkan terendah pada Cluster 3 (0,20). Uji Kruskal–Wallis menunjukkan perbedaan distribusi skor digitalisasi yang signifikan pada empat cluster utama (H = 14,479; p = 0,0023). Temuan ini memberikan gambaran awal mengenai variasi pemanfaatan kanal digital pada kelompok UMKM berdasarkan kedekatan geografis, tetapi tidak membuktikan bahwa lokasi menyebabkan perbedaan digitalisasi. Secara praktis, hasil dapat digunakan sebagai informasi pendukung untuk memetakan kebutuhan asesmen, merancang pelatihan kolektif bagi UMKMyang berdekatan, dan membantu pemerintah menentukan prioritas pendampingan. Penelitian lanjutan perlu menggunakan sampel yang lebih besar dan representatif serta indikator digitalisasi yang mengukur intensitas dan kualitas penggunaan kanal digital.

Kata Kunci: DBSCAN; Pengelompokan Spasial; Digitalisasi UMKM; Analisis Klaster; Usaha Mikro, Kecil, dan Menengah

Abstract−This study addresses the uneven adoption of digital technology among Micro, Small, and Medium Enterprises (MSMEs) in Depok City and the limited information linking their geographic distribution with their digitalization characteristics. Such information is important for planning area-based business support, particularly when training and mentoring resources are limited. This study aims to identify the spatial distribution patterns of MSMEs in Depok City using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm and describe the digitalization profiles of the resulting groups. An exploratory quantitative approach was applied to a dataset of 50 MSMEs across 11 districts. This number was used for exploratory analysis based on the available dataset and is not intended to represent the entire MSME population of Depok City. Clustering was performed exclusively using latitude and longitude coordinates with the Haversine distance metric, while digitalization was examined after cluster formation (post-hoc profiling) through indicators of social media, marketplace, and website adoption. K-Distance Graph analysis and sensitivity testing selected Eps = 2.1 km and min_samples = 4. DBSCAN identified four clusters containing 45 MSMEs, while five observations were classified as noise. Evaluation produced a Silhouette Score of 0.3223 and a Davies-Bouldin Index of 0.6698, indicating that cluster separation should be interpreted cautiously. The highest average digitalization score was found in Cluster 0 (1.72), while the lowest was found in Cluster 3 (0.20). The Kruskal–Wallis test indicated a statistically significant difference in digitalization score distributions across the four main clusters (H = 14.479; p = 0.0023). These findings provide preliminary insight into variations in digital channel adoption among geographically clustered MSMEs, but they do not establish that location causes differences in digitalization. In practice, the results may support needs assessment, collective training for geographically proximate businesses, and the prioritization of MSME mentoring. Further research should use larger and more representative samples and expand digitalization indicators to measure the intensity and quality of digital channel use.

Keywords: DBSCAN; Spatial Clustering; MSME Digitalization; Cluster Analysis; Micro, Small, and Medium Enterprises

1. PENDAHULUAN

Usaha Mikro, Kecil, dan Menengah (UMKM) merupakan tulang punggung perekonomian Indonesia. Kementerian Koordinator Bidang Perekonomian melaporkan bahwa UMKM menyumbang lebih dari 60% Produk Domestik Bruto (PDB), menyerap hampir 97% tenaga kerja, dan jumlahnya mencapai lebih dari 64 juta unit usaha [1]. Peran tersebut juga nyata di tingkat kota. Kota Depok adalah kota satelit di selatan Jakarta dengan penduduk lebih dari

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 124
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 124−138
DOI 10.47065/josh.v8i1.11269
https://ejurnal.seminar-id.com/index.php/josh/

dua juta jiwa dan memadukan karakter perkotaan dan perdesaan, dengan UMKM yang bergerak pada beragam sektor seperti makanan dan minuman, fesyen, kerajinan tangan, dan jasa[2].

Pemanfaatan teknologi digital, mulai dari media sosial, marketplace, situs web, hingga pembayaran nontunai, menjadi sarana penting bagi UMKM untuk memperluas pemasaran dan menjaga daya saing. Namun, adopsinya di Kota Depok belum merata. Survei terhadap 377 pelaku UMKM di Depok menunjukkan bahwa 48% responden hanya memakai uang tunai, hanya 34% yang menggunakan dompet digital, dan 81% mengaku tidak pernah menerima sosialisasi digitalisasi dari pemerintah [2]. Pemerintah Kota Depok merespons kondisi ini dengan memperkuat pelatihan literasi digital dan pemasaran berbasis kecerdasan buatan bagi pelaku UMKM . Karena sumber daya pembinaan terbatas sementara UMKM tersebar di sebelas kecamatan[3], program semacam ini membutuhkan dasar informasi tentang di mana UMKM terkonsentrasi dan seperti apa tingkat digitalisasinya pada setiap konsentrasi tersebut. Informasi itulah yang belum tersedia, dan inilah urgensi penelitian ini: tanpa gambaran spasial, pembinaan digital berisiko tidak tepat sasaran.

Dimensi spasial layak dipertimbangkan karena adopsi digital tidak selalu seragam di ruang geografis. Studi pada UMKM di Eropa menemukan bahwa lokasi memengaruhi peluang adopsi teknologi digital, dengan heterogenitas yang berkaitan dengan kekuatan lingkungan bisnis lokal dan hierarki perkotaan–perdesaan . Temuan tersebut tidak dapat digeneralisasi langsung ke Indonesia, tetapi menunjukkan bahwa lokasi perlu diperlakukan sebagai dimensi analisis, bukan sekadar atribut alamat. Dengan mengelompokkan UMKM berdasarkan kedekatan lokasi, pembinaan dapat dirancang per zona, misalnya melalui pelatihan kolektif atau penyediaan titik layanan, dan tidak terikat pada batas administratif kecamatan.

Penelitian pengelompokan UMKM yang ditemukan penulis umumnya menggunakan atribut usaha, seperti kinerja dan sektor, sebagai dasar pembentukan cluster [4], dengan K-Means sebagai algoritma yang banyak dipakai[5]. Pendekatan ini berguna untuk segmentasi berbasis profil usaha, tetapi tidak menjawab pertanyaan spasial: di mana UMKM berkumpul, dan apakah UMKM yang berdekatan memiliki profil digital yang serupa. Dengan kata lain, informasi lokasi dan informasi digitalisasi umumnya dianalisis secara terpisah.

Untuk data koordinat, pemilihan algoritma perlu mempertimbangkan sifat data geografis. K-Means membagi data ke dalam k cluster berdasarkan pusat (centroid), sehingga jumlah cluster harus ditetapkan di awal, cenderung menghasilkan cluster berbentuk cembung, dan sensitif terhadap pencilan karena setiap titik dipaksa masuk ke salah satu cluster. Hierarchical clustering tidak memerlukan k di awal, tetapi hasilnya bergantung pada metode linkage dan ambang pemotongan dendrogram, tetap menempatkan setiap objek ke dalam suatu cluster, dan pada single linkage dapat menggabungkan kelompok yang terpisah karena efek rantai (chaining). Sebaliknya, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) mendefinisikan cluster sebagai wilayah berkepadatan tinggi yang dipisahkan wilayah berkepadatan rendah; algoritma ini tidak memerlukan jumlah cluster, dapat membentuk cluster berbentuk sembarang, dan melabeli titik terpencil sebagai noise[6]. Sifat-sifat tersebut sesuai dengan data UMKM. Lokasi usaha jarang tersebar dalam pola bulat yang beraturan, dan UMKM yang terpencil justru memberi informasi penting tentang wilayah yang belum terjangkau sehingga tidak semestinya dipaksakan masuk ke suatu kelompok. Selain itu, parameter Eps berdimensi jarak sehingga dapat dinyatakan dalam kilometer dan diterjemahkan menjadi radius layanan pembinaan. Perbandingan K-Means dan DBSCAN pada data sosial-ekonomi dan data ulasan pun telah dilaporkan [7].

DBSCAN telah banyak diterapkan pada data spasial. Pada data point of interest di pusat kota Wuhan, DBSCAN digunakan untuk mengidentifikasi konsentrasi ruang produksi, hunian, dan ekologi [8]. Varian DBSCAN yang dikombinasikan dengan penyaringan adaptif dikembangkan untuk mengelompokkan area perjalanan angkutan umum perkotaan [9], dan peningkatan DBSCAN diusulkan untuk data spasial citra satelit [10]. Di Indonesia, DBSCAN dipakai untuk mengidentifikasi cluster kejadian gempa bumi di wilayah Yogyakarta [11], sedangkan kemampuannya memisahkan titik menyimpang sebagai noise dimanfaatkan pada deteksi anomali [12]. Pada penelitian ini persoalannya berbeda: dengan data yang kecil, tantangan utamanya adalah kepekaan hasil terhadap parameter, bukan beban komputasi. Penelitian sebelumnya telah menerapkan DBSCAN pada data spasial UMKM Kota Depok dan menunjukkan adanya konsentrasi UMKM pada wilayah tertentu [13], tetapi karakteristik digitalisasi pada setiap cluster belum menjadi bagian dari interpretasi.

Penelitian ini memosisikan diri sebagai studi eksploratif dengan 50 pada pengumpulan data. Jumlah ini jauh lebih kecil daripada populasi UMKM Kota Depok yang pada survei tahun 2021 tercatat 5.534 unit [3], sehingga hasilnya tidak dimaksudkan untuk merepresentasikan seluruh UMKM Depok dan belum dapat langsung dijadikan dasar kebijakan wilayah. Pada sampel sekecil ini DBSCAN rentan terhadap perubahan parameter, karena selisih kecil pada Eps atau min_samples dapat mengubah jumlah cluster dan noise [6]. Oleh sebab itu, parameter tidak ditetapkan secara sembarang: K-Distance Graph digunakan untuk mempersempit kandidat Eps, pengujian sensitivitas dilakukan pada lima nilai Eps (1,9–2,3 km), dan kualitas struktur cluster dinilai dengan Silhouette Score [20] serta Davies-Bouldin Indeks [14]. Kualitas tersebut dilaporkan apa adanya, dan temuan diperlakukan sebagai hipotesis awal yang perlu diuji pada data yang lebih besar.

Berdasarkan uraian tersebut, penelitian ini bertujuan (1) mengelompokkan persebaran 50 UMKM di Kota Depok menggunakan DBSCAN berbasis koordinat geografis dengan jarak Haversine, dan (2) mendeskripsikan karakteristik digitalisasi, yaitu penggunaan media sosial, marketplace, dan situs web, pada setiap cluster yang terbentuk. Penelitian memakai rancangan dua tahap. Cluster dibentuk hanya dari koordinat, sedangkan variabel digitalisasi dianalisis setelah cluster terbentuk (post-hoc profiling), agar cluster mencerminkan kedekatan

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 125
This work is licensed under a Creative Commons Attribution 4.0 International License.

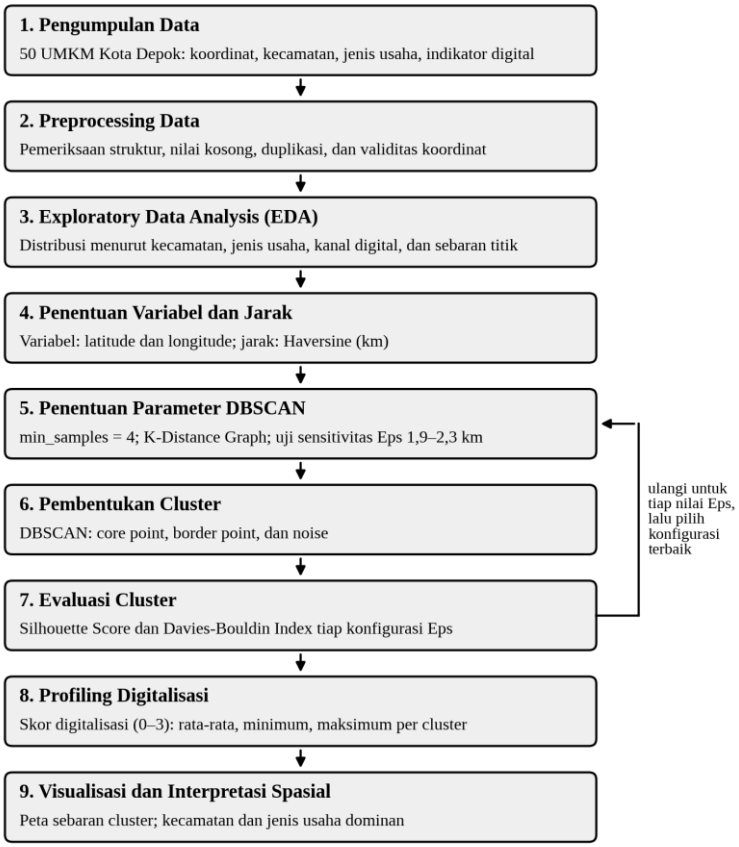
Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 124−138
DOI 10.47065/josh.v8i1.11269
https://ejurnal.seminar-id.com/index.php/josh/

geografis, bukan kemiripan skor digital, dan agar satuan jarak (kilometer) tidak bercampur dengan skor biner. Konsekuensinya, penelitian ini mendeskripsikan variasi digitalisasi antar-cluster dan tidak menguji hubungan sebab-akibat antara kepadatan spasial dan digitalisasi. Kontribusi penelitian ini adalah (i) kerangka dua tahap yang menghubungkan informasi lokasi dan digitalisasi UMKM, (ii) prosedur pemilihan parameter yang dapat ditelusuri, dan (iii) gambaran awal zona UMKM sebagai masukan perencanaan pembinaan berbasis wilayah, dengan batasan sampel sebagaimana dijelaskan di atas.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksploratif dengan metode Density-Based Spatial Clustering of Applications with Noise (DBSCAN) untuk mengidentifikasi pola persebaran UMKM Kota Depok berdasarkan lokasi geografis [6]. Tahapan penelitian disusun mengacu pada kerangka Cross-Industry Standard Process for Data Mining (CRISP-DM) [15] yang disesuaikan dengan kebutuhan analisis spasial. Sembilan tahap yang ditempuh disajikan pada Gambar 1.



Gambar 1. Tahapan penelitian

Tahap pertama adalah pengumpulan data. Dataset terdiri atas 50 UMKM di 11 kecamatan Kota Depok yang dipilih melalui kriteria pemilihan, sumber daftar UMKM, sumber koordinat, dan periode pengumpulan data. Setiap UMKM dicatat nama, jenis usaha, alamat, kecamatan, koordinat, serta tiga indikator digital, yaitu penggunaan media sosial, marketplace, dan situs web, yang diverifikasi melalui pencarian kanal digital internet. Hanya koordinat yang dipakai membentuk cluster. Kecamatan, jenis usaha, dan indikator digital dipakai setelah cluster terbentuk untuk mendeskripsikan karakteristiknya, sehingga cluster mencerminkan kedekatan lokasi dan bukan kemiripan profil usaha[16].

Tahap kedua adalah preprocessing untuk memeriksa struktur data, tipe variabel, nilai kosong, duplikasi, dan validitas koordinat. Seluruh pengolahan menggunakan Python di Google Colab dengan pustaka pandas dan scikit-learn [17]. Tahap ketiga adalah Exploratory Data Analysis (EDA) untuk memperoleh gambaran awal distribusi UMKM menurut kecamatan, jenis usaha, kepemilikan kanal digital, dan sebaran titik koordinat sebelum DBSCAN diterapkan [18].

Tahap keempat menetapkan variabel dan ukuran jarak. Koordinat latitude dan longitude dipertahankan sebagai variabel numerik, dan jarak antar-UMKM dihitung dengan formula Haversine agar Eps dapat dinyatakan

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 126
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 124−138
DOI 10.47065/josh.v8i1.11269
https://ejurnal.seminar-id.com/index.php/josh/

dalam kilometer [19]. Tahap kelima menentukan parameter DBSCAN melalui K-Distance Graph dan uji sensitivitas Eps. Tahap keenam membentuk cluster dengan DBSCAN sehingga setiap UMKM berstatus core point, border point, atau noise [6]. Tahap ketujuh mengevaluasi struktur cluster dengan Silhouette Score dan Davies-Bouldin Index[14], dan hasil evaluasi seluruh konfigurasi Eps dibandingkan untuk memilih konfigurasi akhir.

Tahap kedelapan memberi profil digitalisasi pada setiap cluster melalui rata-rata, minimum, dan maksimum skor digitalisasi, ditambah kecamatan dan jenis usaha dominan. Tahap kesembilan memvisualisasikan hasil dalam peta sebaran titik yang diberi label cluster untuk interpretasi spasial[8].

2.2 Kajian Pustaka Metode

2.2.1 Jarak Geografis (Haversine)

Koordinat geografis berada pada permukaan bola sehingga jarak Euclidean pada satuan derajat tidak merepresentasikan jarak sebenarnya. Formula Haversine menghitung jarak lingkaran besar (great-circle distance) antara dua titik. Jarak antara dua UMKM dihitung dengan Persamaan (1) dan (2).

$$a = \sin^2\left(\frac{\Delta\varphi}{2}\right) + \cos\varphi_1 \cos\varphi_2 \sin^2\left(\frac{\Delta\lambda}{2}\right) \quad (1)$$

$$d = 2R \arcsin\sqrt{a} \quad (2)$$

Pada persamaan tersebut, d adalah jarak antara dua titik (km), R adalah jari-jari bumi (≈ 6.371 km), φ adalah latitude, λ adalah longitude (keduanya dalam radian), $\Delta\varphi = \varphi_2 - \varphi_1$, dan $\Delta\lambda = \lambda_2 - \lambda_1$. Pada scikit-learn, metrik haversine menuntut koordinat dan Eps dalam radian, sehingga Eps dalam kilometer dibagi R [20].

2.2.2 Algoritma DBSCAN

DBSCAN membentuk cluster dari wilayah berkepadatan tinggi yang dipisahkan wilayah berkepadatan rendah dengan dua parameter, yaitu Eps sebagai radius maksimum lingkungan suatu titik dan min_samples sebagai jumlah minimum titik dalam lingkungan tersebut[6]. Lingkungan titik p didefinisikan pada Persamaan (3), dan syarat core point dinyatakan pada Persamaan (4).

$$N_{Eps}(p) = \{q \in D \mid dist(p,q) \leq Eps\} \quad (3)$$

$$| N_{Eps}(p) | \geq min_samples \quad (4)$$

Pada persamaan tersebut, D adalah himpunan data dan dist(p, q) adalah jarak Haversine pada Persamaan (2). Titik yang memenuhi Persamaan (4) disebut core point. Border point berada dalam lingkungan suatu core point tetapi tidak memenuhi Persamaan (4), sedangkan titik lainnya berstatus noise dan diberi label −1 [18]. Berbeda dengan K-Means yang memerlukan jumlah cluster dan sensitif terhadap pencilan, DBSCAN tidak memerlukan jumlah cluster dan memisahkan pencilan sebagai noise. Sifat ini membuat DBSCAN banyak digunakan pada data spasial, misalnya titik lokasi perkotaan dan kejadian gempa bumi.

2.2.3 Penentuan Parameter

Hasil DBSCAN peka terhadap Eps dan min_samples, sehingga keduanya perlu ditetapkan dengan prosedur yang dapat ditelusuri [6]. Prosedur pada penelitian ini terdiri atas empat langkah. Pertama, min_samples ditetapkan 4 karena data berdimensi dua dan nilai tersebut direkomendasikan untuk data dua dimensi. Kedua, K-Distance Graph dengan k = min_samples dibuat dengan mengurutkan jarak setiap titik ke tetangga ke-k terdekatnya. Titik tekuk (knee) pada kurva menandai peralihan dari titik berkepadatan tinggi ke titik terpencil dan dipakai untuk mempersempit kandidat Eps [13]. Ketiga, uji sensitivitas dilakukan pada Eps 1,9; 2,0; 2,1; 2,2; dan 2,3 km, yaitu kisaran di sekitar titik tekuk dengan selang 0,1 km. Untuk setiap nilai dicatat jumlah cluster, jumlah dan persentase noise, Silhouette Score, serta Davies-Bouldin Index. Keempat, konfigurasi akhir dipilih dari kandidat pada kisaran titik tekuk yang memiliki Silhouette Score tertinggi dan Davies-Bouldin Index terendah, dengan persentase noise dipantau agar tidak berlebihan. Dengan prosedur ini, pemilihan Eps tidak bertumpu pada satu percobaan, dan kepekaan hasil terhadap Eps dilaporkan secara terbuka pada bagian hasil.

2.2.4 Evaluasi Kualitas Cluster

Silhouette Score mengukur seberapa baik suatu objek cocok dengan cluster-nya dibandingkan dengan cluster terdekat lainnya [20]. Untuk objek i, nilainya dihitung dengan Persamaan (5), dan skor keseluruhan adalah rata-rata seluruh objek pada Persamaan (6).

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (5)$$

$$S = \frac{1}{n}\sum_{i=1}^{n} s(i) \quad (6)$$

Pada persamaan tersebut, a(i) adalah rata-rata jarak objek i ke objek lain dalam cluster yang sama, b(i) adalah rata-rata jarak objek i ke objek pada cluster terdekat lainnya, dan n adalah jumlah objek yang dievaluasi.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 127
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 124−138
DOI 10.47065/josh.v8i1.11269
https://ejurnal.seminar-id.com/index.php/josh/

Nilai s(i) berada pada rentang −1 sampai 1, dan nilai yang makin mendekati 1 menunjukkan cluster yang makin kompak dan terpisah [20]. Sebagai panduan interpretasi, nilai lebih dari 0,70 menunjukkan struktur kuat, 0,51–0,70 struktur wajar, 0,26–0,50 struktur lemah, dan 0,25 atau kurang menunjukkan struktur yang tidak substansial. Davies-Bouldin Index (DBI) mengukur rata-rata kemiripan setiap cluster dengan cluster yang paling mirip dengannya, sehingga nilai yang lebih rendah menunjukkan cluster yang lebih terpisah [14]. DBI dihitung dengan Persamaan (7) dan (8).

$$R_{ij} = \frac{S_i + S_j}{M_{ij}} \tag{7}$$

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq i} R_{ij} \tag{8}$$

Pada persamaan tersebut, k adalah jumlah cluster, S_i adalah dispersi cluster i (rata-rata jarak anggota ke pusat cluster), dan M_{ij} adalah jarak antara pusat cluster i dan j.

2.2.5 Skor Digitalisasi

Tingkat digitalisasi setiap UMKM dinyatakan dengan skor digitalisasi, yaitu jumlah tiga indikator biner (1 jika kanal digunakan dan 0 jika tidak), seperti pada Persamaan (9).

$$Skor_i = MS_i + MP_i + WB_i \tag{9}$$

Pada persamaan tersebut, MS, MP, dan WB berturut-turut adalah indikator penggunaan media sosial, marketplace, dan situs web oleh UMKM i. Skor berkisar 0–3, dan skor yang lebih tinggi menunjukkan pemanfaatan kanal digital yang lebih beragam. Skor ini tidak mengukur kualitas atau intensitas penggunaan kanal. Pada setiap cluster dihitung rata-rata, minimum, dan maksimum skor secara deskriptif. Karena jumlah anggota tiap cluster kecil, perbedaan antar-cluster ditafsirkan sebagai gambaran awal dan bukan hasil uji statistik.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil secara berurutan, meliputi karakteristik dan eksplorasi dataset, penentuan parameter DBSCAN, pembentukan serta evaluasi cluster, profil digitalisasi, interpretasi dan implikasi pembinaan, pembahasan noise, perbandingan dengan penelitian terdahulu, serta keterbatasan penelitian. Seluruh temuan bersifat eksploratif karena analisis menggunakan 50 UMKM dan tidak dimaksudkan untuk merepresentasikan seluruh UMKM Kota Depok.

3.1 Karakteristik Dataset

Dataset terdiri atas 50 UMKM dengan 12 atribut yang dirangkum pada Tabel 1. Pemeriksaan dengan fungsi info() pada pustaka pandas menunjukkan bahwa seluruh atribut terisi pada 50 observasi, sehingga tidak ditemukan nilai kosong dan seluruh observasi dipertahankan. Tujuh atribut bertipe numerik, yaitu No, Latitude, Longitude, Media Sosial, Marketplace, Website, dan Skor Digitalisasi, sedangkan lima atribut lainnya bertipe kategorikal, yaitu ID UMKM, Nama UMKM, Jenis Usaha, Alamat, dan Kecamatan.

Tabel 1. Struktur dan Kelengkapan Dataset UMKM

No	Atribut	Tipe Data	Jumlah data terisi
1	No	Integer	50
2	ID UMKM	kategorikal	50
3	Nama UMKM	Kategorikal	50
4	Jenis Usaha	Kategorikal	50
5	Alamat	Kategorikal	50
6	Kecamatan	Kategorikal	50
7	Latitude	Float	50
8	Longitude	Float	50
9	Media Sosial	Integer	50
10	Marketplace	Integer	50
11	Website	Integer	50
12	Skor Digitalisasi	Integer	50

Tabel 1 memperlihatkan tiga kelompok atribut. Atribut identitas dan karakteristik usaha (ID UMKM, nama, jenis usaha, alamat, dan kecamatan) dipakai untuk mengenali unit usaha dan menafsirkan hasil. Atribut spasial (Latitude dan Longitude) merupakan satu-satunya variabel pembentuk cluster. Atribut digitalisasi (media sosial, marketplace, website, dan skor digitalisasi) tidak ikut membentuk cluster dan hanya dipakai untuk menyusun profil setiap cluster setelah cluster terbentuk. Statistik deskriptif variabel numerik yang relevan disajikan pada Tabel 2.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 128
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 124−138
DOI 10.47065/josh.v8i1.11269
https://ejurnal.seminar-id.com/index.php/josh/

Tabel 2. Statistik Deskriptif Variabel Numerik

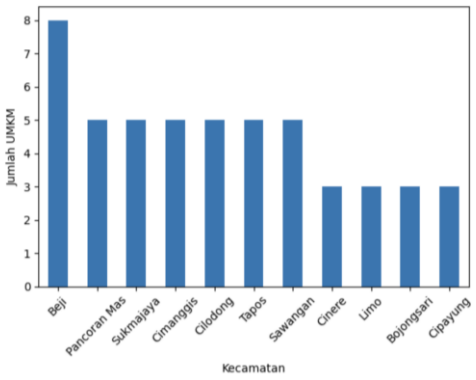
Variabel	Minimum	Maksimum	Rata-rata
Latitude	−6,441076	−6,324781	−6,391300
Longitude	106,730918	106,882024	106,817275
Media Sosial	0	1	0,68
Marketplace	0	1	0,32
Website	0	1	0,32
Skor Digitalisasi	0	3	1,32

Tabel 2 menunjukkan bahwa koordinat 50 UMKM berada pada rentang latitude −6,441076 sampai −6,324781 dan longitude 106,730918 sampai 106,882024. Rentang tersebut setara dengan sekitar 12,9 km pada arah utara–selatan dan 16,7 km pada arah barat–timur. Informasi ini diperlukan untuk menafsirkan parameter Eps pada Subbab 3.3, karena Eps 2,1 km setara dengan sekitar 13% dari lebar sebaran barat–timur. Rata-rata Media Sosial sebesar 0,68 berarti 34 UMKM (68%) memiliki media sosial usaha, sedangkan rata-rata Marketplace dan Website sebesar 0,32 berarti masing-masing 16 UMKM (32%). Skor digitalisasi berkisar 0 sampai 3 dengan rata-rata 1,32, sehingga rata-rata UMKM dalam dataset memakai 1,32 dari tiga kanal digital. Skor ini hanya menghitung jenis kanal yang dimiliki dan tidak mengukur intensitas penggunaannya.

3.2 Eksplorasi Data

3.2.1 Distribusi UMKM Berdasarkan Kecamatan

Gambar 2 menunjukkan bahwa seluruh 11 kecamatan terwakili dalam dataset. Kecamatan Beji memiliki jumlah terbanyak, yaitu 8 UMKM (16%), diikuti enam kecamatan dengan masing-masing 5 UMKM (10%), yaitu Pancoran Mas, Sukmajaya, Cimanggis, Cilodong, Tapos, dan Sawangan. Empat kecamatan lainnya, yaitu Cinere, Limo, Bojongsari, dan Cipayung, masing-masing diwakili 3 UMKM (6%).Distribusi UMKM pada 11 kecamatan disajikan pada Gambar 2.

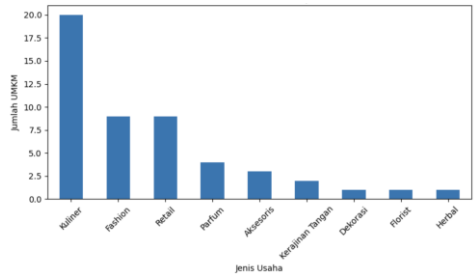


Gambar 2. Distribusi UMKM Berdasarkan Kecamatan

Selisih jumlah antarkecamatan relatif kecil sehingga tidak ada kecamatan yang mendominasi dataset. Namun, angka ini mencerminkan cakupan pengumpulan data, bukan populasi UMKM aktual, sehingga tidak dapat dibaca sebagai tingkat konsentrasi UMKM per kecamatan. Selain itu, DBSCAN bekerja pada kedekatan titik dan bukan pada batas administratif, sehingga UMKM dari kecamatan yang berbeda dapat menjadi anggota cluster yang sama.

3.2.2 Distribusi UMKM Berdasarkan Jenis Usaha

Gambar 3 memperlihatkan bahwa kuliner merupakan jenis usaha terbanyak dengan 20 UMKM (40%), diikuti fashion dan retail masing-masing 9 UMKM (18%), parfum 4 UMKM (8%), aksesoris 3 UMKM (6%), dan kerajinan tangan 2 UMKM (4%). Komposisi jenis usaha dalam dataset disajikan pada Gambar 3.



Gambar 3. Distribusi UMKM Berdasarkan Jenis Usaha

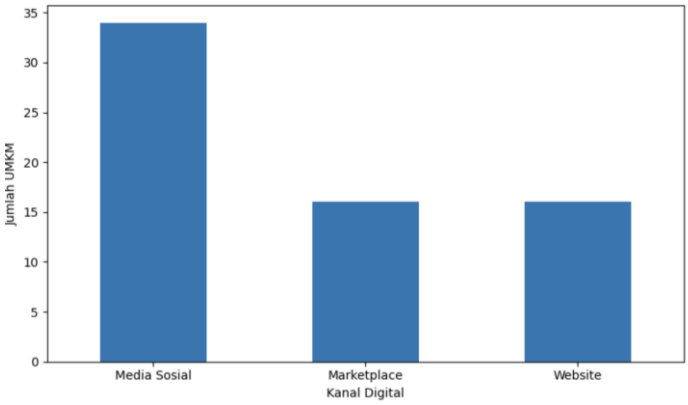
Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 129
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 124−138
DOI 10.47065/josh.v8i1.11269
https://ejurnal.seminar-id.com/index.php/josh/

Dekorasi, florist, dan herbal masing-masing hanya 1 UMKM (2%). Tiga jenis usaha terbesar mencakup 38 dari 50 UMKM (76%). Dominasi ini perlu diperhatikan dalam menafsirkan skor digitalisasi, karena apabila jenis usaha berkaitan dengan pemanfaatan kanal digital, perbedaan skor antar-cluster dapat mencerminkan komposisi jenis usaha dan bukan semata-mata lokasi.

3.2.3 Kepemilikan Kanal Digital

Gambar 4 menunjukkan bahwa media sosial merupakan kanal yang paling banyak dimiliki, yaitu 34 UMKM (68%), sedangkan marketplace dan website masing-masing dimiliki 16 UMKM (32%). Jumlah UMKM yang memiliki masing-masing kanal digital disajikan pada Gambar 4.

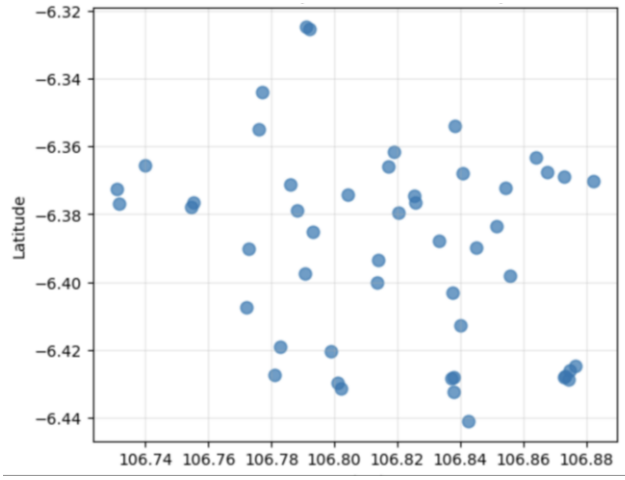


Gambar 4. Kepemilikan Kanal Digital UMKM

Satu UMKM dapat memiliki lebih dari satu kanal, sehingga jumlah ketiga kanal (66 penggunaan) tidak sama dengan jumlah UMKM yang terdigitalisasi, dan jumlah itu pula yang menjadi total skor digitalisasi pada Subbab 3.5. Karena DBSCAN dijalankan hanya dengan koordinat, kepemilikan kanal tidak memengaruhi pembentukan cluster dan baru digunakan sesudahnya untuk membuat profil cluster.

3.2.4 Persebaran Spasial UMKM

Posisi 50 UMKM berdasarkan koordinat disajikan pada Gambar 5, dengan sumbu horizontal sebagai longitude dan sumbu vertikal sebagai latitude.



Gambar 5. Persebaran Spasial UMKM Kota Depok

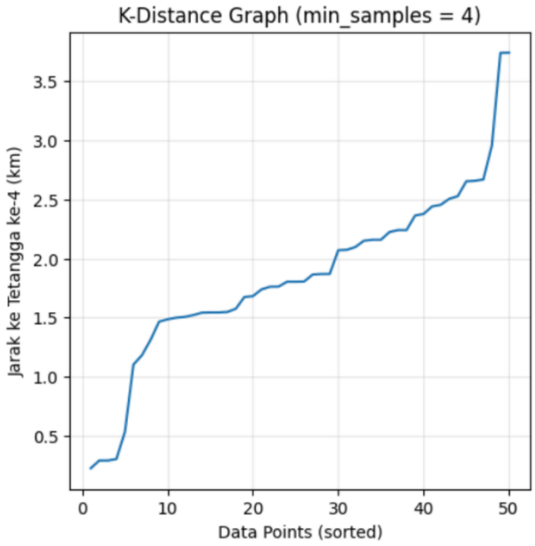
Gambar 5 memperlihatkan bahwa titik-titik UMKM tidak tersebar merata. Sejumlah titik berada sangat berdekatan, misalnya di sudut tenggara pada sekitar longitude 106,87 dan latitude −6,43, sedangkan titik lain berdiri sendiri, terutama di bagian utara dan barat. Variasi kedekatan ini menjadi dasar penerapan DBSCAN. Namun, Gambar 5 belum dapat menentukan jumlah maupun batas cluster karena hal itu bergantung pada nilai Eps dan min_samples, dan antartitik juga tidak tampak celah kepadatan yang tegas. Hal ini relevan untuk menafsirkan kualitas cluster.

3.3 Penentuan Parameter DBSCAN

Nilai min_samples ditetapkan 4, dan K-Distance Graph untuk k = 4 disajikan pada Gambar 6. Kurva K-Distance menunjukkan perubahan kemiringan pada beberapa bagian dan tidak memperlihatkan satu titik tekuk yang sepenuhnya tegas.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 130
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 124−138
DOI 10.47065/josh.v8i1.11269
https://ejurnal.seminar-id.com/index.php/josh/



Gambar 6. K-Distance Graph (min_samples = 4)

Pada bagian awal, jarak tetangga ke-4 meningkat dari nilai yang relatif kecil menuju sekitar 1,5 km. Kurva kemudian meningkat lebih landai pada kisaran sekitar 1,5–2,0 km, sebelum kembali naik pada bagian akhir yang mewakili titik-titik dengan tetangga lebih jauh. Titik tekuk yang paling tajam di ujung kurva tidak otomatis menjadi pilihan terbaik karena dapat menyebabkan sebagian besar titik tergabung dalam cluster besar dan hanya sedikit titik menjadi noise. Oleh karena itu, K-Distance Graph digunakan untuk membatasi rentang kandidat, sedangkan nilai Eps akhir ditentukan melalui uji sensitivitas pada lima kandidat antara 1,9 dan 2,3 km. Uji sensitivitas pada lima nilai Eps dengan min_samples tetap 4 disajikan pada Tabel 3.

Tabel 3. Hasil Uji Sensitivitas Parameter Eps (min_samples = 4)

Eps (km)	Jumlah cluster	Jumlah noise	Persentase noise	Silhouette Score	Davies-Bouldin Index
1,9	3	11	22%	0,3071	0,7109
2,0	3	11	22%	0,3071	0,7109
2,1 (terpilih)	4	5	10%	0,3223	0,6698
2,2	4	4	8%	0,2782	0,7026
2,3	4	3	6%	0,2719	0,7011

Hasil pada Tabel 3 menunjukkan bahwa Eps 2,1 km memberikan nilai Silhouette Score tertinggi (0,3223) dan Davies-Bouldin Index terendah (0,6698) di antara lima konfigurasi yang diuji, dengan empat cluster dan lima titik noise (10%). Pada Eps 1,9 dan 2,0 km, hanya terbentuk tiga cluster dan 11 titik menjadi noise (22%). Ketika Eps dinaikkan menjadi 2,2 km, jumlah cluster tetap empat, tetapi noise berkurang menjadi empat titik (8%); pada Eps 2,3 km, noise turun menjadi tiga titik (6%). Namun, kedua konfigurasi tersebut menghasilkan Silhouette Score yang lebih rendah, masing-masing 0,2782 dan 0,2719. Dengan demikian, Eps 2,1 km dipilih karena memberikan hasil evaluasi relatif terbaik pada rentang kandidat yang diuji, bukan karena menghasilkan noise paling sedikit. Pemilihan ini tetap bersifat kontekstual dan tidak membuktikan bahwa konfigurasi tersebut merupakan satu-satunya solusi yang tepat.

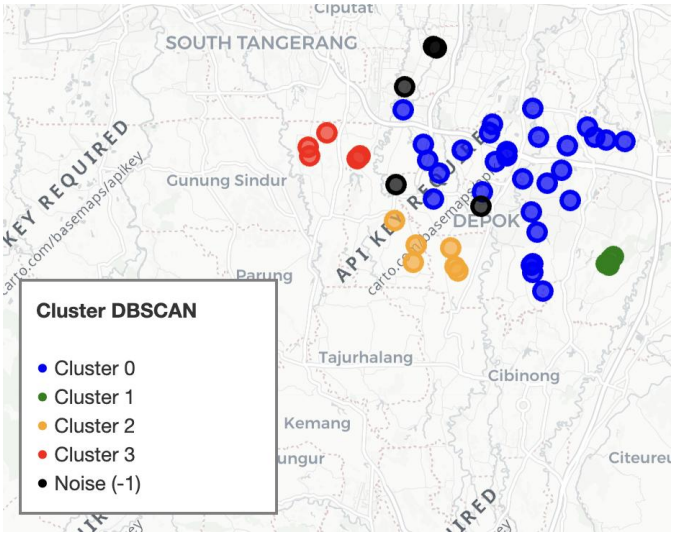
Sebagai pembanding konseptual, apabila 50 titik tersebar acak dan merata pada wilayah sekitar 200 km² [4], jumlah tetangga yang diharapkan dalam radius 2,1 km berada pada orde yang sebanding dengan ambang kepadatan DBSCAN. Perhitungan sederhana ini belum memperhitungkan bentuk wilayah, efek tepi, ketidakmerataan populasi, atau mekanisme pemilihan sampel. Karena itu, perbandingan tersebut tidak dapat digunakan sebagai uji statistik bahwa cluster terbentuk secara kebetulan; fungsinya hanya mengingatkan bahwa kepadatan yang teramati pada sampel kecil perlu ditafsirkan hati-hati. Simulasi sebaran acak yang mempertimbangkan batas wilayah Depok diperlukan untuk menguji apakah pola tersebut berbeda dari pola acak.

3.4 Hasil Pengelompokan dan Evaluasi Kualitas Cluster

Penerapan DBSCAN dengan Eps 2,1 km dan min_samples 4 pada 50 UMKM menghasilkan empat cluster yang mencakup 45 UMKM (90%) serta lima UMKM berstatus noise (10%). Cluster 0 merupakan kelompok terbesar, yaitu 29 UMKM (58% dari seluruh data), disusul Cluster 2 sebanyak 6 UMKM (12%), serta Cluster 1 dan Cluster 3 masing-masing sebanyak 5 UMKM (10%). Angka tersebut menggambarkan komposisi dataset yang dianalisis, bukan proporsi populasi UMKM di setiap wilayah.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 131
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 124−138
DOI 10.47065/josh.v8i1.11269
https://ejurnal.seminar-id.com/index.php/josh/



Gambar 7. Peta Sebaran Hasil Pengelompokan DBSCAN

Peta pada Gambar 7 memperlihatkan Cluster 0 sebagai kelompok terbesar yang mencakup titik-titik dari beberapa kecamatan di bagian utara-tengah hingga timur dataset. Cluster 1 terdiri atas lima titik yang sangat berdekatan di Tapos, sedangkan Cluster 2 dan Cluster 3 masing-masing beranggotakan enam dan lima UMKM pada area selatan. Lima titik noise berada di luar wilayah yang memenuhi kriteria kepadatan minimum pada konfigurasi terpilih. Karena DBSCAN menentukan kelompok berdasarkan jarak dan kepadatan titik, batas cluster tidak harus mengikuti batas administratif kecamatan.

Evaluasi dilakukan pada 45 titik non-noise dengan label empat cluster; lima titik noise tidak diperlakukan sebagai cluster tersendiri dalam penghitungan metrik. Silhouette Score sebesar 0,3223 berada pada rentang 0,26–0,50 yang, berdasarkan pedoman pada Subbab 2.2.4 [31], menunjukkan pemisahan cluster yang relatif lemah. Davies-Bouldin Index sebesar 0,6698 memberikan informasi tambahan tentang dispersi dan kedekatan antarcluster, tetapi tidak boleh ditafsirkan menggunakan ambang universal bahwa nilai di bawah 1 pasti menunjukkan pemisahan yang baik. Metrik ini juga sensitif terhadap bentuk dan ukuran cluster, sedangkan DBSCAN dapat menghasilkan kelompok yang tidak bulat. Secara keseluruhan, kedua metrik dan visualisasi menunjukkan bahwa struktur spasial dapat diidentifikasi, tetapi pemisahannya belum kuat. Hasil ini lebih tepat dipakai sebagai pemetaan eksploratif daripada segmentasi wilayah yang definitif.

3.5 Profil Digitalisasi pada Setiap Cluster

Profil digitalisasi setiap cluster dan kelompok noise disajikan pada Tabel 4. Tabel 4 memperlihatkan perbedaan profil digitalisasi antar-kelompok. Cluster 0 memiliki rata-rata tertinggi, yaitu 1,72, dengan total 50 dari 66 poin digitalisasi pada seluruh dataset (75,8%). Cluster 1 dan Cluster 3 masing-masing memiliki rata-rata 0,40 dan 0,20. Pada kedua cluster tersebut, empat dari lima UMKM memiliki skor nol; dengan demikian, 8 dari 10 UMKM di kedua kelompok belum menggunakan satu pun dari tiga kanal yang diukur. Sebaliknya, skor Cluster 0 tersebar dari 0 hingga 3, dengan skor 2 paling banyak ditemukan (16 UMKM). Temuan ini menunjukkan bahwa rata-rata kelompok tidak menggambarkan kondisi setiap anggota secara seragam. Skor digitalisasi merupakan jumlah kepemilikan media sosial, marketplace, dan website, sehingga tidak menunjukkan intensitas penggunaan, kualitas pengelolaan kanal, atau kontribusinya terhadap penjualan.

Tabel 4. Jumlah Anggota dan Skor Digitalisasi pada Setiap Kelompok

Kelompok	Jumlah UMKM	Persentase	Rata-rata skor	Minimum	Maksimum	Total skor
Cluster 0	29	58%	1,72	0	3	50
Cluster 1	5	10%	0,40	0	2	2
Cluster 2	6	12%	1,17	0	3	7
Cluster 3	5	10%	0,20	0	1	1
Noise (−1)	5	10%	1,20	0	3	6
Seluruh data	50	100%	1,32	0	3	66

Total skor = rata-rata skor × jumlah UMKM (dibulatkan ke bilangan bulat).

Sebagai analisis tambahan terhadap distribusi skor, perbedaan skor antar-empat cluster utama diuji menggunakan Kruskal–Wallis. Uji ini sesuai untuk membandingkan distribusi skor diskret pada kelompok berukuran kecil tanpa mengasumsikan distribusi normal. Jika hasil omnibus menunjukkan perbedaan, perbandingan terencana antara Cluster 0 dan Cluster 1, Cluster 2, Cluster 3, serta gabungan Cluster 1 dan Cluster

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 132
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 124−138
DOI 10.47065/josh.v8i1.11269
https://ejurnal.seminar-id.com/index.php/josh/

3 diperiksa dengan Mann–Whitney dan koreksi Holm. Hasil uji ini bersifat eksploratif; hasilnya tidak mengubah fakta bahwa skor digitalisasi tidak digunakan dalam pembentukan cluster.

Tabel 5. Hasil Uji Perbedaan Skor Digitalisasi antar-Cluster

Uji	Perbandingan	Statistik	Nilai p	Ukuran efek
Kruskal–Wallis	Cluster 0, 1, 2, dan 3	H = 14,479 (df = 3)	0,0023	$\varepsilon^2 = 0{,}280$
Mann–Whitney (Holm)	Cluster 0 vs Cluster 1	U = 122,0	0,0204	r = 0,445
Mann–Whitney (Holm)	Cluster 0 vs Cluster 2	U = 115,0	0,1965	r = 0,222
Mann–Whitney (Holm)	Cluster 0 vs Cluster 3	U = 133,0	0,0055	r = 0,539
Mann–Whitney (Holm)	Cluster 0 vs gabungan Cluster 1 dan 3	U = 255,0	0,0007	r = 0,601

Pada taraf signifikansi 0,05, uji Kruskal–Wallis menunjukkan perbedaan distribusi skor digitalisasi pada empat cluster utama (H = 14,479; df = 3; p = 0,0023), dengan ukuran efek epsilon-kuadrat sekitar 0,280. Pada perbandingan terencana setelah koreksi Holm, Cluster 0 berbeda dari Cluster 1 (p terkoreksi = 0,0204), Cluster 3 (p terkoreksi = 0,0055), dan gabungan Cluster 1 serta Cluster 3 (p terkoreksi = 0,0007), tetapi tidak berbeda signifikan dari Cluster 2 (p terkoreksi = 0,1965). Temuan ini sejalan dengan perbedaan deskriptif pada Tabel 4, terutama antara Cluster 0 dan kelompok dengan skor terendah. Meskipun demikian, hasil hanya berlaku pada sampel yang dianalisis. Ukuran kelompok yang kecil, teknik pengambilan sampel yang tidak acak, dan kemungkinan perbedaan komposisi jenis usaha membatasi generalisasi; hasil tersebut juga tidak membuktikan bahwa kedekatan geografis menyebabkan tingkat digitalisasi tertentu.

3.6 Interpretasi Cluster dan Implikasi bagi Pembinaan UMKM

Komposisi wilayah dan jenis usaha pada setiap kelompok disajikan pada Tabel 6 untuk membantu menafsirkan perbedaan skor digitalisasi pada Tabel 4.

Tabel 6. Komposisi Kecamatan dan Jenis Usaha pada Setiap Kelompok

Kelompok	Kecamatan (jumlah UMKM)	Jenis usaha (jumlah UMKM)
Cluster 0 (29)	Beji 8; Sukmajaya 5; Cimanggis 5; Cilodong 5; Pancoran Mas 3; Limo 3	Kuliner 12; fashion 6; parfum 3; aksesoris 3; retail 1; kerajinan tangan 1; dekorasi 1; florist 1; herbal 1
Cluster 1 (5)	Tapos 5	Kuliner 2; retail 2; fashion 1
Cluster 2 (6)	Cipayung 3; Sawangan 3	Kuliner 3; fashion 2; parfum 1
Cluster 3 (5)	Bojongsari 3; Sawangan 2	Retail 4; kuliner 1
Noise (5)	Cinere 3; Pancoran Mas 2	Kuliner 2; retail 2; kerajinan tangan 1

Catatan: komposisi kecamatan dan jenis usaha pada Tabel 6 diverifikasi melalui tabulasi silang langsung terhadap label cluster pada dataset dan hasil notebook.

Cluster 0. Cluster ini beranggotakan 29 UMKM dari enam kecamatan (Beji, Sukmajaya, Cimanggis, Cilodong, Pancoran Mas, dan Limo) dengan sembilan jenis usaha, sebagaimana tampak pada Tabel 6. Hanya satu UMKM retail berada di cluster ini, sedangkan kuliner (12 UMKM) dan fashion (6 UMKM) mendominasi. Rata-rata skor 1,72 dengan rentang 0–3 menunjukkan bahwa kanal digital sudah banyak dimanfaatkan, tetapi tidak merata di dalam cluster. Karena ukurannya besar dan komposisinya beragam, Cluster 0 sebaiknya tidak diperlakukan sebagai satu kelompok sasaran. Pembinaan lebih tepat disegmentasi di dalam cluster, misalnya UMKM berskor 0–1 yang memerlukan pengenalan kanal dan UMKM berskor 2–3 yang siap naik ke tahap marketplace dan website.

Cluster 1. Cluster ini terdiri atas 5 UMKM yang seluruhnya berasal dari Kecamatan Tapos dan terkonsentrasi sangat rapat di bagian tenggara (Gambar 7). Rata-rata skor 0,40 dengan 4 dari 5 UMKM berskor nol menunjukkan bahwa kanal digital hampir belum dimanfaatkan. Kedekatan lokasi antaranggota membuat pembinaan kolektif di satu titik layanan layak dilakukan, misalnya kelas pembuatan akun media sosial usaha, foto produk, dan pembayaran nontunai. Hasil ini menggambarkan lima UMKM yang diamati dan belum dapat digeneralisasi ke seluruh UMKM di Tapos.

Cluster 2. Cluster ini beranggotakan 6 UMKM, masing-masing tiga dari Cipayung dan Sawangan, dengan rata-rata skor 1,17 dan rentang 0–3. Jenis usaha yang paling banyak adalah kuliner (3 UMKM), disusul fashion (2 UMKM) dan parfum (1 UMKM). Skor yang bervariasi dalam jarak yang berdekatan membuka peluang pembelajaran sebaya, yaitu UMKM berskor tinggi membantu anggota berskor rendah dalam satu cluster, sehingga biaya pembinaan lebih rendah daripada pelatihan terpisah untuk setiap UMKM.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 133
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 124−138
DOI 10.47065/josh.v8i1.11269
https://ejurnal.seminar-id.com/index.php/josh/

Cluster 3. Cluster ini terdiri atas 5 UMKM dari Bojongsari (3 UMKM) dan Sawangan (2 UMKM) dengan rata-rata skor 0,20, di mana 4 dari 5 UMKM berskor nol. Karakter pembeda cluster ini ialah dominasi retail (4 dari 5 UMKM), sedangkan pada cluster lain kuliner dominan atau berimbang dengan retail. Sebanyak 8 dari 9 UMKM retail dalam dataset berada di luar Cluster 0, dan 6 di antaranya berada pada Cluster 1 dan Cluster 3 yang berskor terendah (Tabel 6). Pola ini mengindikasikan bahwa rendahnya skor Cluster 3 mungkin berkaitan dengan jenis usaha dan bukan semata-mata lokasi, sehingga keduanya tidak dapat dipisahkan dengan data yang ada.

Perbandingan Cluster 0 dan Cluster 3. Kedua cluster memiliki kebutuhan pembinaan yang berlawanan. Cluster 0 sudah memakai kanal digital, sehingga tantangannya ialah memperdalam pemanfaatan melalui marketplace, website, dan pembayaran digital. Cluster 3 hampir seluruhnya belum memulai, sehingga tantangannya ialah pengenalan dasar dan pemahaman manfaat kanal bagi usaha retail. Satu program pelatihan seragam bagi kedua cluster berisiko terlalu dasar bagi Cluster 0 dan terlalu sulit bagi Cluster 3. Usulan program untuk setiap kelompok dirangkum pada Tabel 7.

Tabel 7. Usulan Prioritas dan Program Pembinaan per Kelompok

Kelompok	Temuan utama	Prioritas	Usulan program
Cluster 0	29 UMKM di enam kecamatan; rata-rata 1,72; skor 0–3; kuliner dan fashion dominan	Percepatan (naik kelas)	Pelatihan lanjutan marketplace, katalog atau website, dan pembayaran digital; segmentasi internal (skor 0–1 vs 2–3); UMKM berskor 3 sebagai mentor sebaya
Cluster 1	5 UMKM di Tapos; rata-rata 0,40; 4 dari 5 berskor 0	Pengenalan dasar	Kelas kolektif di satu titik layanan Tapos; pembuatan akun media sosial usaha, foto produk, dan pembayaran nontunai
Cluster 2	6 UMKM di Cipayung dan Sawangan; rata-rata 1,17; skor 0–3	Pemerataan antaranggota	Pembelajaran sebaya dalam cluster; pelatihan kanal pesan-antar dan marketplace bagi usaha kuliner
Cluster 3	5 UMKM di Bojongsari dan Sawangan; rata-rata 0,20; retail 4 dari 5; 4 dari 5 berskor 0	Tertinggi	Pembinaan khusus retail (katalog dan stok daring, marketplace); pendampingan langsung; identifikasi hambatan sebelum program
Noise	5 UMKM di Cinere dan Pancoran Mas; rata-rata 1,20; skor 0–3	Layanan individual	Layanan daring atau kunjungan individual, tanpa titik layanan zona; UMKM berskor tinggi sebagai mentor lintas wilayah

Tabel 7 merupakan usulan awal yang diturunkan dari profil 50 UMKM yang dianalisis, bukan rekomendasi kebijakan yang telah tervalidasi. Sebelum diterapkan, usulan perlu dikonfirmasi melalui penambahan sampel dan wawancara dengan pelaku UMKM serta instansi pembina. Eps 2,1 km juga tidak boleh langsung dianggap sebagai radius layanan operasional, sebab jarak Haversine merupakan jarak garis lurus dan tidak mempertimbangkan jaringan jalan, waktu tempuh, maupun hambatan akses.

3.7 Pembahasan Kelompok Noise dan Skor Digitalisasinya

Kelompok noise berisi 5 UMKM dari Cinere (3 UMKM) dan Pancoran Mas (2 UMKM) dengan rata-rata skor 1,20 (Tabel 4). Nilai ini lebih tinggi daripada Cluster 1 (0,40) dan Cluster 3 (0,20), hampir sama dengan Cluster 2 (1,17), dan lebih rendah daripada Cluster 0 (1,72). Terdapat empat penjelasan yang dapat dipertimbangkan, baik secara metodologis, statistik, teoretis, maupun praktis.

Dari sisi metodologis, status noise hanya menandakan bahwa suatu UMKM tidak memiliki cukup tetangga dalam radius 2,1 km. Skor digitalisasi tidak pernah dimasukkan ke dalam algoritma, sehingga tidak ada alasan mekanis bagi kelompok noise untuk berskor rendah atau tinggi. Dari sisi statistik, kelompok ini hanya terdiri atas 5 UMKM dengan total skor 6, sehingga rata-ratanya tidak stabil. Satu UMKM berskor 3 menambah 0,6 pada rata-rata kelompok, dan selisih 0,8 terhadap Cluster 1 setara dengan hanya 4 poin skor, yang dapat berasal dari satu atau dua UMKM. Rentang skor 0–3 pada kelompok noise juga menunjukkan bahwa kelompok ini tidak homogen.

Dari sisi teoretis, terdapat dua dugaan yang berlawanan. Dugaan pertama ialah pengaruh kedekatan antarpelaku usaha, yaitu UMKM yang berdekatan lebih mudah saling meniru praktik digital sehingga cluster padat berskor lebih tinggi. Dugaan kedua ialah kompensasi lokasi, yaitu UMKM yang terisolasi memiliki lebih sedikit pelanggan yang lewat dan lebih sedikit rekan usaha di sekitarnya, sehingga lebih bergantung pada kanal daring untuk menjangkau pelanggan. Studi pada UMKM di Eropa menunjukkan bahwa lokasi memengaruhi peluang adopsi teknologi digital dengan heterogenitas yang berkaitan dengan lingkungan bisnis lokal [5], tetapi arah pengaruh tersebut belum tentu sama di Kota Depok. Data penelitian ini tidak mendukung salah satu dugaan secara tegas. Cluster 0 yang padat memang berskor paling tinggi, tetapi Cluster 1 dan Cluster 3 yang juga padat berskor paling rendah, dan kelompok noise yang terisolasi berada di antara keduanya. Pola ini lebih sesuai dengan heterogenitas wilayah dan jenis usaha daripada dengan hubungan sederhana antara kepadatan dan digitalisasi.

Secara praktis, UMKM noise tidak dapat dilayani dengan pendekatan zona karena tidak memiliki tetangga yang cukup dekat untuk dikumpulkan dalam satu titik layanan. Pendekatan yang lebih sesuai ialah layanan daring

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 134
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 124−138
DOI 10.47065/josh.v8i1.11269
https://ejurnal.seminar-id.com/index.php/josh/

atau kunjungan individual, sedangkan UMKM berskor tinggi di kelompok ini dapat dijadikan mentor lintas wilayah. Faktor lain yang mungkin berperan, misalnya letak Cinere di tepi barat laut Kota Depok yang berbatasan dengan DKI Jakarta dan Tangerang Selatan, tidak diukur dalam penelitian ini dan hanya dapat dijadikan dugaan untuk penelitian berikutnya.

3.8 Implikasi Penggunaan Koordinat sebagai Satu-satunya Variabel Pembentuk Cluster

Penggunaan koordinat sebagai satu-satunya variabel pembentuk cluster berarti bahwa kelompok merepresentasikan kedekatan geografis, bukan kesamaan tingkat digitalisasi. Oleh sebab itu, skor pada Tabel 4 merupakan profil pascapengelompokan, bukan kriteria yang membentuk cluster. Variasi skor 0–3 pada Cluster 0, Cluster 2, dan kelompok noise menegaskan bahwa anggota yang berdekatan tidak selalu memiliki profil digital yang sama. Silhouette Score dan Davies-Bouldin Index dalam analisis ini menilai struktur spasial, bukan homogenitas digitalisasi. Penggabungan koordinat dan indikator digital dapat dipertimbangkan pada penelitian berikutnya, tetapi membutuhkan penanganan perbedaan skala serta pembobotan yang dapat dipertanggungjawabkan.

Kedua, kepadatan yang terbaca dari data adalah kepadatan sampel dan bukan kepadatan populasi. Sampel berjumlah 50 UMKM dari populasi yang tercatat 5.534 unit pada survei tahun 2021 [2], dan teknik pemilihannya dijelaskan pada Subbab 2.1. Apabila pengambilan data lebih banyak dilakukan di kawasan tertentu, kepadatan titik di kawasan itu akan tampak tinggi tanpa menunjukkan sentra UMKM yang sebenarnya. Besarnya Cluster 0 (58% data) dapat sebagian berasal dari faktor ini.

Ketiga, cluster berimpit dengan batas kecamatan dan jenis usaha tertentu. Seluruh anggota Cluster 1 berasal dari Tapos, Cluster 2 hanya dari Cipayung dan Sawangan, Cluster 3 hanya dari Bojongsari dan Sawangan, dan Cluster 3 didominasi retail (Tabel 6). Tanpa variabel kontrol, pengaruh lokasi, kecamatan, dan jenis usaha terhadap skor digitalisasi tidak dapat dipisahkan. Keempat, jarak Haversine adalah jarak garis lurus yang tidak memperhitungkan jalan, sungai, atau jalan tol, sehingga dua UMKM yang tampak berdekatan dapat memiliki waktu tempuh yang jauh berbeda. Kelima, pada sampel kecil keanggotaan cluster ditentukan oleh beberapa titik saja. Perpindahan satu titik antara status noise dan anggota cluster mengubah jumlah noise dan rata-rata skor, sehingga hasil harus dibaca bersama kepekaan parameter pada Tabel 3.

Di sisi lain, penggunaan koordinat murni memberi keuntungan, yaitu hasilnya dapat direplikasi, Eps dapat dinyatakan dalam kilometer dan diterjemahkan menjadi radius layanan, serta UMKM terpencil tidak dipaksakan masuk ke suatu kelompok. Keuntungan tersebut dipertahankan dengan konsekuensi bahwa klaim penelitian dibatasi pada pemetaan kedekatan lokasi yang dilengkapi profil digitalisasi.

3.9 Perbandingan dengan Penelitian Terdahulu

Tabel 8 menunjukkan bahwa studi DBSCAN pada data spasial UMKM Kota Depok juga melaporkan empat cluster, tetapi mengidentifikasi empat titik noise dari 50 UMKM, sedangkan analisis ini mengidentifikasi lima titik noise.Posisi penelitian ini terhadap penelitian-penelitian sebelumnya dirangkum pada Tabel 8.

Tabel 8. Perbandingan dengan Penelitian Terdahulu

Penelitian	Objek dan data	Temuan utama	Perbedaan dengan penelitian ini
DBSCAN pada UMKM Kota Depok [21]	UMKM Kota Depok; data spasial (50 UMKM)	Empat cluster dan empat noise; konsentrasi UMKM pada wilayah tertentu	Tidak menafsirkan digitalisasi; penelitian ini menambah profil digitalisasi pada setiap cluster
Pengelompokan UMKM [6]–[8]	UMKM; atribut usaha seperti kinerja dan sektor	Segmentasi berdasarkan profil usaha, umumnya dengan K-Means	Tidak memakai lokasi; penelitian ini memakai koordinat dan noise
Digitalisasi UMKM di Eropa [5]	UMKM; lokasi dan adopsi digital	Lokasi memengaruhi peluang adopsi; heterogenitas berkaitan dengan lingkungan bisnis lokal	Konteks berbeda; penelitian ini tidak menguji faktor penyebab
DBSCAN pada data titik kepentingan [14]	Pusat kota Wuhan; data POI	Konsentrasi ruang produksi, hunian, dan ekologi	Bukan UMKM dan tanpa atribut digital; data jauh lebih besar
DBSCAN pada kejadian gempa [17]	Wilayah Yogyakarta; kejadian gempa	Identifikasi cluster kejadian gempa	Bidang berbeda; penelitian ini memakai data kecil dan menekankan kepekaan parameter

Informasi yang tersedia dalam ringkasan studi terdahulu belum cukup untuk memastikan apakah selisih tersebut disebabkan oleh nilai Eps, min_samples, sumber koordinat, atau tahap prapemrosesan yang berbeda. Oleh karena itu, selisih satu titik tidak dapat diatribusikan pada satu penyebab tertentu. Kontribusi utama penelitian ini

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 135
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 124−138
DOI 10.47065/josh.v8i1.11269
https://ejurnal.seminar-id.com/index.php/josh/

ialah penambahan profil digitalisasi pascapengelompokan: rata-rata skor pada Cluster 0 sebesar 1,72, sedangkan Cluster 3 sebesar 0,20. Dengan demikian, analisis memperlihatkan variasi kanal digital antar-kelompok spasial tanpa menyatakan bahwa lokasi menjadi penyebab perbedaan tersebut.

Dibandingkan dengan penelitian pengelompokan UMKM yang memakai atribut usaha seperti kinerja dan sektor, penelitian ini menambahkan dimensi lokasi sehingga hasilnya dapat langsung dihubungkan dengan perencanaan pembinaan berbasis wilayah. Konsekuensi dari perbedaan variabel ialah bahwa hasil kedua pendekatan tidak dapat dibandingkan secara langsung. Pengelompokan berbasis atribut usaha menjawab pertanyaan tentang jenis UMKM, sedangkan penelitian ini menjawab pertanyaan tentang lokasi UMKM berkumpul. Temuan bahwa skor digitalisasi bervariasi antar-kelompok spasial searah dengan temuan pada UMKM di Eropa bahwa adopsi digital tidak seragam secara spasial. Namun, penelitian tersebut dan penelitian ini berbeda konteks, ukuran data, dan cara pengukuran, serta penelitian ini tidak menguji faktor penyebabnya, sehingga kesesuaian arah tersebut tidak dapat dijadikan konfirmasi.

Pada penelitian DBSCAN di bidang lain, algoritma ini dipakai untuk data berukuran besar, misalnya titik kepentingan perkotaan dan kejadian gempa bumi [11], dan pengembangan terbaru banyak diarahkan pada efisiensi komputasi. Penelitian ini berbeda karena menggunakan data kecil, sehingga tantangan utamanya ialah kepekaan parameter dan bukan beban komputasi. Karena itu kontribusi metodologisnya ialah pelaporan prosedur pemilihan Eps yang dapat ditelusuri, uji sensitivitas, dan pembanding kepadatan acak, yang umumnya tidak dilaporkan pada studi data besar. Secara keseluruhan, kontribusi penelitian ini ialah kerangka dua tahap yang menghubungkan kedekatan lokasi dan profil digitalisasi UMKM dengan batasan sampel yang dinyatakan secara terbuka.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa DBSCAN dapat digunakan \sebagai pendekatan eksploratif untuk memetakan kedekatan geografis UMKM di Kota Depok. Pada 50 data yang dianalisis, konfigurasi Eps 2,1 km dan min_samples 4 menghasilkan empat cluster yang mencakup 45 UMKM serta lima titik noise. Profil pascapengelompokan memperlihatkan variasi skor digitalisasi, dengan rata-rata tertinggi pada Cluster 0 (1,72) dan terendah pada Cluster 3 (0,20). Temuan tersebut menegaskan bahwa pemetaan spasial dapat membantu menggambarkan konsentrasi dan sebaran UMKM sekaligus menunjukkan kelompok yang layak ditelaah lebih lanjut, tetapi belum membuktikan bahwa kedekatan geografis menyebabkan perbedaan tingkat digitalisasi. Dengan demikian, kontribusi utama penelitian ini adalah kerangka analisis dua tahap yang memisahkan pembentukan cluster berdasarkan koordinat dari interpretasi profil digitalisasi, disertai pelaporan pemilihan parameter dan evaluasi cluster secara transparan. Dari sisi kebijakan, hasil ini dapat dimanfaatkan sebagai informasi awal untuk merancang pembinaan UMKM berbasis wilayah, bukan sebagai dasar tunggal untuk menetapkan penerima program. Pemerintah daerah dan pemangku kepentingan dapat memvalidasi lokasi cluster dan titik noise bersama dinas yang membidangi UMKM serta pelaku usaha setempat, kemudian memetakan kebutuhan pendampingan seperti penguatan media sosial, marketplace, situs web, dan kanal digital lainnya. Kelompok dengan skor digitalisasi rendah dapat diprioritaskan untuk asesmen kebutuhan dan pendampingan awal, sedangkan usaha yang berada di luar cluster padat perlu tetap diperhatikan agar tidak terabaikan akibat pendekatan berbasis konsentrasi lokasi. Penetapan intervensi tetap perlu mempertimbangkan kondisi usaha secara langsung, sebab data penelitian ini belum mengukur kualitas pemanfaatan kanal digital maupun faktor penyebab perbedaan antarkelompok. Penelitian ini memiliki keterbatasan pada jumlah sampel yang hanya 50 UMKM, teknik pemilihan sampel yang tidak acak, sensitivitas DBSCAN terhadap parameter, serta skor digitalisasi yang terbatas pada kepemilikan kanal. Oleh karena itu, hasilnya tidak dapat digeneralisasi ke seluruh UMKM Kota Depok. Penelitian selanjutnya disarankan menggunakan sampel yang lebih besar dengan pengambilan berstrata proporsional menurut kecamatan dan jenis usaha; menguji kestabilan cluster melalui bootstrap serta membandingkan DBSCAN dengan HDBSCAN atau pendekatan clustering spasial-atribut; dan menggunakan simulasi Monte Carlo serta Moran's I untuk menilai apakah pola yang ditemukan berbeda dari sebaran acak dan apakah digitalisasi menunjukkan autokorelasi spasial. Pengukuran digitalisasi juga perlu diperluas dengan intensitas penggunaan, frekuensi pembaruan konten, transaksi pembayaran digital, dan kontribusi penjualan daring. Penambahan variabel skala dan lama usaha serta karakteristik pelaku, penggunaan jarak jaringan jalan, dan validasi melalui wawancara dengan dinas terkait dapat memperkuat ketepatan interpretasi dan relevansi rekomendasi kebijakan.

5. PERNYATAAN PENULIS

5.1 Kontribusi Penulis

Konseptualisasi: Haris Satria Putra, Yoshua Ardy Putra;
Metodologi: Haris Satria Putra, Yoshua Ardy Putra;
Perangkat Lunak: Haris Satria Putra;
Validasi Haris Satria Putra;
Analisis Formal: Haris Satria Putra, Yoshua Ardy Putra;

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 136
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 124−138
DOI 10.47065/josh.v8i1.11269
https://ejurnal.seminar-id.com/index.php/josh/

Investigasi: Haris Satria Putra, Yoshua Ardy Putra;
Sumber Daya: Haris Satria Putra, Yoshua Ardy Putra;
Kurasi Data: Haris Satria Putra;
Penulisan Draf Awal: Haris Satria Putra;
Penulisan, Peninjauan, dan Penyuntingan: Haris Satria Putra, Yoshua Ardy Putra;
Visualisasi: Haris Satria Putra, Yoshua Ardy Putra;
Seluruh penulis telah membaca dan menyetujui versi manuskrip yang telah diterbitkan.

5.2 Ketersediaan Data

Data yang disajikan dalam penelitian ini tersedia berdasarkan permintaan kepada penulis korespondensi.

5.3 Pendanaan

Penelitian ini didanai oleh Politeknik Negeri Jakarta melalui skema Penelitian Asisten Ahli (PAA) tahun 2026 dengan nomor hibah penelitian 343/DST/PL3.14/B/PT.00.03/2026.

5.4 Pernyataan Dewan Peninjau Institusional

Tidak Berlaku (Not applicable)

5.5 Pernyataan Persetujuan Berdasarkan Informasi

Tidak Berlaku (Not applicable)

5.6 Pernyataan Konflik Kepentingan

Para penulis menyatakan bahwa mereka tidak memiliki kepentingan keuangan yang bersaing atau hubungan pribadi yang diketahui yang dapat dianggap memengaruhi pekerjaan yang dilaporkan dalam artikel ini.

5.7 Penggunaan AI Generatif dan Teknologi Berbantuan AI dalam Proses Penulisan

Para penulis menggunakan AI Generatif untuk meningkatkan kejelasan penulisan dalam artikel ini. Para penulis telah meninjau dan menyunting konten yang dihasilkan dengan bantuan AI serta bertanggung jawab penuh atas versi akhir publikasi.

REFERENCES

[1] H. Limanseto, “Pemerintah Dorong UMKM Naik Kelas, Tingkatkan Kontribusi terhadap Ekspor Indonesia,” Kementerian Koordinator Bidang Perekonomian Republik Indonesia, Kabupaten Tangerang, pp. 1–5, Jan. 2025. [Online]. Available: https://ekon.go.id/publikasi/detail/6152/pemerintah-dorong-
[2] R. J. Saputro, D. Winarni, and R. R. Tutupoho, “1 , 2 , 3,” JAProf J. Adm. Prof., vol. 04, no. 1, pp. 57–64, 2023, [Online]. Available: https://doi.org/10.32722/jap.v4i1.6084
[3] Badan Pusat Statistik Kota Depok, “Kota Depok Dalam Angka,” Https://Depokkota.Bps.Go.Id/Id, vol. XXI, Depok, Feb. 2025. [Online]. Available: https://depokkota.bps.go.id/id/publication/2025/02/28/3afc7179869757450705db62/kota-depok-dalam-angka-2025.html
[4] N. Ameliana, N. Suarna, and W. Prihartono, “Analisis Data Mining Pengelompokkan Umkm Menggunakan Algoritma K-Means Clustering Di Provinsi Jawa Barat,” JATI (Jurnal Mhs. Tek. Inform., vol. 8, no. 3, pp. 3261–3268, 2024, doi: 10.36040/jati.v8i3.9655.
[5] D. Marcelina, A. Kurnia, and T. Terttiaavini, “Analisis Klaster Kinerja Usaha Kecil dan Menengah Menggunakan Algoritma K-Means Clustering,” MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 3, no. 2, pp. 293–301, 2023, doi: 10.57152/malcom.v3i2.952.
[6] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, “DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN,” ACM Trans. Database Syst., vol. 42, no. 3, 2017, doi: 10.1145/3068335.
[7] S. Mutiah, Y. Hasnataeni, A. Fitrianto, and L. M. R. D. Jumansyah, “Perbandingan Metode Klastering K-Means dan DBSCAN dalam Identifikasi Kelompok Rumah Tangga Berdasarkan Fasilitas Sosial Ekonomi di Jawa Barat Dalam era digital saat ini , jumlah data yang tersedia dari berbagai bidang , termasuk sosial dan ekonomi , terus,” vol. 09, no. September, pp. 247–260, 2024.
[8] X. Tu, C. Fu, A. Huang, H. Chen, and X. Ding, “DBSCAN Spatial Clustering Analysis of Urban ‘Production–Living–Ecological’ Space Based on POI Data: A Case Study of Central Urban Wuhan, China,” Int. J. Environ. Res. Public Health, vol. 19, no. 9, 2022, doi: 10.3390/ijerph19095153.
[9] X. Zhang, L. Lauber, H. Liu, J. Shi, J. Wu, and Y. Pan, “Research on the method of travel area clustering of urban public transport based on Sage-Husa adaptive filter and improved DBSCAN algorithm,” PLoS One, vol. 16, no. 12 December, 2021, doi: 10.1371/journal.pone.0259472.
[10] A. Ramadhan, F. Achmad, I. Zulkarnain, and M. Aritsugi, “Evaluation of K-Means, DBSCAN, and Hierarchical Clustering for Strategic Segmentation of Tourism SMEs in Rembang, Indonesia,” J. Tek. Inform., vol. 6, no. 3, pp. 1605–1630, 2025, doi: 10.52436/1.jutif.2025.6.3.4602.
[11] W. Ajitomo and I. Pratama, “View of Penerapan Metode Dbscan untuk Identifikasi Kluster Gempa Bumi di Daerah Yogyakarta.pdf,” J. Inform. dan Rekayasa Perangkat Lunak, vol. 6 No 1, pp. 40–46, 2024, [Online]. Available: https://doi.org/10.36499/jinrpl.v6i1.9214
[12] P. Chapman et al., “Crisp-Dm,” in CRISP-DM 1.0 Step by step data mining guide, vol. 78, CRISP-DM consortium, 2000,

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 137
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 124−138
DOI 10.47065/josh.v8i1.11269
https://ejurnal.seminar-id.com/index.php/josh/

pp. 1–78. [Online]. Available: http://www.crisp-dm.org/CRISPWP-0800.pdf
[13] H. S. Putra, Y. A. Putra, and I. R. Triyanto, "Pemetaan Persebaran UMKM Kota Depok Menggunakan Clustering Berbasis Data Spasial," vol. vol 15 no, 2026, [Online]. Available: https://doi.org/10.32722/snrtbisnis.v15i01.7699
[14] N. A. Yolandari, Arnita, and Fanny Ramadhani, "Analisis Perbandingan K-Means Dan Dbscan Dalam Pengelompokan Data Travel Review Ratings Menggunakan Evaluasi Silhouette Index Dan Davies-Bouldin Index," J. Inform. dan Tek. Elektro Terap., vol. 13, no. 3, 2025, doi: 10.23960/jitet.v13i3.6884.
[15] M. M. Juita and R. Rahmi, "Multilingual Sentiment Analysis of RCTI+ Reviews Utilising Orange, ChatGPT, and Naïve Bayes," J. Inf. Technol. Comput. Sci., vol. 10, no. 3, pp. 138–151, 2026, doi: 10.25126/jitecs.2025103787.
[16] A. Saputra and R. A. Sahputra, "Penentuan Kluster UMKM Sektor Perdagangan dan Perikanan Melalui Pendekatan Metode Clustering Data Mining di Kabupaten Aceh Barat," J. Optim., vol. 9, no. 2, p. 173, 2023, doi: 10.35308/jopt.v9i2.8463.
[17] J. Budiasto, H. Ismanto, A. Latif, and Susanto, PYTHON UNTUK DATA SCIENCE Panduan Praktis Menguasai Analisis Data. Yogyakarta: Literasi Langsung Terbit, 2025.
[18] R. Ramadhani, R. Ramadhanu, and T. Hidayat, "Exploratory Data Analysis (EDA) untuk Mengetahui Distribusi Data Kualitas Susu Sapi," J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer), vol. 23, no. 1, p. 68, 2024, doi: 10.53513/jis.v23i1.9500.
[19] G. Boeing, Clustering to Reduce Spatial Data Set Size. 2018, pp. 1–7. doi: 10.2139/ssrn.3145515.
[20] M. Shutaywi and N. N. Kachouie, "Silhouette analysis for performance evaluation in machine learning with applications to clustering," Entropy, vol. 23, no. 6, pp. 1–17, 2021, doi: 10.3390/e23060759.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 138
This work is licensed under a Creative Commons Attribution 4.0 International License.