

Analisis Segmentasi Sentra Wisata Kuliner untuk Optimalisasi Omzet UMKM di Surabaya Menggunakan Metode *Agglomerative Hierarchical Clustering*

Nabilah Selayanti^a Shafira Amanda Putri^b Tresna Maulana Fahrudin^c

^{a,b,c}Universitas Pembangunan Nasional "Veteran" Jawa Timur

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 29 Juli 2024

Revisi Akhir: 10 September 2024

Diterbitkan Online: 31 Desember 2024

KATA KUNCI

Agglomerative Hierarchical Clustering,
UMKM, Sentra Wisata Kuliner, Korelasi
Cophenetic, *Silhouette Coefficient*.

KORESPONDENSI

Tresna Maulana Fahrudin
Universitas Pembangunan Nasional "Veteran"
Jawa Timur
Email : tresna.maulana.ds@upnjatim.ac.id

ABSTRACT

Peran Usaha Mikro, Kecil dan Menengah (UMKM) menjadi salah satu peranan yang dominan dalam struktur perekonomian Indonesia. UMKM menghadapi tantangan dalam hal keragaman karakteristik dan kondisi usaha yang berbeda-beda, salah satunya adalah Sentra Wisata Kuliner (SWK). Meskipun banyak SWK terletak di lokasi strategis dengan fasilitas yang memadai, mereka belum memberikan pendapatan optimal bagi para pelaku usaha. Pada penelitian ini dilakukan analisis pengelompokan SWK Usaha Mikro, Kecil, dan Menengah (UMKM) di kota Surabaya Tujuannya adalah untuk mengelompokkan SWK berdasarkan kesamaan karakteristik seperti luas sentra, kapasitas, jumlah pelaku usaha, dan produktivitas menggunakan metode *Agglomerative Hierarchical Clustering* (AHC). Setelah melalui tahapan pra-pemrosesan data, penentuan metode *cluster* terbaik menggunakan korelasi *cophenetic*, dan validasi jumlah *cluster* optimal dengan *silhouette coefficient*, diperoleh hasil pengelompokan yang membagi SWK menjadi 3 *cluster* berdasarkan pengukuran jarak menggunakan *single linkage*, *average linkage*, *complete linkage*, dan *ward linkage*. *Complete linkage* memberikan performa yang baik yakni nilai *cophenetic* sebesar 0.8734 dan nilai *silhouette coefficient* sebesar 0.4864. Interpretasi *cluster* yang didapatkan yakni *cluster* 1 menunjukkan stabilitas dan aktivitas ekonomi tinggi, *cluster* 2 mencakup mencakup sentra yang kurang berkembang, dan *cluster* 3 menunjukkan karakteristik sebagai pusat-pusat usaha yang sangat besar dan berhasil. Hasil pengelompokan ini dapat digunakan sebagai dasar untuk merancang strategi dan program pengembangan UMKM secara lebih efektif dan berkelanjutan.

DOI: <https://doi.org/10.46961/jommit.v8i2>

1. PENDAHULUAN

Usaha Mikro, Kecil, dan Menengah (UMKM) memiliki peran yang sangat signifikan dalam ekonomi Indonesia, baik dalam hal penciptaan lapangan kerja maupun jumlah bisnis yang ada. UMKM mengacu pada usaha-usaha yang dilakukan oleh individu, kelompok, badan usaha kecil, atau bahkan rumah tangga [1]. Secara fundamental, UMKM merupakan fondasi utama dalam struktur ekonomi masyarakat Indonesia, terutama dalam upaya meningkatkan kemandirian ekonomi dan pembangunan. Ini merupakan strategi penting yang dilakukan negara berkembang seperti Indonesia untuk mempercepat

pertumbuhan ekonomi dan meningkatkan kesejahteraan masyarakat melalui sektor ekonomi yang kuat [2].

Peran Usaha Mikro, Kecil dan Menengah (UMKM) menjadi salah satu peranan yang dominan dalam struktur perekonomian Indonesia. Hal ini dapat dilihat dari data Kementerian Koperasi pada tahun 2019, UMKM memiliki peran yang sangat besar dalam perekonomian Indonesia. Jumlah UMKM mencapai 99,99% dari total unit usaha di Indonesia. UMKM juga memberikan kontribusi yang signifikan terhadap Produk Domestik Bruto (PDB), yaitu sekitar 60.51%. Selain itu, UMKM juga berperan dalam penyerapan tenaga kerja, mencapai sekitar 67% dari total tenaga kerja yang tersedia dan berhasil

menghimpun sekitar 60.4% dari total investasi yang ada pada tahun 2020 [3].

Meskipun memiliki peran penting, UMKM menghadapi tantangan dalam hal keragaman karakteristik dan kondisi usaha yang berbeda-beda. Karakteristik kewirausahaan yang dimiliki oleh individu pemilik UMKM merupakan salah satu faktor yang berpengaruh terhadap keberhasilan bisnis yang dijalankan [4]. Perbedaan karakteristik UMKM dengan bisnis lainnya ditentukan oleh sifat-sifat pribadi seorang wirausahawan. Berdasarkan konsep kewirausahaan, yang mengacu pada kemampuan berpikir konsep kewirausahaan, yang mengacu pada kemampuan berpikir secara kreatif dan bertindak inovatif, kemampuan tersebut menjadi landasan, sumber daya, dorongan, serta proses dalam menghadapi tantangan kehidupan [5].

Sentra Wisata Kuliner (SWK) merupakan salah satu dari bentuk UMKM yang sedang berjuang di tengah merebaknya bisnis kuliner modern saat ini. Di Surabaya, beberapa SWK didirikan untuk meningkatkan perekonomian warga dan pariwisata kota juga menghadapi permasalahan serupa. Meskipun banyak SWK terletak di lokasi strategis dengan fasilitas yang memadai, mereka belum memberikan pendapatan optimal bagi para pelaku usaha. Hal ini disebabkan oleh masalah seperti promosi yang kurang, manajemen yang tidak efisien, dan kurangnya inovasi dalam produk dan layanan. Beberapa SWK bahkan mengalami "mati suri" meskipun telah mendapatkan dukungan anggaran yang signifikan dari pemerintah [6].

Untuk memahami dan mengelompokkan Sentra Wisata Kuliner UMKM berdasarkan kesamaan karakteristik usaha mereka, analisis *hierarchical clustering* dapat menjadi salah satu metode yang efektif [7]. Dengan menggunakan teknik ini, SWK dapat dikelompokkan secara bertingkat menjadi *cluster-cluster* yang memiliki karakteristik seperti luas sentra usaha, kapasitas, jumlah pelaku usaha, dan tingkat produktivitas yang serupa. *Clustering* ini memberikan gambaran yang lebih jelas mengenai profil dan kebutuhan setiap kelompok SWK sehingga memudahkan pemerintah atau pemangku kepentingan dalam merancang program pembinaan atau pengembangan yang sesuai.

Oleh karena itu, penelitian ini dilakukan dengan menerapkan analisis *hierarchical clustering* pada Sentra Wisata Kuliner (SWK) di Surabaya. Dengan menganalisis pengelompokan SWK berdasarkan karakteristik usaha mereka, diharapkan dapat diperoleh informasi yang berharga untuk mengidentifikasi potensi dan tantangan yang dihadapi oleh setiap kelompok SWK. Hasil penelitian ini dapat menjadi landasan bagi pemerintah atau lembaga terkait untuk merancang strategi dan program yang lebih efektif dalam mengembangkan SWK di Surabaya secara berkelanjutan.

2. TINJAUAN PUSTAKA

2.1. Usaha Mikro Kecil (UMKM)

Usaha Mikro, Kecil, dan Menengah (UMKM) merupakan sektor bisnis yang mencakup usaha mikro, kecil, dan menengah. UMKM berperan penting dalam perekonomian dengan kontribusi pada pertumbuhan ekonomi, penciptaan lapangan kerja, dan promosi inklusi sosial. Pemerintah dan lembaga lainnya sering mendukung pengembangan UMKM melalui pelatihan kewirausahaan, akses ke modal, bimbingan, dan akses pasar. Dukungan ini bertujuan meningkatkan pertumbuhan UMKM,

mengurangi pengangguran, dan memperkuat ekonomi inklusif [8]. UMKM harus diberdayakan karena memiliki peran strategis dalam mewujudkan perekonomian nasional yang seimbang, berkembang, dan berkeadilan [9].

2.2. Sentra Wisata Kuliner (SWK)

Sentra Wisata Kuliner (SWK) merujuk pada pusat atau kumpulan tempat di mana orang dapat menikmati berbagai jenis makanan dan minuman. Destinasi yang populer bagi yang ingin menikmati pengalaman kuliner yang unik dari berbagai daerah, karena setiap tempat memiliki tradisi kuliner yang berbeda – beda [10]. Pemerintah Kota Surabaya bersama Dinas Koperasi Usaha Kecil dan Menengah dan Perdagangan sebenarnya telah memiliki kebijakan yang sangat baik dalam mengelola permasalahan PKL kuliner dengan membuat sentra wisata kuliner sebagai solusi atas permasalahan terganggunya keindahan kota, terhambatnya aktivitas pejalan kaki di trotoar, dan lalu lintas menjadi tidak lancar [11].

2.3. Normalisasi

Data yang memiliki selisih rentang yang signifikan dapat menghasilkan model dengan akurasi rendah. Normalisasi data merupakan salah satu upaya untuk mengatasi masalah ini. Proses normalisasi tidak mengubah informasi yang terkandung dalam data, namun membantu menyesuaikan rentang nilainya sehingga memudahkan dalam analisis dan penggunaan model. Salah satu metode yang digunakan adalah *Z-Score* normalization. Metode ini dihitung dengan melakukan pencarian nilai ukuran penyimpangan data dari hasil nilai dari rata – rata yang diukur dalam satuan standar deviasi. Berikut adalah rumus metode normalisasi *Z-Score* [12].

$$Z = \frac{x - \mu}{\sigma} \quad (1)$$

Keterangan:

z = Nilai *Z-score*

x = Nilai yang diamati

μ = Rata – rata

σ = Standar deviasi

2.4. Agglomerative Hierarchical Clustering (AHC)

Pada tahap ini, proses *clustering* dilakukan menggunakan metode *Agglomerative Hierarchical Clustering* (AHC). Terdapat tiga metode hierarki dalam AHC yang dapat digunakan yaitu *single linkage*, *complete linkage*, dan *average linkage*. *Single linkage* mengelompokkan berdasarkan jarak terdekat antara anggota *cluster*, *complete linkage* berdasarkan jarak terjauh antara anggota *cluster*, *average linkage* berdasarkan jarak rata-rata antar objek dari semua pasangan titik data dalam dua *cluster* [13]. sedangkan *ward linkage* menggunakan pendekatan di mana *sum of squares* (SSE) dari jarak setiap titik data ke *centroid cluster* diminimalkan, sehingga mengurangi peningkatan varian total saat dua *cluster* digabungkan [14].

a) *Single Linkage*

$$d_{(uv)w} = \min(d_{uw}, d_{vw}) \quad (2)$$

b) *Complete Linkage*

$$d_{(uv)w} = \max(d_{uw}, d_{vw}) \quad (3)$$

c) *Average Linkage*

$$d_{(uv)w} = \frac{d_{uw} + d_{vw}}{n_{(uv)}n_w} \quad [15] \quad (4)$$

Dimana :

- d_{uw} = Jarak antara *cluster* U dan *cluster* W
- d_{vw} = Jarak antara *cluster* V dan *cluster* W
- $d_{(uv)w}$ = Jarak antara *cluster* baru UV dan *cluster* W
- $n_{(uv)}$ = Jumlah titik dalam *cluster* baru UV
- n_w = Jumlah titik dalam *cluster* W

d) *Ward Linkage*

$$SSE = \sum_{i=1}^n (X_i - \bar{X}) - (X_i - \bar{X}) \quad (5)$$

Di mana :

- X_i = Data ke -i dari variabel X
- $cluster$ =
- \bar{X} = Vektor kolom berisi rata-rata nilai pengamatan dalam *cluster*
- n = Jumlah pengamatan

Total jarak terdekat dihitung menggunakan rumus:

$$I = I_1 + I_2 + \dots + I_n \quad (6)$$

Jarak antara pengamatan c_1, c_2 , dan pengamatan $(c_{1,2})$ dengan metode ward yaitu sebagai berikut:

$$I_{(c_1c_2)c_{1,2}} = \frac{n_{c_1} + n_{c_2}}{n_{c_1c_2} + n_{c_{1,2}}} I_{c_1c_{1,2}} + \frac{n_{c_2} + n_{c_{1,2}}}{n_{c_1c_2} + n_{c_{1,2}}} I_{c_2c_{1,2}} \quad (7)$$

Di mana :

- $I_{(c_1c_2)c_{1,2}}$ = Jarak antara *cluster* 1,2 dan (1,2)
- $I_{c_1c_{1,2}}$ = Jarak antara *cluster* 1 dan (1,2)
- $I_{c_2c_{1,2}}$ = Jarak antara *cluster* 2 dan (1,2)
- $I_{c_1c_2}$ = Jarak antara *cluster* 1 dan 2
- $n_{c_1}, n_{c_2}, n_{c_{1,2}}$ = Jumlah pengamatan pada *cluster* ke 1, 2, dan (1,2) secara berurutan [16]

2.5. Koefisien Korelasi Cophenetic

Dalam perbandingan metode statistika, pencarian model terbaik diperlukan. Setelah mendapatkan hasil dari analisis *cluster* hierarki, uji validitas *cluster* diperlukan untuk mengevaluasi kebaikan model hasil analisis *cluster*. Salah satu ukuran yang dapat digunakan untuk menguji validitas hasil *cluster* adalah koefisien korelasi. Koefisien korelasi *cophenetic* adalah ukuran korelasi antara elemen-elemen matriks jarak *Euclidean* asli yang tidak mirip dengan elemen-elemen yang dihasilkan oleh matriks [17]. Berikut merupakan rumus untuk menghitung koefisien korelasi *cophenetic*:

$$r_{coph} = \frac{\sum_{i<j} (d_{ij} - \bar{d})(d_{c_{ij}} - \bar{d}_c)}{[\sum_{i<j} (d_{ij} - \bar{d})^2][\sum_{i<j} (d_{c_{ij}} - \bar{d}_c)^2]} \quad (8)$$

Keterangan:

r_{coph} = Koefisien korelasi *cophenetic*

d_{ij} = Jarak *squared euclidean* antara objek i dan j

\bar{d} = Rata – rata jarak *squared euclidean* antara objek i dan j

$d_{c_{ij}}$ = Jarak *cophenetic* antara objek i dan j

\bar{d}_c = Rata – rata jarak *cophenetic* antara objek i dan j

Koefisien korelasi *cophenetic* memiliki rentang nilai antara -1 dan 1. Ketika nilai r_{coph} mendekati 1, menunjukkan bahwa hasil *clustering* memiliki korelasi yang baik dengan matriks jarak *euclidean* asli, sehingga dapat dianggap sebagai solusi yang baik dari proses *clustering* [18].

2.6. Silhouette Coefficient

Evaluasi dari hasil algoritma Hierarchical penting untuk memahami seberapa baik *cluster* dilakukan. Salah satu metode evaluasi yang umum digunakan adalah *Silhouette Coefficient*. Metode evaluasi ini menggabungkan metode cohesion dan separation untuk mengukur kedekatan objek dalam sebuah *cluster* dan seberapa jauh *cluster* terbaik yang memiliki rata – rata *Silhouette Coefficient* tertinggi. Berikut adalah langkah – langkah dalam menghitung *Silhouette Coefficient*:

- a) Melakukan perhitungan jarak rata – rata antara suatu data, seperti data i, dan semua data lain yang terdapat dalam klaster yang sama (ai).

$$a(i) = \frac{1}{|A| - 1} \sum_j \in_{A, j \neq i} d(i, j) \quad (6)$$

Di mana j adalah data lain yang termasuk dalam satu *cluster* A, dan d(i, j) adalah jarak antara data i dan j.

- b) Hitunglah rata – rata jarak dari data i dengan semua data dalam *cluster* lain, dan pilihlah nilai terkecil dari hasil perhitungan tersebut.

$$d(i, C) = \frac{1}{|A| - 1} \sum_j \in C d(i, j) \quad (7)$$

Di mana d(i,C) adalah jarak rata – rata dari data i dengan semua objek dalam *cluster* lain C, di mana $A \neq C$

$$b(i) = \min C \neq A d(i, C) \quad (8)$$

- c) Rumus *Silhouette Coefficient* adalah

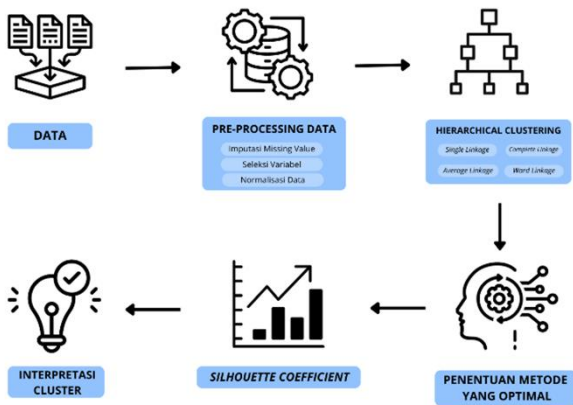
$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (9)$$

Silhouette Coefficient berkisar antara -1 hingga 1, menunjukkan kesesuaian data dalam sebuah *cluster*. Rata – rata *Silhouette Coefficient* mendekati 1 mengindikasikan *cluster* yang baik, sementara nilai mendekati -1 menunjukkan *cluster* yang buru [19].

3. KONSEP PERANCANGAN

Dalam penelitian ini, digunakan metode *Agglomerative Hierarchical Clustering* (AHC). Adapun tahapannya seperti yang

ditunjukkan pada Gambar 1. Pada tahapan *pre-processing* ini akan dilakukan beberapa langkah untuk menghasilkan dataset yang bersih sehingga dapat digunakan dalam tahap berikutnya. Dalam tahapan ini dilakukan memahami tipe data, pemeriksaan terhadap keberadaan *missing value* atau hilang yang hilang dalam dataset, pemilihan variabel yang relevan, dan normalisasi data. Nilai – nilai yang hilang akan diganti dengan nilai rata – rata dari kolom tersebut. Salah satu aspek dari pemilihan variabel adalah penghapusan variabel yang tidak digunakan karena telah terdapat variabel tambahan yang mencakup informasi yang sama atau berkaitan langsung dengan variabel lainnya. Proses seleksi variabel ini penting untuk menyederhanakan analisis dan memastikan bahwa hanya variabel yang paling relevan yang digunakan dalam pemodelan. Normalisasi data bertujuan untuk membuat semua variabel penelitian berada dalam jangkauan yang sama dan memperkecil perbedaan antar variabel penelitian.



Gambar 1. Tahapan Proses *Hierarchical Clustering*

Proses selanjutnya metode dalam analisis *cluster* yang digunakan untuk mengelompokkan data berdasarkan kemiripan karakteristik. Metode – metode tersebut meliputi *Single Linkage*, *Average Linkage*, *Complete Linkage*, dan *Ward Linkage*. *Single Linkage* mengukur jarak antara *cluster* dengan memperhatikan jarak terdekat antara anggota individu di kedua *cluster*, sementara *Average Linkage* menggunakan rata-rata jarak antara semua pasangan anggota *cluster* yang berbeda. Di sisi lain, *Complete Linkage* memperhitungkan jarak terjauh antara anggota di kedua *cluster*, sementara *Ward Linkage* meminimalkan peningkatan varians saat menggabungkan *cluster*.

Tahapan selanjutnya adalah pemilihan metode *cluster* terbaik dengan menggunakan metrik korelasi *cophenetic*. Korelasi *cophenetic* digunakan untuk mengevaluasi seberapa baik dendrogram hasil *clustering* merepresentasikan jarak asli antara data. Tahapan ini melibatkan perhitungan korelasi *cophenetic* untuk masing-masing metode *cluster*, termasuk *Single Linkage*, *Complete Linkage*, *Average Linkage*, dan *Ward Linkage*. Dengan membandingkan nilai korelasi *cophenetic* dari berbagai metode, peneliti dapat menentukan metode *cluster* terbaik yang paling sesuai dengan struktur dan karakteristik data yang ada.

Langkah selanjutnya dalam analisis adalah memvalidasi jumlah *cluster* yang optimal menggunakan *Silhouette Coefficient*. Setelah menentukan metode *cluster* terbaik, tahapan berikutnya adalah menemukan jumlah *cluster* yang paling representatif dalam proses *cluster* menggunakan

metode terbaik. Metode validasi *cluster* yang digunakan dalam penelitian ini adalah koefisien siluet.

Penerapan algoritma *Agglomerative Hierarchical Clustering* untuk melakukan analisis *cluster* terhadap dataset. Menentukan jumlah *cluster* yang diinginkan, dalam penelitian ini menetapkan jumlah *cluster* sebagai 2. Selanjutnya, melatih model *cluster* menggunakan data yang telah dinormalisasi. Setelah model dilatih, mendapatkan label *cluster* untuk setiap sampel data. Kemudian menggabungkan label *cluster* tersebut dengan data asli.

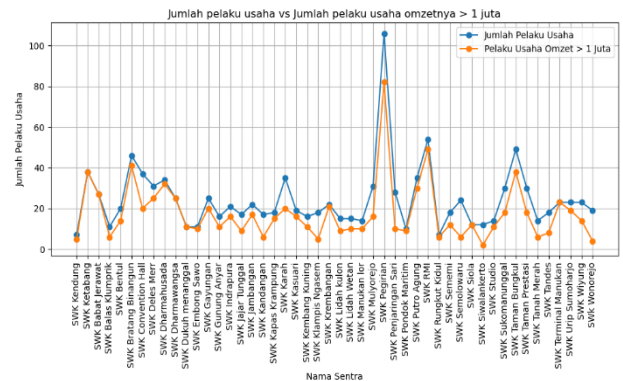
4. HASIL DAN PEMBAHASAN

4.1 Data Penelitian

Data yang akan dianalisis dalam penelitian adalah data jumlah sentra wisata kuliner yang didapatkan dari situs resmi Pemerintah Kota Surabaya yang dapat diakses melalui tautan berikut: <https://opendata.surabaya.go.id/dataset/1500-5245-135>. Data berjumlah 48 *instances* dan memiliki 11 variabel.

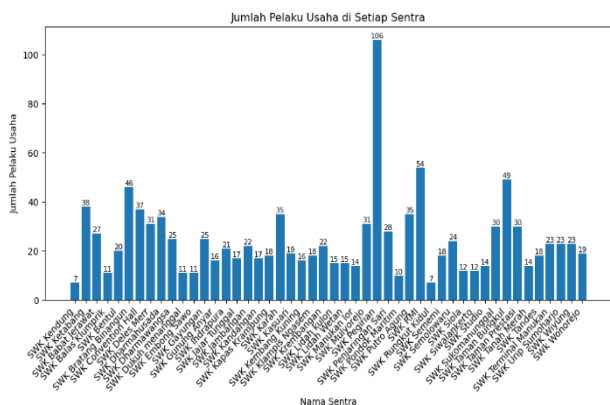
4.2 Statistika Deskriptif

Pada tahap ini dilakukan beberapa statistika deskriptif pada data yang bertujuan untuk memberikan gambaran umum tentang karakteristik data yang akan dianalisis lebih lanjut.



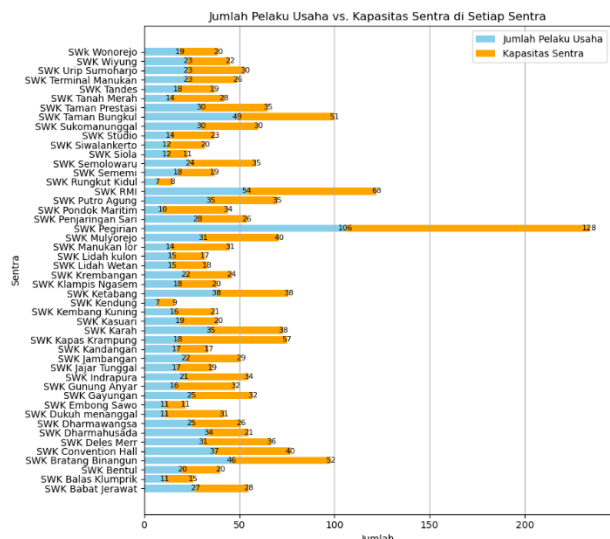
Gambar 2. Grafik Perbandingan Jumlah Pelaku Usaha Dan Jumlah Pelaku Usaha Dengan Omzet > 1 Juta

Berdasarkan Gambar 2, menggambarkan perbandingan antara jumlah pelaku usaha dan jumlah pelaku usaha dengan omzet lebih dari 1 juta pada setiap sentra. Garis biru mewakili jumlah total pelaku usaha, sementara garis oranye mewakili pelaku usaha dengan omzet lebih dari 1 juta. Meskipun keduanya memiliki pola yang serupa, dengan puncak dan lembah yang hampir sama, namun garis oranye selalu berada di bawah garis biru, yang menunjukkan bahwa hanya sebagian kecil pelaku usaha yang memiliki omzet lebih dari 1 juta.



Gambar 3. Histogram Jumlah Pelaku Usaha di Setiap Sentra

Pada Gambar 3, menampilkan variasi dalam jumlah pelaku usaha di setiap sentra. Terlihat bahwa beberapa sentra menonjol dengan jumlah pelaku usaha yang lebih tinggi. Sebagian besar sentra memiliki jumlah pelaku usaha yang relatif rendah hingga sedang, namun terdapat beberapa sentra dengan konsentrasi pelaku usaha yang jauh lebih tinggi.



Gambar 4. Grafik Batang Jumlah Pelaku Usaha dengan Kapasitas Sentra

Berdasarkan Gambar 4, menampilkan perbandingan jumlah pelaku usaha dengan kapasitas sentra, terdapat dua jenis batang dalam grafik tersebut, batang biru menunjukkan jumlah pelaku usaha di setiap sentra, sementara batang kuning menunjukkan kapasitas sentra yang tersedia. Sentra dengan batang biru lebih rendah dari batang kuning menunjukkan kapasitas yang belum terisi sepenuhnya, sementara sentra dengan batang biru lebih tinggi menunjukkan bahwa kapasitas sudah melebihi batasnya.

4.3 Analisis Cluster

4.3.1. Pre-Processing Data

Pada tahap ini, untuk menghasilkan dataset yang bersih dilakukan beberapa langkah sehingga dapat digunakan dalam

tahap berikutnya. Dari data yang telah terkumpul, tidak semua variabel digunakan dalam penelitian ini. Hanya 4 variabel yang dipertahankan, yaitu Luas Sentra, Kapasitas, Jumlah Pelaku Usaha, dan Jumlah Pelaku Usaha yang Omzetnya lebih dari 1 juta.

Pada data tersebut terdapat beberapa nilai yang hilang (*missing value*), oleh karena itu dilakukan imputasi nilai dengan menggunakan rata-rata (*mean*). Setelah proses imputasi, data terlihat seperti Tabel 1 berikut.

Tabel 1. Dataset Setelah Cleaning

No	Luas Sentra (m2)	Kapasitas	Jumlah Pelaku Usaha	Jumlah Pelaku Usaha Omzetnya > 1 juta
1	213,63000	9	7	5
2	569,30000	38	38	38
3	1412,00000	28	27	27
4	560,00000	15	11	6
5	227,50000	20	20	14
...
44	150,00000	19	18	8
45	488,00000	26	23	23
46	1134,00000	30	23	19
47	357,75000	22	23	14
48	350,75000	20	19	4

Setelah melakukan imputasi untuk menangani nilai-nilai yang hilang (*missing value*), tahap selanjutnya adalah melakukan transformasi data dengan teknik normalisasi. Data yang telah diimputasi tersebut akan dinormalisasikan menggunakan metode z-score atau StandardScaler sehingga menghasilkan nilai yang memiliki skala dari 0 sampai 1.

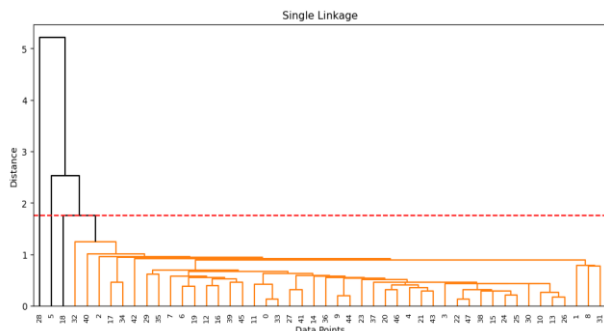
Tabel 2. Dataset Setelah Normalisasi

No	Luas Sentra (m2)	Kapasitas	Jumlah Pelaku Usaha	Jumlah Pelaku Usaha Omzetnya > 1 juta
1	-1,162807	1,128038	-1,090876	-0,898956
2	-0,241640	0,423571	0,860114	1,462481
3	1,940910	-0,111466	0,167827	0,675335
4	-0,265727	-0,807015	-0,839135	-0,827398
5	-1,126885	-0,539496	-0,272719	-0,254928
...
44	-1,327606	-0,593000	-0,398589	-0,684280
45	-0,452203	-0,218474	-0,083914	0,389100
46	1,220904	-0,004459	-0,083914	0,102866
47	-0,789544	-0,432489	-0,083914	-0,254928
48	-0,807673	-0,539496	-0,335654	-0,970515

4.3.2. Metode Single Linkage

Metode *Single Linkage* dalam pengelompokan data didasarkan pada konsep jarak minimum atau jarak terdekat antara

objek-objek yang akan dikelompokkan. Hasil dari penerapan metode pengelompokan *Single Linkage* pada data ini dapat dilihat pada gambar berikut.



Gambar 5. Dendrogram Metode *Single Linkage*

Cluster 1 : SWK Kendung, SWK Ketabang, SWK Babat Jerawat, SWK Balas Klumprik, SWK Bentul, SWK Convention Hall, SWK Deles Merr, SWK Dharmahusada, SWK Dharmawangsa, SWK Dukuh menanggal, SWK Embong Sawo, SWK Gayungan, SWK Gunung Anyar, SWK Indrapura, SWK Jajar Tunggal, SWK Jambangan, SWK Kandangan, SWK Kapas Krampung, SWK Karah, SWK Kasuari, SWK Kembang Kuning, SWK Klampis Ngasem, SWK Krembangan, SWK Lidah kulon, SWK Lidah Wetan, SWK Manukan lor, SWK Mulyorejo, SWK Penjaringan Sari, SWK Pondok Maritim, SWK Putro Agung, SWK RMI, SWK Rungkut Kidul, SWK Sememi, SWK Semolowaru, SWK Siola, SWK Siwalankerto, SWK Studio, SWK Sukomanunggal, SWK Taman Bungkul, SWK Taman Prestasi, SWK Tanah Merah, SWK Tandes, SWK Terminal Manukan, SWK Urip Sumoharjo, SWK Wiyung, SWk Wonorejo

Cluster 2 : SWK Bratang Binangun

Cluster 3 : SWK Pegirian

4.3.3. Metode Average Linkage

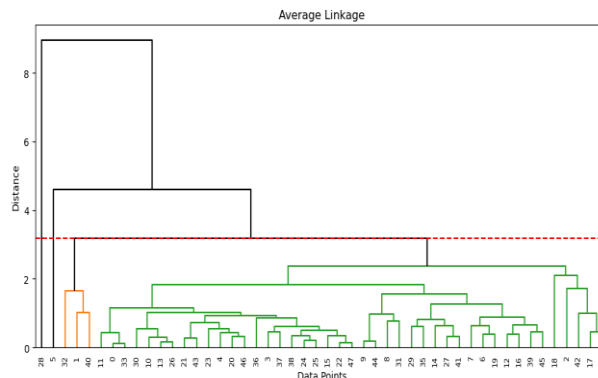
Metode *Average Linkage* dalam pengelompokan data beroperasi dengan menghitung jarak antara dua kelompok (*cluster*) berdasarkan rata-rata jarak antar objek. Jarak ini dihitung dengan meminimalkan rata-rata jarak antara setiap objek dalam kelompok yang digabungkan. Berikut adalah hasil pengelompokan data menggunakan metode *Average Linkage*.

Cluster 1 : SWK Kendung, SWK Ketabang, SWK Babat Jerawat, SWK Balas Klumprik, SWK Bentul, SWK Convention Hall, SWK Deles Merr, SWK Dharmahusada, SWK Dharmawangsa, SWK Dukuh menanggal, SWK Embong Sawo, SWK Gayungan, SWK Gunung Anyar, SWK Indrapura, SWK Jajar Tunggal, SWK Jambangan, SWK Kandangan, SWK Kapas Krampung, SWK Karah, SWK Kasuari, SWK Kembang Kuning, SWK Klampis Ngasem, SWK Krembangan, SWK Lidah kulon, SWK Lidah Wetan, SWK Manukan lor, SWK Mulyorejo, SWK Penjaringan

Sari, SWK Pondok Maritim, SWK Putro Agung, SWK RMI, SWK Rungkut Kidul, SWK Sememi, SWK Semolowaru, SWK Siola, SWK Siwalankerto, SWK Studio, SWK Sukomanunggal, SWK Taman Bungkul, SWK Taman Prestasi, SWK Tanah Merah, SWK Tandes, SWK Terminal Manukan, SWK Urip Sumoharjo, SWK Wiyung, SWk Wonorejo

Cluster 2 : SWK Bratang Binangun

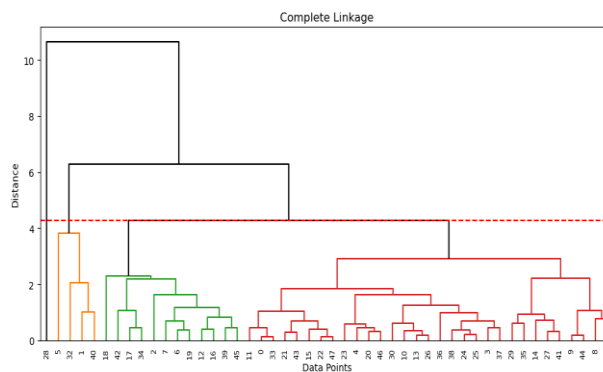
Cluster 3 : SWK Pegirian



Gambar 6. Dendrogram Metode *Average Linkage*

4.3.4. Metode Complete Linkage

Metode *Complete Linkage* dalam pengelompokan data menggunakan jarak terjauh antara anggota dari satu kelompok dengan anggota kelompok lainnya. Hasil dari penerapan metode pengelompokan *Complete Linkage* pada gambar berikut.



Gambar 7. Dendrogram Metode *Complete Linkage*

Cluster 1 : SWK Ketabang, SWK Bratang Binangun, SWK RMI, SWK Taman Bungkul

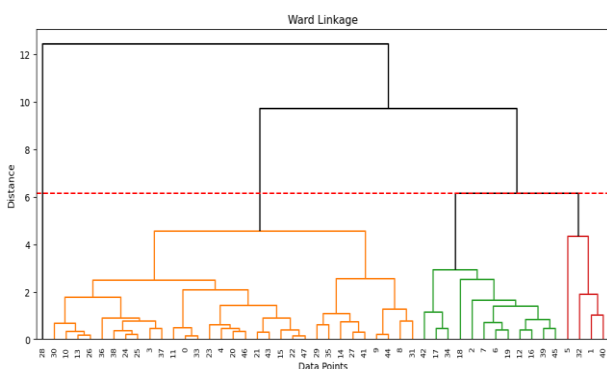
Cluster 2 : SWK Kendung, SWK Babat Jerawat, SWK Balas Klumprik, SWK Bentul, SWK Convention Hall, SWK Deles Merr, SWK Dharmahusada, SWK Dharmawangsa, SWK Dukuh menanggal, SWK Embong Sawo, SWK Gayungan, SWK Gunung Anyar, SWK Indrapura, SWK Jajar Tunggal, SWK Jambangan, SWK Kandangan, SWK Kapas Krampung, SWK Karah, SWK Kasuari, SWK Kembang Kuning, SWK Klampis Ngasem, SWK Krembangan, SWK Lidah kulon, SWK Lidah

Wetan, SWK Manukan lor, SWK Mulyorejo, SWK Penjaringan Sari, SWK Pondok, Maritim, SWK Putro Agung, SWK Rungkut Kidul, SWK Sememi, SWK Semolowaru, SWK Siola, SWK Siwalankerto, SWK Studio, SWK Sukomanunggal, SWK Taman Prestasi, SWK Tanah Merah, SWK Tandes, SWK Terminal Manukan, SWK Urip Sumoharjo, SWK Wiyung, SWK Wonorejo

Cluster 3 : SWK Pegirian

4.3.5. Metode Ward Linkage

Metode Ward Linkage dalam pengelompokan data berfokus pada penggabungan kelompok dengan pertumbuhan variasi terendah, menghasilkan kelompok yang homogen dengan varian yang rendah. Hasil dari penerapan metode pengelompokan Ward Linkage pada gambar berikut.



Gambar 8. Dendrogram Metode Ward Linkage

Cluster 1 : SWK Kendung, SWK Balas Klumprik, SWK Bentul, SWK Dharmahusada, SWK Dharmawangs, SWK Dukuh menanggal, SWK Embong Sawo, SWK Gunung Anyar, SWK Indrapura, SWK Jajar Tunggal, SWK Kasuari, SWK Kembang Kuning, SWK Klampis Ngasem, SWK Krembangan, SWK Lidah kulon, SWK Lidah Wetan, SWK Manukan lor, SWK Mulyorejo, SWK, Penjaringan Sari, SWK Pondok Maritim, SWK Putro Agung, SWK Rungkut Kidul, SWK Semolowaru, SWK Siola, SWK Siwalankerto, SWK Studio, SWK Taman Prestasi, SWK Tandes, SWK Terminal Manukan, SWK Wiyung, SWK Wonorejo

Cluster 2 : SWK Ketabang, SWK Babat Jerawat, SWK Bratang Binangun, SWK Convention Hall, SWK Deles Merr, SWK Gayungan, SWK Jambangan, SWK Kandangan, SWK Kapas Krampung, SWK Karah, SWK RMI, SWK Sememi, SWK Sukomanunggal, SWK Taman Bungkul, SWK Tanah Merah, SWK Urip Sumoharjo

Cluster 3 : SWK Pegirian

4.3.6. Penentuan Metode yang Optimal

Untuk memastikan proses pengelompokan yang optimal, dalam penelitian ini menerapkan validasi untuk mengevaluasi berbagai algoritma Agglomerative Hierarchical Clustering berdasarkan kemampuannya dalam mengukur kesamaan dengan tepat dan menentukan jumlah cluster terbaik. Koefisien korelasi *cophenetic* berperan sebagai indikator validasi untuk menilai baik metode pengukuran jarak maupun algoritma pengelompokan.

Tabel 3. Dataset Penentuan Algoritma yang Optimal

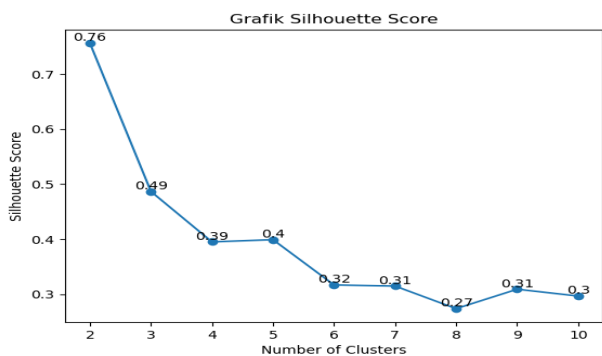
Algoritma	Nilai Cophenetic	Silhouette Coefficient	Proporsi Cluster
Single Linkage	0,9111	0,5461	Cluster 1 : 46 anggota Cluster 2 : 1 anggota Cluster 3 : 1 anggota
Average Linkage	0,9317	0,5461	Cluster 1 : 46 anggota Cluster 2 : 1 anggota Cluster 3 : 1 anggota
Complete Linkage	0,8734	0,4864	Cluster 1 : 4 anggota Cluster 2 : 43 anggota Cluster 3 : 1 anggota
Ward Linkage	0,6368	0,3954	Cluster 1 : 31 anggota Cluster 2 : 16 anggota Cluster 3 : 1 anggota

Berdasarkan Tabel 3, diketahui bahwa algoritma *average linkage* memiliki nilai *cophenetic* tertinggi sebesar 0,9317. Meskipun nilai *cophenetic* yang tinggi menunjukkan kemampuan metode dalam menangkap kemiripan jarak dalam data, hasil *clustering* yang dihasilkan tidak selalu bermakna secara praktis. Dalam kasus ini, *average linkage* menghasilkan cluster dimana cluster 2 dan cluster 3 masing-masing hanya memiliki satu anggota, yang tidak ideal untuk dianalisis lebih mendalam. Nilai *silhouette coefficient* sebesar 0,5461 juga menunjukkan bahwa kualitas *clustering* tidak optimal, mengingat banyaknya cluster dengan hanya satu anggota.

Oleh karena itu, peneliti memilih untuk menggunakan metode *complete linkage*, yang memiliki nilai *cophenetic* tertinggi ketiga sebesar 0,8734 dan nilai *Silhouette Coefficient* sebesar 0,4864. Walaupun nilai *cophenetic*-nya sedikit lebih rendah, *complete linkage* menghasilkan distribusi cluster yang lebih seimbang dan representatif, dimana cluster 1 terdiri dari 4 anggota, cluster 2 terdiri dari 43 anggota, dan cluster 3 terdiri dari 1 anggota. Meskipun nilai *Silhouette Coefficient*-nya lebih rendah dibandingkan dengan *average linkage*, distribusi cluster yang dihasilkan oleh *complete linkage* lebih bermakna dan dapat digunakan untuk analisis lebih lanjut.

4.3.7. Jumlah Cluster

Setelah mengidentifikasi algoritma pengelompokan yang optimal yaitu *complete linkage*, langkah selanjutnya adalah menentukan jumlah cluster yang paling representatif untuk proses pengelompokan *complete linkage*. Untuk memvalidasi efektivitas pembentukan cluster menggunakan metode *Silhouette Coefficient*. Nilai *Silhouette Coefficient* disajikan dalam Gambar berikut.



Gambar 7. Grafik Silhouette Coefficient

Silhouette Coefficient yang berkisar dari -1 hingga 1, menunjukkan efektifitas pembentukan cluster. Jumlah cluster yang representatif dapat dilihat dengan Silhouette Coefficient yang mendekati 1. Berdasarkan Gambar 7, nilai-nilai Silhouette Coefficient menyarankan bahwa 2 cluster mewakili pengelompokan optimal, dengan nilai tertinggi mencapai 0.76. Oleh karena itu, Silhouette Coefficient yang diperoleh termasuk dalam kategori good cluster.

Namun, setelah mempertimbangkan hasil dendrogram dan distribusi cluster, jumlah cluster 2 tidak merepresentasikan pengelompokan yang bermakna. Oleh karena itu, peneliti memilih untuk menggunakan 2 cluster sebagai jumlah yang optimal. Meskipun nilai Silhouette Coefficient untuk 3 cluster adalah 0,49 hasil ini memberikan distribusi cluster yang lebih seimbang dan bermakna untuk analisis lebih lanjut, sesuai dengan hasil validasi menggunakan koefisien korelasi cophenetic pada Tabel 3. Dengan demikian, 3 cluster dianggap sebagai pilihan yang lebih optimal untuk pengelompokan sentra wisata kuliner di Surabaya.

4.4 Interpretasi Cluster

Setelah melakukan pengelompokan data menggunakan tiga metode hierarkis, yaitu single linkage, average linkage, dan complete linkage, ward linkage ditemukan bahwa metode complete linkage memberikan hasil pengelompokan yang paling optimal setelah mempertimbangkan hasil dendrogram dan distribusi cluster. Oleh karena itu, complete linkage dipilih sebagai metode terbaik untuk mengelompokkan data sentra wisata kuliner ini. Selanjutnya, untuk mendeskripsikan karakteristik setiap kelompok atau cluster yang terbentuk dari metode complete linkage, dilakukan analisis dengan melihat nilai rata-rata dari variabel-variabel indikator yang digunakan dalam pengelompokan sentra wisata kuliner tersebut. Pendekatan ini dilakukan untuk memberikan gambaran yang lebih jelas dan rinci mengenai ciri-ciri spesifik yang membedakan setiap cluster yang terbentuk.

Tabel 4. Rata-Rata Cluster terbentuk

Variabel	Cluster 1	Cluster 2	Cluster 3
Luas Sentra (m2)	0,803857	-0,091534	0,720526
Kapasitas	1,186000	-0,232161	5,238910
Jumlah Pelaku Usaha	1,410796	-0,250765	5,139703

Jumlah Pelaku Usaha Omzetnya > 1 juta	Rata-Rata		
	1,712937	-0,266577	4,611064
	1,278397	-0,210259	3,927551

Berdasarkan Tabel 4 diketahui bahwa pada data sentra wisata kuliner UMKM di Surabaya menghasilkan 3 cluster dengan karakteristik yang berbeda. Cluster 1 terdiri dari sentra wisata kuliner yang berkembang baik. Sentra-sentra dalam cluster 1 ini memiliki luas sentra, kapasitas, jumlah pelaku usaha, dan jumlah pelaku usaha dengan omzet lebih dari 1 juta rupiah yang berada di atas rata-rata, yang menunjukkan bahwa mereka adalah pusat usaha yang maju dan stabil. Cluster 2 mencakup sentra wisata kuliner yang kurang berkembang, dengan semua variabel di bawah rata-rata. Sentra-sentra ini mungkin masih dalam tahap awal pengembangan atau menghadapi tantangan dalam menarik dan mempertahankan pelaku usaha. Sementara itu, Cluster 3 terdiri dari sentra wisata kuliner yang sangat sukses dan berpengaruh, dengan kapasitas dan aktivitas ekonomi yang sangat tinggi. Sentra ini menunjukkan keberhasilan signifikan dalam menarik dan mempertahankan pelaku usaha yang berhasil.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis yang diperoleh :

1. Hasil pengelompokan dengan menggunakan metode complete linkage diperoleh tiga cluster. Cluster 1 terdiri dari 4 sentra yang berkembang baik dengan luas sentra, kapasitas, dan jumlah pelaku usaha yang di atas rata-rata, menunjukkan stabilitas dan aktivitas ekonomi tinggi. Cluster 2 mencakup 43 sentra yang kurang berkembang dengan semua variabel di bawah rata-rata, mencerminkan tantangan dalam pengembangan dan daya tarik ekonomi. Dan cluster 3 terdiri dari 1 sentra yang sangat sukses dan berpengaruh dengan kapasitas dan jumlah pelaku usaha yang jauh di atas rata-rata, menunjukkan keberhasilan dan kontribusi signifikan terhadap ekonomi lokal.
2. Metode complete linkage yang dipilih menunjukkan pengelompokan yang lebih seimbang dan representatif dibandingkan metode lain. Meskipun metode average linkage memiliki nilai cophenetic tertinggi, hasil clusteringnya kurang representatif dengan hanya satu anggota di beberapa cluster. Complete linkage memberikan pembagian yang lebih bermakna dengan distribusi anggota yang lebih seimbang di setiap cluster, sehingga lebih baik dalam menggambarkan kondisi nyata dari sentra-sentra wisata kuliner UMKM di Surabaya.

Saran untuk mendukung pengembangan sentra wisata kuliner di Surabaya didasarkan pada analisis clustering. Langkah-langkah strategis mencakup fokus pada pengembangan Cluster 2 yang kurang berkembang dengan peningkatan infrastruktur, program pelatihan, dan strategi promosi. Untuk Cluster 1 yang sudah berkembang, perlu diterapkan langkah-langkah pengembangan berkelanjutan melalui program-program pengembangan dan penggunaan inovasi. Sedangkan untuk Cluster 3 yang sangat sukses, optimalisasi dapat dilakukan melalui ekspansi dan diversifikasi usaha serta memanfaatkan sentra-sentra dalam cluster ini sebagai model pengembangan bagi sentra lain.

Diharapkan strategi ini akan mendukung pertumbuhan ekonomi lokal dan meningkatkan daya tarik wisata kuliner di Surabaya.

DAFTAR PUSTAKA

- [1] Y. Khoiruddin, "Perancangan Aplikasi E-Commerce Pada Umkm Di Desa Gempolkarya," *Abdimas Jurnal Pengabdian Mahasiswa*, vol. 2, no. 1, pp. 719–725, 2023, [Online]. Available: <https://journal.ubpkarawang.ac.id/index.php/AJPM/article/view/3656>
- [2] S. Vinatra, "Peran Usaha Mikro, Kecil, dan Menengah (UMKM) dalam Kesejahteraan Perekonomian Negara dan Masyarakat," *Jurnal Akuntan Publik*, vol. 1, no. 3, pp. 1–08, 2023, doi: 10.59581/jap-widyakarya.v1i1.832.
- [3] A. Subardjo and M. I. Rahmawati, "Inovasi Model Bisnis Inkubasi dan Kolaborasi dalam Meningkatkan Kinerja UMKM Surabaya di Era 4.0," *Jurnal Penelitian Inovatif*, vol. 2, no. 1, pp. 17–24, Jan. 2022, doi: 10.54082/jupin.36.
- [4] S. Almaidah and T. Enderwati, "Analisis Pengaruh Karakteristik Wirausaha, Modal, Motivasi, Pengalaman, Dan Kemampuan Usaha Terhadap Keberhasilan Usaha Pada UKM Penghasil Mete Di Kabupaten Wonogiri," *Edusaintek*, vol. 3, pp. 111–124, 2019, [Online]. Available: <http://prosiding.unimus.ac.id>
- [5] Maisaroh, "Kajian Karakteristik Kewirausahaan Terhadap Keberhasilan Usaha UKM (Studi Kasus Sentra Industri Konveksi Dusun Mlangi Dan Sawahan Nogotirto Gamping Sleman Yogyakarta)," 2019. doi: <https://doi.org/10.32424/jeba.v2i1i2>.
- [6] Y. Ermawati, M. Sodikin, and E. Supeni, "Strategi Pemberdayaan UMKM Berbasis Sentra Wisata Kuliner Di Surabaya," *Seminar Nasional Teknologi dan Multidisiplin Ilmu*, vol. 2, pp. 390–404, 2022, [Online]. Available: <https://prosiding.stekom.ac.id/index.php/semnastekmu>
- [7] K. Pratama Simanjuntak and U. Khaira, "Pengelompokan Titik Api di Provinsi Jambi dengan Algoritma Agglomerative Hierarchical Clustering," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 1, pp. 7–16, 2021, doi: <https://doi.org/10.57152/malcom.v1i1.6>.
- [8] M. T. Febriyanto, I. Harris, D. Sundiman, M. R. Pradana, and E. Lestari, "Pelatihan Kewirausahaan dan Peningkatan Kualitas Manajemen dan Tata Kelola Keuangan Bagi Pelaku UMKM di Lingkungan PKK Tiban Global Batam," *Jurnal Pengabdian Kepada Masyarakat*, vol. 2, pp. 271–279, 2019, [Online]. Available: <http://ejournal.bsi.ac.id/ejournal/index.php/abdimas>
- [9] Y. R. Suci, "Perkembangan Umkm (Usaha Mikro Kecil Dan Menengah) Di Indonesia," 2017. [Online]. Available: <https://ejournal.upp.ac.id/index.php/Cano/article/view/1239>
- [10] Y. Ermawati, M. Sodikin, and E. Supeni, "Strategi Pemberdayaan UMKM Berbasis Sentra Wisata Kuliner Di Surabaya," *Seminar Nasional Teknologi Dan Multidisiplin Ilmu Semnastekmu*, vol. 2, no. Vol 2 No 2 (2022): SEMNASTEKMU, pp. 390–404, 2022, [Online]. Available: <https://prosiding.stekom.ac.id/index.php/semnastekmu>
- [11] M. Sholeh, D. Andayati, and R. Yuliana Rachmawati, "Data Mining Model Klasifikasi Menggunakan Algoritma K-Nearest Neighbor Dengan Normalisasi Untuk Prediksi Penyakit Diabetes," 2022. doi: <https://doi.org/10.36342/teika.v12i02.2911>.
- [12] R. O. Pratikto and N. Damastuti, "Klasterisasi Menggunakan Agglomerative Hierarchical Clustering Untuk Memodelkan Wilayah Banjir," 2018. [Online]. Available: <http://publishing-widyagama.ac.id/ejournal-v2/index.php/jointecs>
- [13] N. Ulinnuh and R. Veriani, "Analisis Cluster dalam Pengelompokan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode Complete Linkage, Average Linkage dan Ward," *Jurnal Nasional Informatika dan Teknologi Jaringan*, vol. 5, 2020, doi: 10.30743/infotekjar.v5i1.2464.
- [14] D. Iyan Yulianti, T. Iman Hermanto, and M. Defriani, "Analisis Clustering Donor Darah dengan Metode Agglomerative Hierarchical Clustering," *RESOLUSI: Rekayasa Teknik Informatika dan Informasi*, vol. 3, no. 6, pp. 285–290, 2023, [Online]. Available: <https://djournals.com/resolusi>
- [15] Imasiani, I. Purnamasari, and A. Fidia Deny Tisna, "Perbandingan Hasil Analisis Cluster Dengan Menggunakan Metode Average Linkage Dan Metode Ward (Studi Kasus: Kemiskinan Di Provinsi Kalimantan Timur Tahun 2018)," *Jurnal EKSPONENSIAL*, vol. 13, no. Vol 13 No 1 (2022), pp. 9–18, 2022, doi: <https://doi.org/10.30872/eksponensial.v13i1.875>.
- [16] A. F. Dewi and K. Ahadiyah, "Agglomerative Hierarchy Clustering Pada Penentuan Kelompok Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Pendidikan," *Zeta - Math Journal*, vol. 7, no. 2, pp. 57–63, Nov. 2022, doi: 10.31102/zeta.2022.7.2.57-63.
- [17] S. I. Pratiwi, T. Widiharih, and A. R. Hakim, "Analisis Klaster Metode Ward Dan Average Linkage Dengan Validasi Dunn Index Dan Koefisien Korelasi Cophenetic (Studi Kasus: Kecelakaan Lalu Lintas Berdasarkan Jenis Kendaraan Tiap Kabupaten/Kota di Jawa Tengah Tahun 2018)," *JURNAL GAUSSIAN*, vol. 8, pp. 486–495, 2019, doi: <https://doi.org/10.14710/j.gauss.8.4.486-495>.
- [18] R. N. Puspita, "Perbandingan Metode Analisis Cluster Hirarki Pada Data Margin Perdagangan Dan Pengangkutan (Mpp) Komoditas Strategis Di Indonesia," : *Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, vol. 3, no. <https://lebesgue.lppmbinabangsa.id/index.php/home/issue/view/8>, pp. 206–223, 2022, doi: 10.46306/lb.v3i1.
- [19] R. A. Farissa, R. Mayasari, and Y. Umidah, "Perbandingan Algoritma K-Means dan K-Medoids

Untuk Pengelompokan Data Obat dengan Silhouette Coefficient,” 2021. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>