

Analisis Pemasaran Mikro Perumahan Terjangkau Indonesia Menggunakan Deteksi Komunitas dengan Algoritma Leiden

Subhan Nooriansyah¹, Yusuf Amrozi², Faridatun Nadziroh³, Arrio Eka Firmansyah⁴

^{1,2,4}Sistem Informasi, Fakultas Sain dan Teknologi, UIN Sunan Ampel Surabaya, Indonesia

³Teknik Telekomunikasi, Jurusan Teknik Elektro, Politeknik Elektronika Negeri Surabaya (PENS), Indonesia

subhan.nooriansyah@uinsa.ac.id

Abstract

Indonesia's population is projected to reach approximately 287.19 million by 2026. Demographic shifts between 2025 and 2026 could drive increased urbanization in major metropolitan areas, subsequently influencing provincial housing markets. The focus of housing purchase analysis can be expanded to the affordable housing sector by considering factors such as province, property type, price, and property status (subsidized vs. commercial). This study proposes a network clustering method for subsidized affordable housing at the national level, employing community detection techniques to support micro-marketing analysis. Data were obtained from Kaggle, followed by cleaning, feature engineering, and graph construction for community detection using the Leiden algorithm. Clustering results were evaluated using the Silhouette score (to measure inter-cluster separation), the Davies-Bouldin Index (DBI) (to assess intra-cluster proximity), and Normalized Mutual Information (NMI) and Adjusted Rand Index (ARI) (to evaluate community detection performance). The Leiden algorithm yielded the highest modularity (0.976) and the lowest conductance (0.005) among four benchmark methods, indicating the formation of well-defined network communities. It outperformed other methods by achieving the highest NMI and an ARI score of 1.0, although its Silhouette (-0.0196) and DBI (0.685) values were lower than those of KMeans and HDBSCAN. The Leiden method successfully formed cohesive communities—specifically within cluster 437, which included clusters 0 through 9, where affordable and commercial housing were mixed.

Keywords: Affordable Housing, Micro-Marketing Analytics, Clustering, Community Detection, Leiden Algorithm

Abstrak

Populasi Indonesia diproyeksikan mencapai sekitar 287,19 juta jiwa pada tahun 2026. Perubahan demografis dari tahun 2025 hingga 2026 dapat meningkatkan urbanisasi di wilayah metropolitan utama. Hal ini selanjutnya dapat memengaruhi pasar pembelian perumahan di setiap provinsi. Fokus pembelian rumah dapat diperluas ke ranah perumahan terjangkau dengan mempertimbangkan faktor provinsi, jenis properti, harga, serta status properti (bersubsidi atau komersial). Penelitian ini mengusulkan metode pengelompokan jaringan pada perumahan terjangkau dengan kategori bersubsidi pada skala nasional, menggunakan teknik deteksi komunitas untuk mendukung analisis pemasaran mikro. Dalam proses penelitian ini, data dikumpulkan melalui unduhan dari Kaggle. Kemudian, data dibersihkan, dilakukan penentuan rekayasa fitur, dan konstruksi graf untuk deteksi komunitas menggunakan algoritma Leiden. Selanjutnya, hasil pengelompokan dievaluasi menggunakan skor Silhouette untuk mengukur tingkat pemisahan antarkluster, menghitung kedekatan antar data dalam satu kelompok kluster dengan Davies-Bouldin Index (DBI), serta Normalized Mutual Information (NMI) dan Adjusted Rand Index (ARI) untuk mengevaluasi deteksi komunitas. Algoritma Leiden menghasilkan nilai modularitas tertinggi (0.976) dan konduktans terendah (0.005) dibandingkan dengan empat metode pembandingan lainnya; hal ini menunjukkan terbentuknya komunitas jaringan yang terdefinisi dengan baik. Hasil ini mengungguli metode lain dengan mencapai nilai NMI tertinggi dan skor ARI sebesar 1.0, meskipun nilai S -0.0196 dan BDI 0.685

menunjukkan nilai yang lebih rendah dibandingkan dengan KMeans dan HDBSCAN. Metode Leiden masih dapat membentuk komunitas yang utuh melalui klaster 437, di mana hampir semua klaster 0 hingga 9 menunjukkan perumahan terjangkau yang bercampur dengan perumahan komersial.

Kata kunci: Perumahan Terjangkau, Analitik Pemasaran Mikro, Klasterisasi, Deteksi Komunitas, Algoritma Leiden

© 2026 Author

Creative Commons Attribution 4.0 International License



1. Latar Belakang

Indonesia menghadapi tantangan demografis yang krusial dengan jumlah penduduk sekitar 287,19 juta jiwa, berdasarkan proyeksi tahun 2026 dari Badan Pusat Statistik (BPS) [1]. Semakin bertambah jumlah penduduk dan berubah struktur demografis, seperti proporsi penduduk usia produktif dari Generasi Z dan Milenial yang meningkat. Pada saat yang sama, jumlah Generasi X sedikit menurun, sementara proporsi kelompok Baby Boomer berkurang secara perlahan [2]. Dengan perubahan rasio generasi ini, generasi usia produktif berpotensi membeli sebuah rumah berdasarkan faktor-faktor pendukung, seperti proses perpindahan dari rumah orang tua karena pernikahan dan pekerjaan, ataupun faktor lain [3]. Sehingga, pergeseran demografis ini bisa meningkatkan keuntungan dalam pasar pembelian rumah. Tetapi generasi produktif besar kemungkinan tidak mempunyai penghasilan yang layak untuk membeli rumah [4]. Sehingga generasi produktif ini berfokus untuk mencari perumahan terjangkau. Pembelian rumah dapat dikategorikan sebagai hunian terjangkau bergantung pada indikator seperti pendapatan, lokasi, investasi, kredit, subsidi dan pengaruh makro ekonomi yang lain [5]. Berdasarkan faktor-faktor tersebut bisa digunakan dalam proses analitik pemasaran mikro untuk pengelompokan perumahan terjangkau.

Dalam proses pengelompokan segmentasi pemasaran bisa menerapkan metode konvensional klasterisasi yaitu seperti k-Means. Metode ini bisa melakukan segmentasi profil pembelian rumah. Berdasarkan penelitian terkait segmentasi profil pembelian rumah dengan metode perkembangan klaster yang lain seperti Gaussian Mixture Model (GMM) dan Hierarchical Clustering (HC) [6]. Hasil eksperimen menunjukkan bahwa k-Means berhasil mengungguli kedua metode tersebut dalam memisahkan klaster pemasaran berdasarkan profil pembelian rumah. Tetapi konsep yang diterapkan dengan k-Means berfokus pada pemasaran makro karena menerapkan konsep titik pusat atau *centroid*. Selain itu, dataset spasial yang dikelompokkan ke dalam pemasaran mikro bisa ditemukan terpisah dengan klaster yang lain. Kondisi tersebut diartikan sebagai derau atau *noise*. Dalam kondisi ini metode klasterisasi seperti k-Means bisa melakukan klaster, tetapi akan ditemukan data yang tidak masuk ke

dalam klaster dan itu bukan klaster. Kondisi tersebut bisa menerapkan Density-Based Spatial Clustering (DBSCAN). Metode ini berhasil mengungguli k-Means dalam pengelompokan transaksi e-Commerce berdasarkan harga dan jumlah. DBSCAN berhasil dalam menerapkan market mikro dibandingkan dengan k-Means. Namun, DBSCAN dapat terjebak dalam jumlah dimensi yang besar atau *curse of dimensionality* karena nilai pembanding kepadatan yang tidak berubah dan sensitif pada klaster yang terpengaruh oleh derau [7]. Untuk menghadapi kondisi tersebut, DBSCAN ditingkatkan dengan menerapkan Hierarchical pada DBSCAN menjadi HDBSCAN. Metode ini mampu memecah komponen kepadatan menjadi subkelompok dalam struktur *minimum spanning tree*. HDBSCAN dapat mengelompokkan data perumahan terjangkau secara lebih efektif dibandingkan DBSCAN, namun metode ini memiliki beban komputasi yang tinggi untuk kumpulan data berskala besar [8]. Sehingga metode klaster konvensional ketika dihadapkan pada data yang besar maka komputasi yang dibutuhkan tidak kecil.

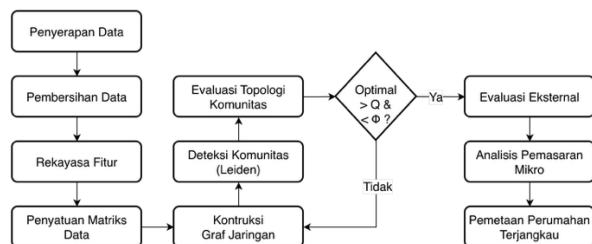
Pengelompokan metode klaster konvensional dapat diperluas menjadi deteksi komunitas, karena pengelompokan yang saling terhubung dalam jaringan yang bermakna disebut sebagai pengelompokan jaringan atau *network clustering*. Algoritma pertama adalah Louvain, yang membangun jaringan lingkungan ketetanggaan atau *neighborhood network* sebelum mengelompokkan simpul-simpul ke dalam grup. Algoritma Louvain digunakan untuk segmentasi guna menghasilkan klaster yang padat dan saling terhubung, mulai dari pemasaran umum hingga pemasaran berskala hampir mikro [9]. Metode lain untuk deteksi komunitas adalah algoritma Leiden, yaitu pengembangan dari algoritma Louvain yang memiliki dasar serupa dalam hal graf jaringan, namun menggunakan rumus yang berbeda untuk menghubungkan simpul-simpul dalam setiap graf [10]. Algoritma Leiden lebih efektif dibandingkan algoritma Louvain dalam menciptakan graf jaringan yang optimal serta mempertahankan jaringan yang telah terbentuk. Proses Leiden dalam membentuk sebuah komunitas yang sempurna tidak membutuhkan komputasi yang tinggi seperti HDBSCAN dan DBSCAN pada data yang besar. Sehingga proses Leiden dapat dikategorikan sebagai

KMeans, dengan kondisi data yang besar hingga kecil. Selain itu, komunitas yang dibangun lebih sempurna dibandingkan komunitas di Louvain. Metode Leiden dapat digunakan untuk membentuk komunitas wisata yang sempurna berdasarkan interaksi pengunjung [11].

Tujuan penelitian ini adalah menerapkan algoritma Leiden untuk deteksi komunitas guna mengelompokkan jaringan perumahan terjangkau di Indonesia bagi keperluan analisis pemasaran mikro. Hasil metode ini dievaluasi menggunakan tiga metrik: pertama, skor *Silhouette* (S) untuk mengukur pemisahan antar-klaster; kedua, nilai maksimum dan minimum modularitas untuk mengukur graf jaringan yang terbentuk; ketiga, *Normalized Mutual Information* (NMI) dan *Adjusted Rand Index* (ARI) untuk mengukur kemampuan deteksi komunitas tersebut dalam memfasilitasi pertukaran informasi antarklaster. Hasil pengelompokan ini dapat dimanfaatkan untuk analisis pemasaran mikro perumahan terjangkau di Indonesia yang dikategorikan menjadi perumahan bersubsidi dan perumahan komersial. Selain itu, hasil klasterisasi Leiden tersebut dapat digunakan untuk menganalisis perumahan bersubsidi dan non-bersubsidi.

2. Metodologi

Penelitian ini menerapkan alur proses analisis pemasaran mikro dengan menggunakan metode deteksi komunitas. Metode ini mengimplementasikan algoritma Leiden. Secara keseluruhan, metodologi penelitian ini disajikan pada Gambar 1.



Gambar 1 Alur Proses Analisis Pemasaran Mikro dengan Menggunakan Deteksi Komunitas

Pada tahap pertama, penyerapan data dilakukan dari Kaggle sebagai tempat berkumpulnya data terbuka. Tahap kedua, pembersihan data, diterapkan untuk membersihkan data yang berpotensi mengganggu proses klasterisasi. Tahap ketiga, rekayasa fitur berguna untuk menentukan fitur-fitur yang diterapkan pada proses klasterisasi pada deteksi komunitas. Tahap keempat, penyusunan matriks data, digunakan untuk menggabungkan semua hasil dari proses rekayasa fitur. Tahap kelima, Konstruksi Graf Jaringan digunakan untuk proses pembuatan konstruksi graf berdasarkan matriks data sebagai awal proses pembentukan komunitas. Tahap keenam, deteksi komunitas menggunakan algoritma Leiden berfungsi untuk mengatur dan mengatur ulang modularitas komunitas. Tahap ketujuh, evaluasi

topologi komunitas dan pengecekan optimalisasi algoritma Leiden berfungsi untuk mengatur bentuk komunitas agar optimal. Pada tahap kedelapan, dalam proses ini dilakukan evaluasi eksternal untuk menguji Deteksi Komunitas Leiden dengan metode seperti KMeans, DBSCAN, HDBSCAN, dan metode Louvain. Pada tahap kesembilan yaitu Analisis Pemasaran Mikro berfungsi untuk menjelaskan bentuk graf jaringan yang membentuk subkomunitas-subkomunitas. Pada tahap kesepuluh, sebagai tahap terakhir memberikan bentuk visualisasi pemasaran mikro untuk perumahan terjangkau di Indonesia.

2.1 Penyerapan Data

Data mengenai perumahan terjangkau dan bersubsidi di Indonesia diperoleh dari Kaggle, sebuah wadah komunitas data terbuka yang menyajikan data hasil pengumpulan pihak ketiga. Kualitas data ini dinilai oleh Kaggle dengan skor kegunaan (*usability*) sebesar 10,00 [10]. Data perumahan terjangkau di Indonesia ini terdiri dari 89.000 baris dan 24 kolom; kolom-kolom tersebut mencakup informasi mengenai lokasi, harga, rata-rata luas lahan, jenis properti, serta kategori kecamatan. Berikut ini adalah data tabular pada Tabel 1 menunjukkan data numerik.

Lintang	Bujur	Harga (Rp)	Luas tanah (m2)	Luas bangunan (m2)
-7.2333	109.0119	166000000	66.0	30.000
-7.3634	109.2802	166000000	66.0	36.000
-7.6533	110.6786	166000000	60.0	30.000
-10.2247	123.6289	185000000	91.0	36.000
-6.2890	106.6174	3164845045	45.0	138.375

Tabel 1 Data Numerik dari 24 kolom

Tabel 2 menunjukkan tipe properti, status data bersubsidi dan tidak bersubsidi, atau komersial. Data yang disajikan tidak mencakup seluruh data; hanya data perwakilan untuk teks dan data kategorikal dari 24 kolom.

Nama Perumahan	Tipe Properti	Status Subsidi	Provinsi	Kota Kabupaten
Grand Satya Bumiayu 3 The Garden	Rumah Tapak	subsidi	JAWA TENGAH	KAB BREBES
GRAND MIRANDA SUMBANG	Rumah Tapak	subsidi	JAWA TENGAH	KAB BANYUM AS
GRHA NUSANTARA KUNCEN	Rumah Tapak	subsidi	JAWA TENGAH	KAB KLATEN
Sunly Permata Bello 6	Rumah Tapak	subsidi	NUSA TENGGARA TIMUR	KOTA KUPANG
Asterra Business Park	Rumah Tapak	komersi l	BANTEN	KAB TANGERA NG

Tabel 2 Data Teks dan Kategorial dari 24 Kolom

Berdasarkan tabel tersebut, data dapat dilihat berdasarkan lokasi perumahan terjangkau antarprovinsi di Indonesia, seperti yang ditunjukkan pada Gambar 2.



Gambar 2 Pemetaan Data Perumahan Terjangkau Antarprovinsi di Indonesia

2.2 Pembersihan Data

Sebelum data digunakan untuk deteksi komunitas dengan algoritma Leiden. Data perumahan terjangkau antarprovinsi di Indonesia dianalisis secara eksploratori dengan menerapkan metode *Exploratory Data Analysis* (EDA) [12]. Hasil EDA digunakan untuk menentukan data yang akan dihapus. Ada 3 hal yang diterapkan dalam proses EDA sebagai berikut: data lokasi yang berada di luar pulau, data yang kosong dalam data tabular, serta harga rumah yang terlalu tinggi dan luas rumah yang terlalu besar. Sehingga proses pembersihan data dapat diterapkan tanpa membuang data secara acak. Setelah itu, hasil tersebut dapat meningkatkan efisiensi proses klasterisasi serta membentuk visualisasi pemetaan perumahan terjangkau.

2.3 Rekayasa Fitur

Pada tahapan ini, 24 kolom tidak semuanya digunakan dalam proses klasterisasi. Sehingga, rekayasa fitur diterapkan untuk membuat metode klaster menjadi efisien dalam proses. Dalam penerapan rekayasa fitur, dilakukan analisis fitur yang tepat untuk klaster. Berikut ini rekayasa fitur yang diterapkan, yaitu fitur pertama, rasio bangunan terhadap tanah sebagai fitur interaksi dengan fitur luas (rasio bangunan terhadap tanah). Fitur kedua adalah kapasitas dan harga per ruangan sebagai interaksi fitur kamar (total ruangan, harga per ruangan). Fitur ketiga: transformasi logaritmik pada harga dan luas, sehingga dapat mengetahui variansi ekstrem pada harga rumah, luas bangunan, luas rumah, dan luas tanah (log harga rumah, log luas tanah, log luas bangunan, log harga per m² tanah). Fitur keempat adalah ekstraksi spasial geografis yang diubah ke radian yang bertujuan untuk analisis spasial dalam proses klasterisasi.

2.4 Penyatuan Matriks Data

Hasil rekayasa fitur dari keempat bagian tersebut akan digabungkan menjadi matriks dengan menerapkan formula (1).

$$M \in \mathbb{R}^{N \times F} \quad (1)$$

Matriks ini dibutuhkan untuk membangun konstruksi graf.

2.5 Konstruksi Graf

Metode Leiden menerapkan metode dasar seperti k-nearest neighbor (kNN) [13]. Metode ini membentuk sekumpulan graf yang saling terhubung. Di mana jarak spasial setiap graf dihitung menggunakan formula (2) [14].

$$d(x_i, x_j) = \|x_i - x_j\|_2 = \sqrt{\sum_{k=1}^d (x_{i,k} - x_{j,k})^2} \quad (2)$$

Setiap lokasi disimbolkan sebagai latitude dan longitude x_i, x_j . Setelah itu, perhitungan himpunan tetangga terdekat pada setiap graf dilakukan menggunakan formula (3) [14]. Dengan menentukan nilai K optimal menggunakan formula (4) [14].

$$N_K(x_i) = \{x_j \in X \setminus \{x_i\} \mid d(x_i, x_j) \leq d_K(x_i)\} \quad (3)$$

$$|N_K(x_i)| = 5 \quad (4)$$

Langkah terakhir dalam konstruksi graf adalah menghitung matriks ketetanggaan terarah menggunakan formula (5) dan matriks ketetanggaan tidak terarah menggunakan formula (6) [15].

$$A_{i,j}^{\text{dir}} = \begin{cases} 1, & \text{jika } x_j \in N_K(x_i) \\ 0, & \text{lainnya} \end{cases} \quad (5)$$

$$A_{i,j} = \max(A_{i,j}^{\text{dir}}, A_{j,i}^{\text{dir}}) \cdot (1 - \delta_{i,j}) \quad (6)$$

Konstruksi graf yang optimal dapat membuat Leiden membentuk komunitas yang optimal dengan K=5.

2.6 Deteksi Komunitas

Metode deteksi komunitas menggunakan algoritma Leiden [10]. Metode ini dapat membentuk komunitas yang sempurna. Selama proses Leiden, menghitung bentuk modular pada komunitas dilakukan dengan formula (7).

$$Q = \frac{1}{2m} \sum_{i,j \in V} \left(A_{i,j} - \frac{k_i k_j}{2m} \right) \delta(\sigma_i, \sigma_j) \quad (7)$$

Dengan keterangan sebagai berikut, $m = \frac{1}{2} \sum_{i,j} A_{i,j}$ sebagai total jumlah sisi dalam graf. $k_i = \sum_j A_{i,j}$ sebagai derajat simpul i . Kategori klaster ditunjukkan pada $\sigma_i \in \{1, \dots, K\}$. Kemudian terdapat kondisi untuk klaster berdasarkan simbol $\delta(\sigma_i, \sigma_j) = 1$ jika $\sigma_i = \sigma_j$, dan 0. Setelah itu, selama proses Leiden membentuk komunitas, dilakukan perubahan dari simpul i menjadi subkomunitas c' sesuai formula (8).

$$\mathbb{P}(i \rightarrow c') = \frac{\exp\left(\frac{1}{\theta} \Delta Q(i \rightarrow c')\right)}{\sum_{c''} \exp\left(\frac{1}{\theta} \Delta Q(i \rightarrow c'')\right)} \quad (8)$$

2.7 Evaluasi Topologi Komunitas dan Pengecekan Optimalisasi modularitas dan konduktans

Setelah proses konstruksi graf dan pembentukan komunitas dengan Leiden, bentuk komunitas akan diuji dengan perhitungan modularitas menggunakan formula (7) untuk mengetahui bentuk komunitas, serta konduktans menggunakan formula (9) untuk mengetahui hubungan antarkomunitas dan subkomunitas.

$$\phi(c_k) = \frac{\sum_{i \in c_k, j \in c_k} A_{ij}}{\min(\text{vol}(c_k), \text{vol}(V \setminus c_k))} \quad (9)$$

2.8 Evaluasi Eksternal

Metode pengukuran evaluasi eksternal, yaitu KMeans, DBSCAN, HDBSCAN, Louvain, dan Leiden, menerapkan 4 formula, yaitu formula pertama, skor Silhouette (10), untuk mengukur kemiripan sampel antarkluster dengan menghitung jarak Euclidean [16].

$$S = \frac{1}{N} \sum_{i=1}^N \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (10)$$

Formula kedua digunakan untuk menghitung kedekatan antarkluster menggunakan metode Indeks Davies-Bouldin (DBI) dengan formula (11) [16].

$$\text{DBI} = \frac{1}{K} \sum_{j=1}^K \max_{j \neq l} \left(\frac{s_i + s_j}{d(\mu_i, \mu_j)} \right) \quad (11)$$

Formula ketiga Normalized Mutual Information (NMI), digunakan untuk menghitung kuantitas informasi yang saling dibagikan antara dua partisi kluster C dan kluster T (12) [17]. $I(C, T)$ sebagai mutual informasi antar partisi kluster yang dihasilkan C dan partisi referensi T . $M(C, T)$ sebagai faktor normalisasi untuk mutual informasi.

$$\text{NMI}(C, T) = \frac{I(C, T)}{M(C, T)} \quad (12)$$

Formula keempat Adjusted Rand Index (ARI) digunakan untuk mengukur tingkat kesesuaian antara partisi komunitas yang dihasilkan oleh algoritma komunitas. Formula ditunjukkan pada formula (13) [18].

$$\text{ARI}(C, T) = \frac{\text{RI} - \mathbb{E}[\text{RI}]}{\max(\text{RI}) - \mathbb{E}[\text{RI}]} \quad (13)$$

2.9 Analisis Pemasaran Mikro

Pada tahap ini, dalam menganalisis pemasaran mikro, dilihat dari bentuk graf jaringan komunitas menggunakan algoritma Leiden. Selain itu, ditambahkan tabel data dari Penyatuan Matriks Data serta penyajian data berdasarkan indikator sebagai berikut: kategori subsidi dan tidak bersubsidi, provinsi, ukuran rumah terjangkau, dan harga rumah.

2.10 Pemetaan Perumahan Terjangkau

Pada tahap terakhir ini, model klasterisasi dengan metode deteksi komunitas menggunakan algoritma

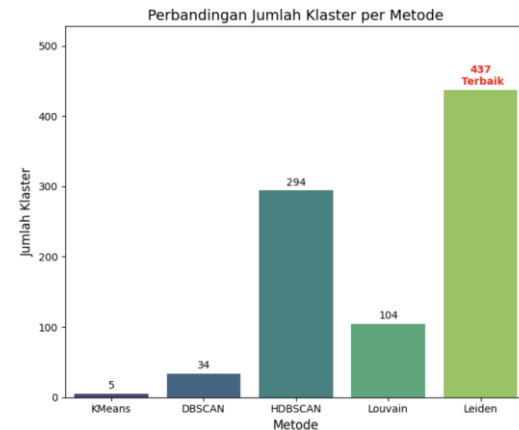
Leiden yang optimal disimpan. Model ini dapat digunakan untuk membuat visualisasi pemetaan perumahan terjangkau berdasarkan hasil analisis pemasaran mikro. Visualisasi diatur sesuai dengan indikator analisis pemasaran mikro. Sehingga, pemetaan ini dapat memberikan gambaran posisi perumahan yang terjangkau. Visualisasi berfokus pada wilayah Indonesia dan bagian ibu kota Indonesia, Jakarta, dan Bandung.

3. Hasil dan Pembahasan

3.1 Hasil

Penelitian ini menyajikan hasil eksperimen dengan mengimplementasikan metode deteksi komunitas menggunakan algoritma Leiden. Hasil eksperimen terbagi menjadi tiga bagian, yaitu lima evaluasi metrik dari lima metode kluster, analisis pemasaran mikro perumahan terjangkau, dan pemetaan pemasaran mikro perumahan terjangkau antarprovinsi di Indonesia.

Hasil eksperimen pertama ditunjukkan pada Gambar 3 yang menunjukkan evaluasi jumlah kluster untuk pemasaran mikro perumahan terjangkau.

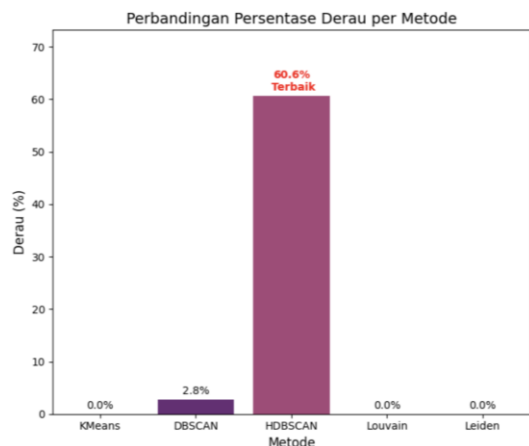


Gambar 3 Perbandingan Jumlah Kluster Leiden dengan KMeans, DBSCAN, HDBSCAN dan Louvain

Kluster yang terbanyak, yaitu Leiden, berjumlah 437. Hal ini menunjukkan bahwa Leiden telah berhasil mengungguli metode kluster lainnya. Terutama, metode yang sensitif terhadap data derau, yaitu HDBSCAN, berada di peringkat 2, sedangkan metode Louvain berada di peringkat 3, yang menerapkan deteksi komunitas tetapi terkendala pada konstruksi graf yang terhubung dengan graf lain secara terbatas.

Pada hasil evaluasi, metrik kedua yang menghitung persentase derau pada kluster ditunjukkan pada Gambar 4. HDBSCAN berhasil menjadi yang terbaik karena dapat memisahkan data kluster dari data derau. Pada metode Leiden, Louvain, dan KMeans, data derau tidak terpisah dari data bukan derau sehingga masuk ke kluster yang sama. Pada peringkat ke-2, yaitu DBSCAN, hanya 2,6 % dari data yang dinyatakan sebagai data derau. Sehingga HDBSCAN

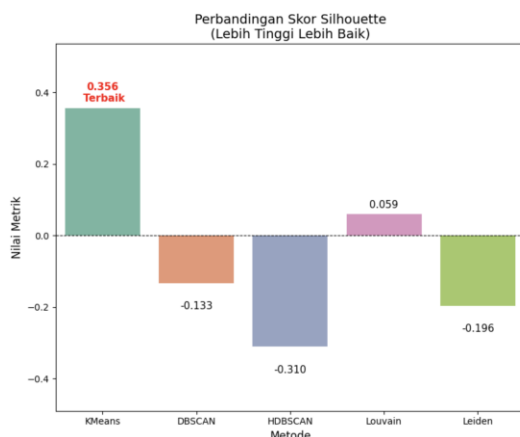
dapat diartikan sebagai sangat sensitif terhadap data luar (derau) dibandingkan dengan Leiden.



Gambar 4 Perbandingan Persentase Derau Leiden dengan Metode lain

Metode Leiden berfokus pada bentuk konstruksi graf yang dapat memberikan hubungan antara graf yang satu dengan yang lain dalam konstruksi graf tersebut. Sehingga data luar dapat diartikan sebagai subkomunitas yang paling ujung atau tidak terhubung secara langsung dengan graf lainnya.

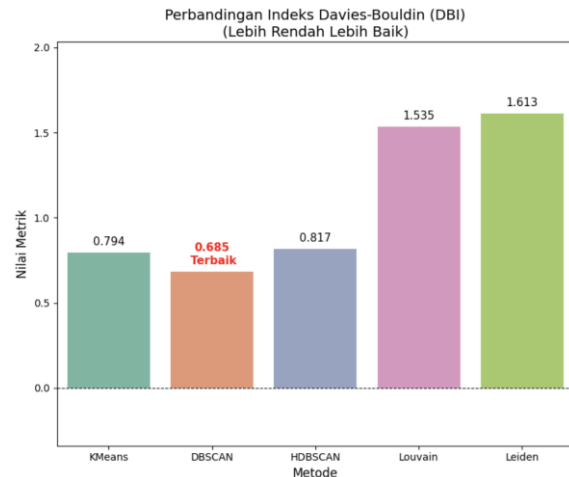
Pada hasil evaluasi metrik ketiga, perhitungan skor *silhouette* ditunjukkan pada Gambar 5. Metode KMeans dapat mengungguli metode Leiden dan terutama ketiga metode lain, yaitu DBSCAN, HDBSCAN, dan metode Louvain. KMeans mempunyai keunggulan dalam menghitung jarak data satu dengan lain, tetapi nilai tersebut menunjukkan jumlah kluster mayor yaitu 5 kluster, yang berfungsi untuk pemasaran makro pada perumahan terjangkau.



Gambar 5 Perbandingan Skor Silhouette

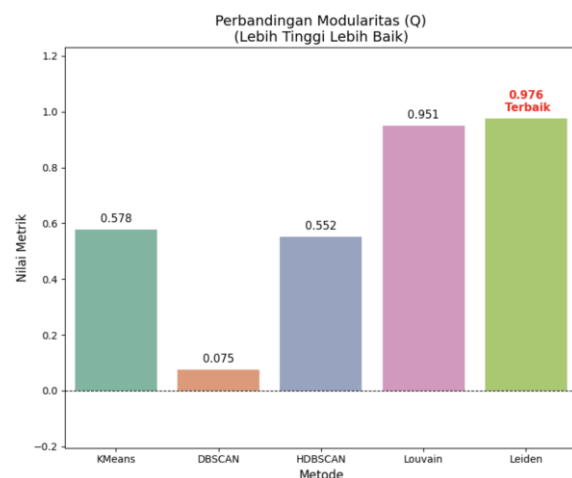
Leiden menunjukkan nilai negatif, yaitu -0.196, karena perhitungan yang digunakan berbeda dari KMeans yang menggunakan jarak Euclidean antarkluster. Metode Leiden menerapkan konstruksi graf untuk menghubungkan graf satu dengan yang lain.

Pada hasil evaluasi metrik keempat yang ditunjukkan pada Gambar 6. Perhitungan DBI terbaik yaitu metode DBSCAN, sedangkan nilai Leiden yang sangat tinggi bukan berarti tidak menunjukkan kualitas internal yang baik. Namun, hal ini menunjukkan bahwa metode Leiden berhasil membentuk konstruksi graf terhubung dari data tersebut.



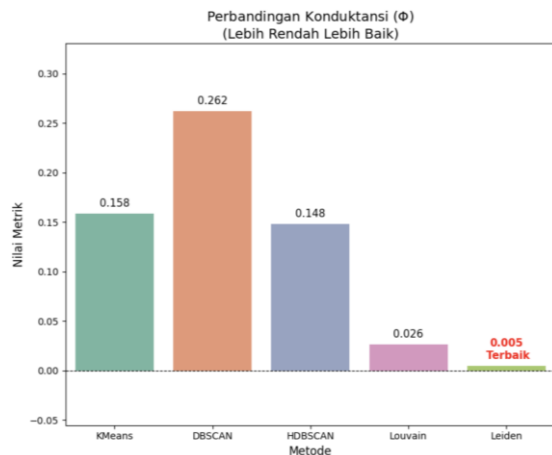
Gambar 6 Perbandingan Indeks Davies-Bouldin (DBI) Leiden dengan metode yang lain

Pada hasil evaluasi metrik kelima dan keenam, pengecekan konstruksi graf yang optimal menunjukkan modularitas dan konduktans yang ditunjukkan pada Gambar 7 dan Gambar 8. Leiden berhasil mencapai nilai modularitas tertinggi dengan komunitas yang terhubung secara maksimal, sedangkan konduktans menunjukkan bahwa setiap konstruksi graf yang terhubung tidak terhubung dengan subkomunitas lain. Jika dibandingkan dengan metode Louvain, ada kondisi yang tidak terhubung.



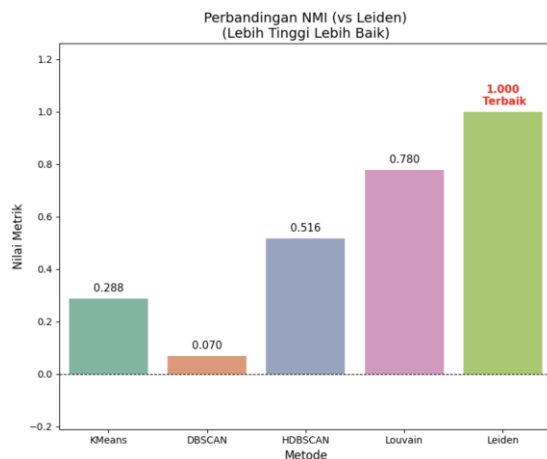
Gambar 7 Perbandingan Modularitas Leiden dengan Metode lain

Pada metode DBSCAN, KMeans, dan HDBSCAN, hubungan antarkluster dapat terputus karena metode ini berfungsi untuk membentuk kluster yang besar, bukan yang kecil, seperti pada Leiden dan Louvain.

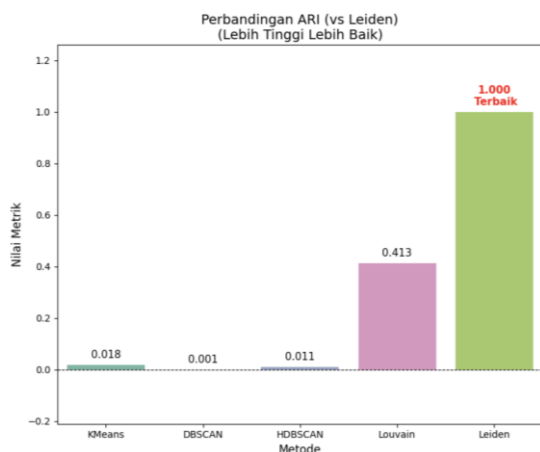


Gambar 8 Perbandingan Konduktansi dengan Metode lain

Evaluasi metrik ketujuh dan kedelapan ditunjukkan pada Gambar 9 dan 10.



Gambar 9 Perbandingan NMI metode lain dengan Leiden

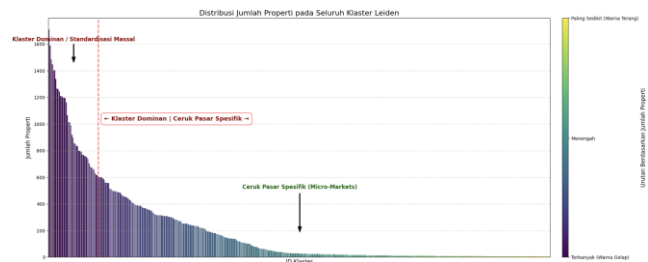


Gambar 10 Perbandingan ARI Leiden dengan metode lain

Metode Leiden yang diusulkan membuat komunitas untuk memetakan pemasaran mikro bagi perumahan terjangkau. Metode ini dapat mengungguli metode klaster konvensional seperti KMeans, serta metode

klaster yang sensitif terhadap data derau seperti DBSCAN dan HDBSCAN. Metode komunitas seperti Louvain tidak dapat mendekati nilai NMI dan ARI seperti Leiden. Hasil dari kedua gambar tersebut menunjukkan bahwa bentuk komunitas Leiden berhasil mendistribusikan konstruksi graf pada hampir seluruh graf meskipun terdapat data derau.

Hasil klaster 437 dari Leiden dapat menunjukkan distribusi jumlah properti dari pemasaran mikro hingga pemasaran makro pada Gambar 11. Yang mana ceruk pasar mikro berada di tengah-tengah grafik, kemudian klaster dominan memisahkan posisi pemasaran makro dan mikro.



Gambar 11 Distribusi Jumlah Properti pada Seluruh Klaster Leiden

Kemudian, berdasarkan distribusi jumlah properti pada grafik, hal tersebut dapat disajikan dalam Tabel 3. Tabel ini disajikan berdasarkan atribut harga jual, luas tanah, luas bangunan, harga per m² tanah, kamar tidur, dan kamar mandi, yang dikelompokkan berdasarkan klaster komersial sebagai bukan subsidi dan klaster 9 sebagai subsidi.

Atribut	Klaster (Komersial)	Klaster (Subsidi)
Harga Jual (IDR)	630,327,075.392	1,206,811,061.906
Luas Tanah (m2)	97.780	90.506
Luas Bangunan (m2)	53.306	66.713
Harga per m2 Tanah (IDR)	6,479,199.181	13,477,540.745
Kamar Tidur	2.207	2.689
Kamar Mandi	1.238	1.996
Total Ruang	3.445	4.685

Tabel 3 Perbandingan Atribut Subsidi dan Komersil pada Klaster 9 dan 0

Harga klaster 9 menunjukkan harga jual rumah terjangkau di atas 1 miliar dengan bantuan subsidi, sedangkan klaster 0 menunjukkan harga jual tanpa bantuan subsidi di bawah 1 miliar. Jumlah perumahan terjangkau yang tergolong cukup mahal memperoleh subsidi, tetapi luas tanahnya tidak seluas perumahan komersial. Tabel 4 menunjukkan rasio antara komersial dan subsidi pada klaster 0 – 6 yang menunjukkan jumlah subsidi lebih kecil dibandingkan dengan klaster. Klaster 378 hingga 437

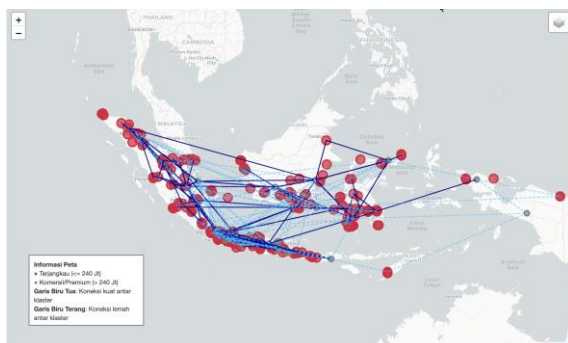
menunjukkan rasio subsidi yang lebih tinggi dibandingkan dengan komersial.

ID Klaster	Non-Subsidi (Komersial)	Subsidi	Total Properti
0	1678	32	1710
3	961	488	1449
4	1361	46	1407
5	1348	55	1403
6	1213	126	1339
...	...	...	...
378	1	7	8
392	2	5	7
407	1	6	7
408	1	6	7
437	2	4	6

Tabel 4 Rincian Porsi Komersial dan Subsidi pada Klaster

Berdasarkan table 4. Ini menunjukkan bahwa pada klaster yang semakin besar, rasio subsidi lebih besar dibandingkan komersial.

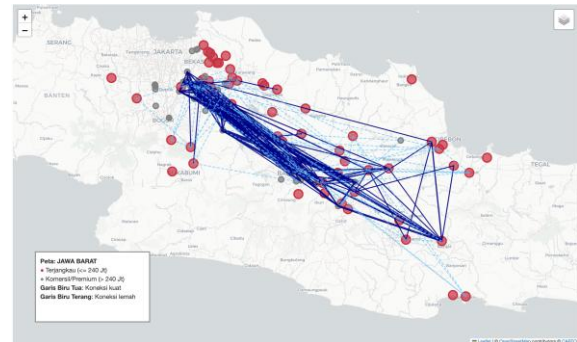
Peta pemasaran mikro untuk perumahan terjangkau dapat dilihat pada Gambar 12 sebagai visualisasi secara keseluruhan provinsi, Pulau Jawa, Sumatra, Kalimantan, Bali, NTT, NTB, Papua, dan Sulawesi, dan pada Gambar 13 sebagai bagian dari Provinsi Jawa Barat.



Gambar 12 Pemetaan Pemasaran Mikro Perumahan Terjangkau Antar provinsi di Indonesia

Pemetaan perumahan terjangkau di atas ditunjukkan pada titik lokasi perumahan berwarna merah dengan harga di bawah 240 juta. Kemudian, warna biru tua menunjukkan koneksi antar titik lokasi sehingga membentuk komunitas rumah terjangkau. Kemudian, lokasi berwarna abu-abu menunjukkan perumahan komersial atau perumahan tanpa subsidi. Setiap provinsi memiliki koneksi komunitas yang lemah dengan provinsi lain. Namun, yang paling lemah terlihat pada perumahan terjangkau di Papua. Jika dibandingkan dengan Sumatra, Kalimantan dan Jawa yang mempunyai hubungan yang kuat. Gambar 13 menunjukkan bagian dari Pulau Jawa, yaitu Jawa

Barat, yang memiliki koneksi kuat dari Jakarta Barat hingga Bandung. Bentuk komunitas ini menunjukkan bahwa daerah perumahan terjangkau berada di luar Jakarta hingga daerah sekitar Bandung dan Kota Bandung. Kota Bandung bisa menjadi wilayah yang masih terjangkau dibandingkan dengan Jakarta secara keseluruhan, karena daerah Jakarta Barat masih terdapat perumahan terjangkau.



Gambar 13 Pemetaan Pemasaran Mikro Perumahan Terjangkau Jawa Barat

3.2 Pembahasan

Berdasarkan hasil evaluasi metrik keenam pada Gambar 3 hingga 10, metode Leiden telah mengungguli modularitas, konduktans, NMI, dan ARI. Hal ini menunjukkan bahwa Leiden dapat membentuk komunitas yang paling terhubung dibandingkan dengan keempat metode tersebut. Namun, Leiden tidak dapat mengungguli metrik skor silhouette dari KMeans dan metrik DBI dari HDBSCAN. Keunggulan ini menunjukkan bahwa Leiden memiliki keunggulan dalam komunitas, yang ditunjukkan oleh kelemahannya terhadap KMeans dan HDBSCAN.

Pada Tabel 3 dan Tabel 4, hasil klaster 0 menunjukkan bahwa bagian pemasaran makro yang timpang secara komersial terhadap perumahan terjangkau bersubsidi. Dengan kata lain, klaster 9 menunjukkan adanya perbandingan rasio yang mulai muncul pada perumahan terjangkau bersubsidi. Sehingga, dari klaster 10 hingga klaster 437, distribusi perumahan terjangkau bersubsidi lebih tinggi dibandingkan perumahan komersial.

Berdasarkan pemetaan pada Gambar 12 dan 13, dapat dianalisis bahwa komunitas yang dibuat oleh Leiden telah menghubungkan subkomunitas yang dapat diartikan sebagai derau. Meskipun ada graf yang tidak terhubung antara Sumatra bagian Banda Aceh, hal ini menunjukkan bahwa nilai modularitas menjadi 1.0 dan konduktans menjadi 0.0. Dengan hasil tersebut, Leiden tidak mencapai nilai yang sempurna, tetapi masih mampu membentuk komunitas yang lebih baik dibandingkan dengan Louvain, KMeans, DBSCAN, dan HDBSCAN.

Metode Leiden dapat digunakan untuk memetakan perumahan terjangkau dalam kategori bersubsidi dan

komersial. Atribut luas tanah dan harga bisa disesuaikan dengan klaster yang dibuat oleh Leiden, karena konstruksi graf yang dibuat akan terhubung dengan graf lain sesuai dengan rekayasa fitur yang diterapkan pada proses awal sebelum klaster. Sehingga, proses Leiden dapat membuat visualisasi pemetaan perumahan terjangkau yang berfokus pada lokasi, harga, kategori bersubsidi, dan luas rumah.

4. Kesimpulan

Penelitian ini menerapkan deteksi komunitas menggunakan algoritma Leiden untuk mengelompokkan pemasaran mikro perumahan terjangkau di Indonesia. Algoritma ini merupakan peningkatan dari algoritma Louvain, di mana komunitas yang dibangun menjadi besar tidak dapat mempertahankan hubungan dengan subkomunitas lain sehingga terputus. Kelebihan Leiden dibandingkan dengan Louvain yaitu keberhasilan dalam membentuk komunitas yang sempurna, di mana selama proses pembentukan komunitas, komunitas tersebut dipecah menjadi graf kecil dan selanjutnya digabungkan kembali sehingga antarsubkomunitas dapat saling terhubung. Komunitas baru yang terhubung ketika bentuk komunitas berfungsi untuk mengukur modularitas pada tingkat maksimum dan minimum dalam suatu jaringan. Setiap simpul dapat dipindahkan ke simpul lain untuk memperoleh struktur jaringan yang optimal setelah deteksi komunitas dilakukan.

Hasil penerapan algoritma Leiden menunjukkan kinerja yang lebih unggul dibandingkan dengan KMeans, DBSCAN, HDBSCAN, dan algoritma Louvain. Evaluasi terhadap algoritma Leiden menghasilkan nilai modularitas tertinggi (0.976) dan nilai konduktans terendah (0.005). Hal ini mengindikasikan terbentuknya komunitas jaringan yang terdefinisi dengan baik, serta tercapainya nilai NMI tertinggi dan nilai ARI sebesar 1.0, berbeda dengan skor Silhouette sebesar -0.0196 dan nilai DBI sebesar 1.613. Proses deteksi komunitas ini berhasil membentuk 437 klaster untuk mengelompokkan pemasaran mikro bagi perumahan terjangkau. Bentuk komunitas yang sempurna ini hanya tidak memiliki koneksi dengan Sumatera, yaitu Banda Aceh. Hal ini dikarenakan konstruksi graf yang sudah dioptimalkan sehingga berfokus pada bagian Sumatera Tengah dan Selatan. Selain itu, pengoptimalan juga terjadi di bagian Jawa Barat, di mana koneksitas berfokus pada Jakarta Barat dan Kota Bandung.

Referensi

- [1] B.-S. Indonesia, "Mid Year Population - Statistical Data." Accessed: Jul. 16, 2026. [Online]. Available: <https://www.bps.go.id/en/statistics-table/2/MTk3NSMy/mid-year-population--thousand-people-.html>

- [2] A. A. Y. Sukmatika, "Perubahan Struktur Demografis Indonesia: Tantangan & Rekomendasi," *HUMANIORUM*, vol. 3, no. 1, pp. 7–15, Jan. 2025, doi: 10.37010/hmr.v3i1.62.
- [3] U. Jannah and J. Adianto, "Getting a Home for Urban Young Adults: A preliminary study of intergenerational housing transitions in Indonesia," vol. 08, no. 02, 2024.
- [4] K. N. Yohanes, H. S. Hamumpuni, and M. P. P. Utami, "The middle-income trap problem faced by millennials and gen z in jakarta: Challenges and mitigating strategies," *Crit. Issue Sustain. Future*, vol. 2, no. 1, pp. 45–65, Feb. 2025, doi: 10.61511/crsusf.v2i1.1803.
- [5] M. Hasanah, H. Permadi, and N. Rista, "Studi Perbandingan Gen Millenial Dan Gen Z Terhadap Daya Beli Dan Minat Konsumen Pada Pt Apfiani Anugerah Properti," *Econ. Digit. Bus. Rev.*, vol. 7, no. 1, pp. 348–373, Sep. 2025, doi: 10.37531/ecotal.v7i1.3115.
- [6] M. D. L. P. Dewi and I. Wasito, "Evaluation of Cluster Models for Creating Profiles of Home Buyers," *Sink. J. Dan Penelit. Tek. Inform.*, vol. 8, no. 4, pp. 2116–2124, Oct. 2024, doi: 10.33395/sinkron.v8i4.13888.
- [7] M. N. Maulana, A. M. R. R. Ilham, and D. I. S. Akis, "Comparative Analysis of K-Means and DBSCAN for Data-Driven E-Commerce Transaction Segmentation Using Price and Quantity Attributes," *Soc. Transform. Data Soc.*, vol. 1, no. 1, pp. 37–51, May 2026, Accessed: Jul. 18, 2026. [Online]. Available: <https://journal.abcollab.id/index.php/STDS/article/view/61>
- [8] R. Ibrahim and A. Aderoba, "Comparison of K-Means and HDBSCAN Clustering Approaches to Enhance Marketing Strategies," vol. 8, no. 3, 2024.
- [9] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech. Theory Exp.*, vol. 2008, no. 10, p. P10008, Oct. 2008, doi: 10.1088/1742-5468/2008/10/P10008.
- [10] V. Traag, L. Waltman, and N. J. van Eck, "From Louvain to Leiden: guaranteeing well-connected communities," Oct. 30, 2019, doi: 10.1038/s41598-019-41695-z.
- [11] A. P. Wardani, T. Trimono, and M. Nasrudin, "Implementation of Multiplex Leiden Algorithm for Clustering Ancol Visitors," *Bit-Tech*, vol. 8, no. 3, pp. 3565–3576, Apr. 2026, doi: 10.32877/bt.v8i3.3608.
- [12] F. Amalia Maresti, G. Mustika Anugraheni, R. A. Hargiyanto, and K. Mustaqim, "PENERAPAN EXPLORATORY DATA ANALYSIS (EDA) DAN ANALISIS RECENCY, FREQUENCY, AND MONETARY (RFM) UNTUK SEGMENTASI PELANGGAN E-COMMERCE," *Competitive*, vol. 19, no. 1, pp. 14–25, Jun. 2024, doi: 10.36618/competitive.v19i1.4059.
- [13] "Scalable Varied-Density Clustering via Graph Propagation." Accessed: Jul. 27, 2026. [Online]. Available: <https://arxiv.org/html/2508.02989v1>
- [14] W.-L. Zhao, H. Wang, and C.-W. Ngo, "Approximate k-NN Graph Construction: a Generic Online Approach," Sep. 17, 2020, *arXiv*: arXiv:1804.03032. doi: 10.48550/arXiv.1804.03032.
- [15] A. Kovalenko, V. Pozdnyakov, and I. Makarov, "Graph Neural Networks with Trainable Adjacency Matrices for Fault Diagnosis on Multivariate Sensor Data," Oct. 20, 2022, *arXiv*: arXiv:2210.11164. doi: 10.48550/arXiv.2210.11164.
- [16] P. M. Hasugian, D. Mathelinea, S. Simamora, and P. B. N. Simangunsong, "Comparative Evaluation of Data Clustering Accuracy through Integration of Dimensionality Reduction and Distance Metric," *MATRIK J. Manaj. Tek. Inform. Dan Rekayasa Komput.*, vol. 24, no. 3, pp. 579–590, Jul. 2025, doi: 10.30812/matrik.v24i3.5057.
- [17] A. D. McCarthy, T. Chen, R. Rudinger, and D. W. Matula, "Metrics matter in community detection," vol. 881, 2020, pp. 164–175. doi: 10.1007/978-3-030-36687-2_14.
- [18] V. Labatut, "Generalized Measures for the Evaluation of Community Detection Methods," *Int. J. Soc. Netw. Min.*, vol. 2, no. 1, p. 44, 2015, doi: 10.1504/IJSNM.2015.069776.