

# HASIL CEK\_Analisis Persebaran UMKM Bidang Jasa Menggunakan Metode AHC Complete Linked

*by Lisna, Seftian, Utaminingsih, Anna*

---

**Submission date:** 11-Aug-2023 02:37PM (UTC+0700)

**Submission ID:** 2144337167

**File name:** aran\_UMKM\_Bidang\_Jasa\_Menggunakan\_Metode\_AHC\_Complete\_Linked.pdf (903.23K)

**Word count:** 3434

**Character count:** 19088

## Analisis Persebaran UMKM Bidang Jasa Menggunakan Metode AHC Complete Linkage

Lisna Zahrotun<sup>1</sup>, Seftian Hadi Nugroho<sup>2</sup>, Utaminingsih Linarti<sup>3</sup>, Anna Hendri Soleha Jones<sup>4</sup>

<sup>1,2,4</sup>Informatika, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

<sup>3</sup>Teknik Industri, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

E-mail: <sup>1</sup>lisna.zahrotun@tif.uad.ac.id, <sup>2</sup>seftian1800018269@webmail.uad.ac.id, <sup>3</sup>utaminingsih.linarti@ie.uad.ac.id, <sup>4</sup>annahendri@tif.uad.ac.id

### 11 Abstrak

Dinas Perindustrian Koperasi Usaha Kecil dan Menengah merupakan salah satu yang bertanggung jawab terhadap pertumbuhan dan kestabilan Usaha Mikro, Kecil dan Menengah (UMKM) yang berada di kota X. Banyaknya UMKM bidang jasa mengakibatkan perlunya penanganan khusus dalam menjaga pertumbuhan dan kestabilan dari UMKM tersebut. Oleh karena itu agar pihak Dinas dapat memberikan strategi dalam menjaga pertumbuhan dan kestabilan UMKM bidang jasa maka diperlukan analisis dan identifikasi persebaran UMKM tersebut. Salah satu teknik yang dapat digunakan dalam melakukan analisis persebaran UMKM adalah pengelompokan. Dimana dari hasil analisis pengelompokan UMKM bidang jasa ini dapat dijadikan sebagai referensi dalam pertumbuhan dan menjaga kestabilan UMKM bidang Jasa. Tahapan dalam penelitian ini adalah Load Data, Data Cleaning dan Data Selection, Data Transformation, Proses Pengelompokan Dengan Agglomerative Hierarchical Clustering (AHC) complete linkage, pengujian menggunakan Silhouette Coefficient, dan Representasi Pengeahuan. Hasil dari penelitian ini menunjukkan bahwa metode AHC complete linkage dapat dilakukan pada UMKM bidang jasa dengan hasil pengujian terbaik yaitu 0,729 dengan jumlah kelompok 2. Dari 2 kelompok yang dihasilkan, kelompok 1 merupakan kelompok UMKM bidang jasa yang perlu mendapatkan perhatian khusus dari Dinas Perindustrian dan Koperasi karena usia berdiri sudah lama namun omsetnya masih dibawah 10 juta.

**Keywords:** Agglomerative Hierarchical Clustering, Silhouette Coefficient

### 1.3 Pendahuluan

Usaha mikro memiliki peran penting dalam perekonomian yang ada di Indonesia, di mana usaha mikro memasok 99,8% lapangan kerja dan merupakan lebih dari 95% dari semua perusahaan di Indonesia [1]. Kerja sama dari industry, pemerintah dan perguruan tinggi berperan penting dalam proses penciptaan inovasi dan aliran pengetahuan di wilayahnya masing masing [2]. Dinas Perindustrian Koperasi Usaha Kecil dan Menengah di kota X merupakan salah satu Dinas yang mengelola pertumbuhan dan keadaan Usaha Mikro, Kecil dan Menengah (UMKM). Pada Dinas ini terdapat data UMKM yang berada di wilayahnya. Berdasarkan wawancara terhadap salah satu pihak Dinas mengemukakan bahwa belum ada pengolahan terhadap database yang telah terkumpul. Pihak Dinas juga mengemukakan bahwa dari database UMKM bidang jasa yang telah terkumpul juga bisa digunakan sebagai dasar pemilihan pelatihan yang tepat, namun belum ada pengolahan data karena data UMKM bidang jasa yang ada sering berubah seiring dengan bertambahnya UMKM. Selain itu banyaknya variabel dan juga data menyebabkan pihak Dinas kesulitan dalam melakukan pengecekan dan pemantauan terhadap UMKM. Agar pihak Dinas dapat melakukan strategi-strategi dalam meningkatkan UMKM di kota maka diperlukan analisis persebaran kondisi UMKM. Salah satu teknik yang dapat digunakan dalam pengelompokan adalah teknik

*clustering* dimana dengan adanya pengelompokan dapat dilakukan analisis persebaran UMKM pada kota X.

Beberapa penelitian terkait UMKM pernah dilakukan sebelumnya diantaranya analisis persebaran UMKM di kota Malang menggunakan K-Means [3], klasifikasi pengkarakteristikan iklan UMKM [4], pemasaran online untuk meningkatkan UMKM [5], dan penentuan strategi promosi UMKM menggunakan metode pengelompokan menggunakan K-Means [6]. Selain K-Means teknik pengelompokan yang dapat digunakan adalah *Agglomerative Hierarchical Clustering (AHC)*. Teknik AHC ini dapat digunakan untuk data dalam bentuk kategori, beberapa penerapan teknik AHC adalah pendeteksian komunitas [7], pemodelan wilayah banjir [8], pengelompokan jumlah penumpang [9]. Pengelompokan UMKM juga dilakukan di Negara India [10].

Salah satu teknik pengelompokan adalah *Agglomerative Hierarchical Clustering (AHC)*. Dimana metode ini cocok untuk data yang bersifat kategorikal. Salah satu penelitian yang menggunakan AHC adalah pengelompokan data mahasiswa yang mendapatkan nilai akurasi 0.8 [11]. Dilihat dari peluang usaha UMKM bidang jasa yang sangat tinggi, maka dalam penelitian ini dilakukan analisis persebaran UMKM bidang jasa menggunakan metode AHC *single linked*. Penelitian ini diharapkan dapat menerapkan keunggulan dari metode dan memenuhi kebutuhan akan pemrosesan data.

## 2. Metode

Tahapan yang dilakukan dalam penelitian ini ditunjukkan pada Gambar 1.

### 2.1. Load Data

Pada proses ini dilakukan *load data* atau *input dataset* berbentuk file excel dengan format xlsx ke dalam aplikasi agar dapat diproses.

### 2.2. Data Cleaning & Selection

Pada tahap ini, dilakukan pembersihan data untuk dijadikan data penelitian terbebas dari data yang tidak relevan dan *noise*.

### 2.3. Data Transformation

Pada tahap ini adalah mengubah dataset yang berupa string menjadi integer berniali 0 dan 1 agar datanya dapat digunakan atau diproses oleh sistem.

### 2.4. AHC Complete Linkage

Pada tahap ini adalah memasukkan data hasil transformasi ke dalam metode *AHC*. Dalam statistik, pengelompokan hirarki (*hierarchical clustering*) adalah metode analisis kelompok yang berusaha untuk membangun sebuah hierarki kelompok [12]. *Agglomerative Hierarchical Clustering (AHC)* merupakan metode pengelompokan hierarki dengan pendekatan bawah – atas (*bottom – up*) [13].

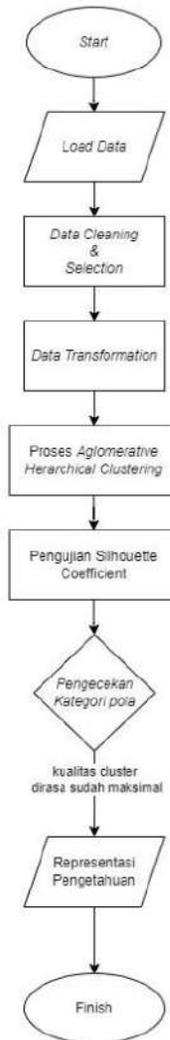
Berikut merupakan langkah-langkah dalam pengelompokan hierarki:

- a) Menghitung matriks kedekatan berdasarkan jenis jarak yang digunakan, untuk penelitian ini menggunakan jarak *euclidian*,

$$\|x - y\|^2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Keterangan:

$x_i$  = value data 1;  $y_i$  = value data 2



Gambar 1. Tahap Penelitian

- b) Mengulangi Langkah 3 sampai 4, hingga hanya 1 kelompok yang tersisa,
- c) Menggabungkan dua kelompok yang memiliki jarak terdekat pada matriks jarak.
- d) Memperbarui matriks kedekatan untuk merefleksikan kedekatan di antara kelompok baru dan kelompok asli yang sudah digabung menggunakan *complete linkage*.

$$d_{UV} = \max\{d_{UV}\}, d_{UV} \in D$$

$d_{uv}$  = jarak antara tetangga terjauh dari kelompok (U dan V);

D = matriks kedekatan.

## 2.4 Silhouette Coefficient

*Cluster validation* adalah cara untuk mengetahui seberapa baik hasil yang diperoleh dari proses *clustering* [14]. Pada penelitian ini menggunakan metode *silhouette coefficient*. *Silhouette coefficient* merupakan validasi yang menggabungkan dua unsur yaitu *cohesion* dan *separation*. Nilai dari *silhouette coefficient* ini mempunyai range dari -1 sampai 1. Nilai 1 dikategorikan dalam *cluster* baik atau *cluster* yang terbentuk merupakan hasil *cluster* yang baik, sedangkan nilai -1 menunjukkan bahwa hasil *cluster* yang terbentuk adalah kurang baik, semakin mendekati 0 menunjukkan semakin sedikit dokumen yang benar pengelompokannya.

## 2.6. Representasi Pengetahuan

Setelah dilakukan perhitungan *silhouette* kemudian dilakukan evaluasi untuk penentuan klaster terbaik.

## 3. Hasil dan Pembahasan

### 3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data UMKM bidang jasa di kota X dengan jumlah 399 data. *Dataset* berupa file excel dengan 40 variabel.

### 3.2. Proses *Cleaning* dan *Selection*

Data set yang diperoleh akan dilakukan proses preproccesing yaitu *cleaning* data, dalam proses ini data akan dicek satu persatu untuk melihat apakah ada data *outlier* atau *noise*. Dari 399 data UMKM setelah dilakukan proses *cleaning* tidak ditemukan adanya *noise* sehingga jumlah data tetap 399 data UMKM. Langkah selanjutnya adalah *selection* atau seleksi variable. Dalam proses ini 40 variabel di seleksi menjadi 12 variabel. 12 Variabel hasil proses *selection* ditunjukkan pada Tabel 1.

Tabel 1. Tabel Hasil Seleksi

Hasil Seleksi	Hasil Pelabelan
Pendidikan Terakhir	pendidikan
Tanggal Pendirian Usaha	tanggal_pendirian_usaha
Kegiatan Usaha	kegiatan_usaha
Tujuan Pemasaran	tujuan_pemasaran
Status Kepemilikan Tanah/Bangunan	kepemilikan_tanah
Sarana Media Elektronik	sarana_media_elektronik
Modal Bantuan Pemerintah	modal_bantuan_pemerintah
Pinjaman Kredit Usaha Rakyat	pinjaman
Omset per-Tahun	omset_pertahun
Kepemilikan Asuransi Kesehatan	asuransi
Tenaga Kerja Laki-laki	tenaga_kerja_laki
Tenaga Kerja Perempuan	tenaga_kerja_perempuan

### 3.3. Proses *Data Transformation*

Tahap selanjutnya selanjutnya setelah yaitu data transformation. Pada tahap ini variable yang sudah terpilih akan dilakukan pengkategorian menggunakan proses One Hot Encoding. Hasil Variabel H Encoding ditunjukkan pada Tabel 2. Sedangkan hasil data transformation ditunjukkan pada Tabel 3.

Tabel 2. Hasil Atribut Baru

No.	Variabel	Hasil Variabel dari One Hot Encoding
1	Pendidikan Terakhir	-, 0, D1, D2, D3, S1, S2, S3, SD, SMA, SMK, SMP
2	Kegiatan Usaha	Penjualan, Produksi
3	Tujuan Pemasaran	Dalam wilayah DIY,

No.	Variabel	Hasil Variabel dari One Hot Encoding
		Dalam wilayah Kota Yogyakarta,
		Dalam wilayah Pulau Jawa,
		Dalam wilayah luar Pulau Jawa (tersebar),
		Luar Negeri (ekspor),
		Luar Negeri (jenis pengiriman biasa),
4	Status Kepemilikan Tanah/Bangunan	Lainnya, Magersari (adat), Milik sendiri, Sewa
5	Sarana Media Elektronik	-, Facebook, Gojek, Grab, Instagram, Shopee, Tokopedia, Twitter, WhatsApp, Lainnya
6	Modal Bantuan Pemerintah	-, Pemda DIY, Pemerintah Pusat, Pemkot Yogyakarta
7	Pinjaman Kredit Usaha Rakyat	-, Bank, Koperasi, Lainnya, Pemerintah
8	Omset per-Tahun	Kurang dari 10 juta,
		10 juta s/d 25 juta,
		25 juta s/d 40 juta,
		40 juta s/d 55 juta,
		55 juta s/d 70 juta,
		70 juta s/d 85 juta,
		85 juta s/d 100 juta,
		100 juta s/d 120 juta,
		120 juta s/d 150 juta,
		Lebih dari 150 juta
9	Kepemilikan Asuransi Kesehatan	-, Asuransi Swasta, BPJS

**Tabel 3.** Tabel Hasil Transformasi

-	0	D1	D2	D3	S1	S2	.....	tenaga_kerja_laki	tenaga_kerja_perempuan
0	0	0	0	1	0	0	.....	2	2
1	0	0	0	0	0	0	.....	1	0
0	0	0	0	0	0	0	.....	0	0
0	0	0	0	0	1	0	.....	1	0
0	0	0	0	0	0	1	.....	5	3
0	0	0	0	0	0	0	.....	2	0
0	0	0	0	0	0	0	.....	0	0
0	0	0	0	0	1	0	.....	0	0
0	0	0	0	0	1	0	.....	4	0
0	0	0	0	0	0	0	.....	0	0
0	0	0	0	0	0	0	.....	0	1
0	0	0	0	0	0	0	.....	0	1
0	0	0	0	0	0	0	.....	0	0
0	0	0	0	0	1	0	.....	0	0
0	0	0	0	0	1	0	.....	1	0
0	1	0	0	0	0	0	.....	0	0
1	0	0	0	0	0	0	.....	3	2
0	0	0	0	0	0	0	.....	0	0
0	0	0	0	0	1	0	.....	0	1
0	0	0	0	0	1	0	.....	0	1
1	0	0	0	0	0	0	.....	0	0
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
0	0	0	0	0	0	0	.....	0	0

### 3.4. Proses AHC *Completed Linked*

Tahap selanjutnya dengan menghitung AHC menggunakan *Completed linked*. Dalam penghitungan ini akan dilakukan penghitungan jarak antar data menggunakan *Euclidian Distance* untuk menghitung jarak terdekat pada setiap umkm. Dikatakan jarak terdekat jika data memiliki jarak mendekati 0 atau merupakan jarak terkecil.

$$\|x - y\|^2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Keterangan:

$x_i$  = value data 1

$y_i$  = value data 2

3 Menghitung jarak *euclidean* pada setiap data umkm. Seperti pada 2 sampel berikut:

a) Menghitung jarak umkm 1 dan 2

$$\begin{aligned} & \sqrt{\begin{pmatrix} (0-1)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + \\ (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (2-3)^2 + (0-1)^2 + (1-0)^2 + (1-1)^2 + \\ (1-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-1)^2 + (0-0)^2 + (1-0)^2 + \\ (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-1)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + \\ (0-0)^2 + (0-0)^2 + (1-1)^2 + (1-1)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (1-1)^2 + \\ (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + \\ (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (1-1)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + \\ (0-1)^2 + (2-1)^2 + (2-0)^2 \end{pmatrix}} \\ & = \sqrt{\begin{pmatrix} 1+0+0+0+1+0+0+0+0+ \\ 0+0+0+1+1+1+0+1+0+ \\ 0+0+0+1+0+1+0+0+0+ \\ 1+0+0+0+0+0+0+0+0+ \\ 0+0+0+0+0+0+0+0+0+ \\ 0+0+0+0+0+0+0+0+0+ \\ 1+0+1+1+4 \end{pmatrix}} \\ & = \sqrt{16} \\ & = 4 \end{aligned}$$

b) Menghitung jarak UMKM 1 dan 3

$$\begin{aligned} \|u - v\|_2 &= \sqrt{\sum_i (u_i - v_i)^2} \\ &= 4,69041576 \end{aligned}$$

*Euclidean distance* dihitung untuk seluruh data hasil transformasi sehingga menghasilkan matrik kedekatan, hasil matrik kedekatan ditunjukkan pada Tabel 4.

12 **Tabel 4.** Matrik Kedekatan

	UMKM 1	UMKM 2			UMKM 6	UMKM 399
UMKM 1	0					
UMKM 2	4	0				
...	...	...				
...	...	...				
...	...	...				
UMKM 6	3,605551275	3,605551175			0	
UMKM 399	7,549834435	6,08276253			7,483314774	0

Selanjutnya adalah menentukan jarak terkecil. Dari matrik kedekatan terdapat umkm 6, 1, dan 2 yang memiliki jarak terkecil maka dilakukan penggabungan dan menghasilkan matrik baru pada Tabel 5.

**Tabel 5.** Hasil Penggabungan Jarak Terdekat

	UMKM 1 UMKM 2 UMKM 6				UMKM 399
UMKM 1 UMKM 2 UMKM 6	0				
...		0			
...		...	0		
...		...	...	0	
UMKM 399		4,582575695	5,830951895	7,681145748	0

Pada data tabel 5 terdapat kekosongan jarak, untuk mengisi kekosongan jarak tersebut diisi menggunakan *complete linkage*. Yaitu dengan mengecek setiap nilai jarak pada setiap data. Untuk mengisi data pada jarak UMKM(1,2,6) dengan UMKM(3), dilakukan dengan perbandingan jarak dari matriks kedekatan UMKM 1 dan 3, UMKM 2 dan 3, UMKM 6 dan 3. Dari ketiga data didapatkan jarak dari matriks diantaranya 4,69041576; 3,741657387; dan 5. Lalu dari ketiga data itu dipilih jarak yang paling jauh atau besar sehingga didapatkan 5 sebagai jarak antara UMKM(1,2,6) dengan UMKM(3). Lalu dilanjutkan dengan menentukan jarak terkecil pada matriks baru dan menggabungkan lagi, kemudian mengisi kekosongan jarak dengan *complete linkage*. Tahap ini dilakukan terus menerus hingga tersisa 2 kelompok. Perulangan ini menghasilkan iterasi pada Tabel 6 hanya menggunakan 7 data UMKM.

**Tabel 6.** Hasil Pengelompokkan

Tahap	Jumlah hasil cluster	Hasil Anggota Cluster	Jarak Terkecil
1	5	UMKM 1,	3,605551275
		UMKM 2,	
		UMKM 6	
2	4	UMKM 3,	4,582575695
		UMKM 7	
3	3	UMKM 1,	5
		UMKM 2,	
		UMKM 6,	
		UMKM 3,	
4	2	UMKM 7	5,916079783
		UMKM 4	
		UMKM 5	

Selanjutnya adalah menguji hasil klaster menggunakan *silhouette coefficient*. Pada pengujian ini akan dilakukan dengan menguji *silhouette* 2 klaster.

- a) Menghitung rata – rata jarak setiap umkm dengan umkm lain yang masih dalam satu klaster dengan persamaan. Hasil menghitung rata-rata ditunjukkan pada Tabel 7.

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in Cd(i, j)} \quad (3)$$

7eterangan :

- a(i) = Perbedaan rata – rata objek (i) ke semua objek lain pada A.
- d(I,j) = Jarak antara i ke j
- A = Cluster

**Tabel 7.** Hasil Perhitungan Rata-Rata Jarak dengan satu kluster

Data	a(i)	Cluster
UMKM 1	4,961450368	0
UMKM 2	4,357492798	0
UMKM 3	4,503662244	0
UMKM 4	5,916079783	1
UMKM 5	5,916079783	1
UMKM 6	4,923604331	0
UMKM 7	6,424621859	0

- b) Lalu dilanjutkan dengan menghitung rata – rata jarak umkm dengan umkm lain yang berbeda kluster. Hasil penghitungan jarak dengan cluster ditunjukkan pada Tabel 8.

$$d(i, C) = \frac{1}{|A|} \sum_{j \in C} d(i, j) \quad (4)$$

$$b(i) = \min_{C \neq A} d(i, C) \quad (5)$$

Keterangan:

d(i,C) = perbedaan rata - rata objek (i) ke semua objek lain pada C.

C = Cluster lain selain cluster A atau cluster C tidak sama dengan A.

b(i) = Rata – rata jarak terkecil umkm dengan setiap umkm pada cluster lain

**Tabel 8.** Hasil Perhitungan Jarak dengan Cluster lain

Data	d(I, 0)	d(I, 1)	b(i)	Cluster
UMKM 1	-	10,41165512	10,41165512	0
UMKM 2	-	9,924428901	9,924428901	0
UMKM 3	-	8,368265443	8,368265443	0
UMKM 4	8,971018069	-	8,971018069	1
UMKM 5	9,455268467	-	9,455268467	1
UMKM 6	-	7,885164805	7,885164805	0
UMKM 7	-	6,756048822	6,756048822	0

- c) Lalu dilanjutkan dengan menghitung nilai silhouette. Hasil penghitungan silhouette ditunjukkan pada Tabel 9.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (6)$$

Keterangan:

s(i) = Silhouette Coefficient

a(i) = Perbedaan rata – rata objek (i) ke semua objek lain pada A

b(i) = Rata – rata jarak terkecil umkm dengan setiap umkm pada cluster lain

**Tabel 9.** Hasil Perhitungan Silhouette

Data	d(I, 0)	d(I, 1)	b(i)	S(i)	Cluster
UMKM 1	-	10,41165512	10,41165512	0,523471503	0
UMKM 2	-	9,924428901	9,924428901	0,56093264	0
UMKM 3	-	8,368265443	8,368265443	0,461816521	0
UMKM 4	8,971018069	-	8,971018069	0,34053418	1
UMKM 5	9,455268467	-	9,455268467	0,37430864	1
UMKM 6	-	7,885164805	7,885164805	0,375586376	0
UMKM 7	-	6,756048822	6,756048822	0,04905633	0
Rata – rata s(i)				0,383672313	

Lalu dilanjutkan dengan melakukan evaluasi pola. Berdasarkan dari percobaan menggunakan sampel data didapatkan nilai *silhouette* sebesar 0,383672313. Yang artinya nilainya mendekati 0 dan 2 *cluster* bukan merupakan *cluster* yang baik.

### 3.5. Pengujian Hasil dari Sistem

Pada penelitian ini menggunakan pengujian *Silhouette coefficient*. Pada system telah dilakukan 9 kali pengujian terhadap pengujian 10-klaster Tabel 10 merupakan hasil *silhouette coefficient* setiap klaster. Hasil pengujian menunjukkan bahwa cluster 2 memiliki nilai kaurasi tertinggi yaitu 0,729.

**Tabel 10.** Hasil Pengujian

Jumlah Cluster	Hasil Silhouette Coefficient
2	0,729
3	0,709
4	0,488
5	0,473
6	0,471
7	0,423
8	0,423
9	0,420
10	0,264

### 3.6. Representasi pengetahuan

Dari hasil pengujian dari sistem didapatkan 2 kelompok sebaran UMKM bidang jasa yang memiliki nilai *silhouette* 0,729. Hasil analisis persebaran UMKM bidang jasa di Kota X pada 2 kelompok tersebut adalah:

- a) Kelompok 1  
 Pada Kelompok 1 ini terdiri dari 16 data UMKM yang memiliki rata-rata lama berdiri 41 tahun, Pendidikan dari pemilik UMKM Sekolah Menengah ke bawah, jumlah karyawan rata rata 1 orang dan omset pertahun rata rata di bawah 10 juta.
- b) Kemompok 2  
 Pada kelompok 2 terdiri dari 383 data UMKM yang memiliki rata-rata lama berdiri 7 tahun, Pendidikan Pemilik UMKM Sekolah Menegah ke atas, jumlah karyawan rata-rata 1 dan ada yang tidak memiliki karyawan serta rata-rata omset pertahunnya di atas 10 juta.

Dari hasil analisis persebaran UMKM 2 kelompok tersebut dapat dijadikan acuan bagi Dinas Perindustrian Koperasi Usah Kecil dan Menengah di kota X dalam meningkatkan UMKM -UMKM yang belum maju yang terlihat pada UMKM di kelompok 1. Salah satu usulan yang bisa dilakukan disini adalah peningkatan pengetahuan untuk menignkatkan omset pertahun dengan memberikan pelatihan-pelatihan kepada UMKM pada kelompok 1.

## 4. Kesimpulan

Kesimpulan dalam penelitian ini adalah bahwa penerapan algoritma AHC *complete linkage* dapat dilakukan pada UMKM bidang jasa menghasilkan 2 cluster yang direkomendasikan dengan nilai akurasi 0,729, dimana kelor<sup>13</sup>ok 1 merupakan kelompok UMKM yang perlu mendapatkan perhatian khusus dari Dinas Perindustrian Koperasi Usah Kecil dan Menengah kota X karena usia berdiri sudah lama namun omsetnya masih dibawah 10 juta.

## Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Lembaga Penelitian dan Pengabdian Masyarakat Universitas Ahmad Dahlan yang telah memberikan hibah pada penelitian ini melalui skem Penelitian Terapan tahun 2022.

## Daftar Pustaka

- [1] I. R. Maksum, A. Y. Sri Rahayu, and D. Kusumawardhani, "A social enterprise approach to empowering micro, small and medium enterprises (SMEs) in Indonesia," *J. Open Innov. Technol. Mark. Complex.*, vol. 6, no. 3, 2020, doi: 10.3390/JOITMC6030050.
- [2] A. H. Jankowiak, "Industrial Clusters As an Important Factor for Development of Micro, Small and Medium-Sized Enterprises (Msmes)," *12Th Int. Days Stat. Econ.*, no. November, pp. 708–718, 2018.
- [3] P. Puntoriza and C. Fibriani, "Analisis Persebaran UMKM Kota Malang Menggunakan Cluster K-means," *JOINS (Journal Inf. Syst.)*, vol. 5, no. 1, pp. 86–94, 2020, doi: 10.33633/joins.v5i1.3469.
- [4] M. T. Satya, N. K. Rahayu, and A. Fidayan, "Pengklasifikasian Karakteristik Iklan UMKM Menggunakan Analisis Cluster," *Jesya (Jurnal Ekon. Ekon. Syariah)*, vol. 3, no. 2, pp. 218–231, 2020, doi: 10.36778/jesya.v3i2.169.
- [5] D. Solihin, A. Ahyani, K. Karolina, L. Pricilla, and I. S. Octaviani, "Pelatihan Pemasaran Online Berbasis Digital Untuk Meningkatkan Penjualan Bisnis Online Pada Umkm Di Desa Cicalengka Kecamatan Pagedangan Kabupaten Tangerang," *Dedik. Pkm*, vol. 2, no. 3, pp. 307–311, 2021, doi: 10.32493/dedikasipkm.v2i3.10726.
- [6] D. Astuti, A. R. Iskandar, and A. Febrianti, "Penentuan Strategi Promosi Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 1, no. 2, pp. 60–72, 2019, doi: 10.20895/inista.v1i2.71.
- [7] I. W. Rahayu *et al.*, "Hierarchical Clustering Untuk Deteksi Komunitas Pada Media Sosial Facebook Analysis and Implementation of Agglomerative Hierarchical Clustering Algorithm for Community Detection in Social Media Facebook," *e-Proceeding Eng.*, vol. 5, no. 1, pp. 1460–1468, 2018.
- [8] R. O. Pratikto and N. Damastuti, "Klasterisasi Menggunakan Agglomerative Hierarchical Clustering Untuk Memodelkan Wilayah Banjir," *JOINTECS (Journal Inf. Technol. Comput. Sci.)*, vol. 6, no. 1, p. 13, 2021, doi: 10.31328/jointecs.v6i1.1473.
- [9] L. Zahrotun, "Analisis Pengelompokan Jumlah Penumpang Bus Trans Jogja Menggunakan metode Clustering K-Means dan Agglomerative Hierarchical Clustering (AHC)," *J. Inform.*, vol. 9, no. 1, pp. 1039–1047, 2015.
- [10] D. Rawat, R. K. Mittal, and V. S. Aggarwal, "Cluster Development Approach in India: an Antidote for Micro, Small, and Medium Enterprises," *Indian J. Commer. Manag. Stud.*, vol. VIII, no. 2, pp. 19–29, 2017, doi: 10.18843/ijcms/v8i2/03.
- [11] B. Harli Trimulya Suandi As and L. Zahrotun, "Penerapan Data Mining Dalam Mengelompokkan Data Riwayat Akademik Sebelum Kuliah dan data Kelulusan Mahasiswa Menggunakan Metode Agglomerative Hierarchical Clustering(Implementation Of Data Mining In Grouping Academic History Data Before Students And Stud)," *J. Teknol. Informasi, Komput. dan Apl.*, vol. 3, no. 1, pp. 62–71, 2021, [Online]. Available: <http://jtika.if.unram.ac.id/index.php/JTIKA/>
- [12] K. E. T. Luthfi, *Algoritma Data Mining*, 1st ed. Yogyakarta: C.V Andi Offset, 2019.
- [13] E. Prasetyo, *DATA MINING-Konsep dan Aplikasi Menggunakan MATLAB*, 1st

- ed. Yogyakarta: Andi Offset, 2012.
- [14] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data An Introduction to Cluster Analysis*. Canada: John Wiley & Sons, Inc, Hoboken, New Jersey.

# HASIL CEK\_Analisis Persebaran UMKM Bidang Jasa Menggunakan Metode AHC Complete Linked

## ORIGINALITY REPORT

18%

SIMILARITY INDEX

16%

INTERNET SOURCES

4%

PUBLICATIONS

4%

STUDENT PAPERS

## PRIMARY SOURCES

1	<a href="http://eprints.uad.ac.id">eprints.uad.ac.id</a> Internet Source	4%
2	Corie Mei Hellyana, Nuzul Imam Fadlilah. "Implementasi Algoritma K-Means Untuk Clusterisasi Jumlah Kunjungan Tamu Asing Pada Hotel Berbintang Menurut Provinsi di Indonesia", EVOLUSI : Jurnal Sains dan Manajemen, 2023 Publication	2%
3	<a href="http://123dok.com">123dok.com</a> Internet Source	2%
4	<a href="http://id.scribd.com">id.scribd.com</a> Internet Source	2%
5	<a href="http://docplayer.info">docplayer.info</a> Internet Source	2%
6	(3-30-03) <a href="http://202.158.39.106/modamo/contents/berita/200189">http://202.158.39.106/modamo/contents/berita/200189</a> Internet Source	1%
7	<a href="http://jtika.if.unram.ac.id">jtika.if.unram.ac.id</a>	

---

Internet Source

1 %

---

8 journal.trunojoyo.ac.id  
Internet Source

1 %

---

9 zombiedoc.com  
Internet Source

1 %

---

10 juti.if.its.ac.id  
Internet Source

1 %

---

11 www.neliti.com  
Internet Source

1 %

---

12 repository.its.ac.id  
Internet Source

1 %

---

13 doku.pub  
Internet Source

1 %

---

14 ejournal.uin-suska.ac.id  
Internet Source

1 %

---

Exclude quotes  On

Exclude matches  < 1%

Exclude bibliography  On

## Analisis Persebaran UMKM Bidang Jasa Menggunakan Metode AHC Complete Linkage

Lisna Zahrotun<sup>1</sup>, Seftian Hadi Nugroho<sup>2</sup>, Utaminingsih Linarti<sup>3</sup>, Anna Hendri Soleliza Jones<sup>4</sup>

<sup>1,2,4</sup>Informatika, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

<sup>3</sup>Tenik Industri, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

E-mail: <sup>1</sup>lisna.zahrotun@tif.uad.ac.id, <sup>2</sup>seftian1800018269@webmail.uad.ac.id,

<sup>3</sup>utaminingsih.linarti@ie.uad.ac.id, <sup>4</sup>annahendri@tif.uad.ac.id

### Abstrak

Dinas Perindustrian Koperasi Usaha Kecil dan Menengah merupakan salah satu yang bertanggung jawab terhadap pertumbuhan dan kestabilan Usaha Mikro, Kecil dan Menengah (UMKM) yang berada di kota X. Banyaknya UMKM bidang jasa mengakibatkan perlunya penanganan khusus dalam menjaga pertumbuhan dan kestabilan dari UMKM tersebut. Oleh karena itu agar pihak Dinas dapat memberikan strategi dalam menjaga pertumbuhan dan kestabilan UMKM bidang jasa maka diperlukan analisis dan identifikasi persebaran UMKM tersebut. Salah satu teknik yang dapat digunakan dalam melakukan analisis persebaran UMKM adalah pengelompokan. Dimana dari hasil analisis pengelompokan UMKM bidang jasa ini dapat dijadikan sebagai referensi dalam pertumbuhan dan menjaga kestabilan UMKM bidang Jasa. Tahapan dalam penelitian ini adalah Load Data, Data Cleaning dan Data Selection, Data Transformation, Proses Pengelompokan Dengan Agglomerative Hierarchical Clustering (AHC) complete linkage, pengujian menggunakan Silhouette Coefficient, dan Representasi Pengetahuan. Hasil dari penelitian ini menunjukkan bahwa metode AHC complete linkage dapat dilakukan pada UMKM bidang jasa dengan hasil pengujian terbaik yaitu 0,729 dengan jumlah kelompok 2. Dari 2 kelompok yang dihasilkan, kelompok 1 merupakan kelompok UMKM bidang jasa yang perlu mendapatkan perhatian khusus dari Dinas Perindustrian dan Koperasi karena usia berdiri sudah lama namun omsetnya masih dibawah 10 juta.

**Keywords:** Agglomerative Hierarchical Clustering, Silhouette Coefficient

### 1. Pendahuluan

Usaha mikro memiliki peran penting dalam perekonomian yang ada di Indonesia, di mana usaha mikro memasok 99,8% lapangan kerja dan merupakan lebih dari 95% dari semua perusahaan di Indonesia [1]. Kerja sama dari industry, pemerintah dan perguruan tinggi berperan penting dalam proses penciptaan inovasi dan aliran pengetahuan di wilayahnya masing masing [2]. Dinas Perindustrian Koperasi Usaha Kecil dan Menengah di kota X merupakan salah satu Dinas yang mengelola pertumbuhan dan keadaan Usaha Mikro, Kecil dan Menengah (UMKM). Pada Dinas ini terdapat data UMKM yang berada di wilayahnya. Berdasarkan wawancara terhadap salah satu pihak Dinas mengemukakan bahwa belum ada pengolahan terhadap database yang telah terkumpul. Pihak Dinas juga mengemukakan bahwa dari database UMKM bidang jasa yang telah terkumpul juga bisa digunakan sebagai dasar pemilihan pelatihan yang tepat, namun belum ada pengolahan data karena data UMKM bidang jasa yang ada sering berubah seiring dengan bertambahnya UMKM. Selain itu banyaknya variabel dan juga data menyebabkan pihak Dinas kesulitan dalam melakukan pengecekan dan pemantauan terhadap UMKM. Agar pihak Dinas dapat melakukan strategi-strategi dalam meningkatkan UMKM di kota maka diperlukan analisis persebaran kondisi UMKM. Salah satu teknik yang dapat digunakan dalam pengelompokan adalah teknik

*clustering* dimana dengan adanya pengelompokan dapat dilakukan analisis persebaran UMKM pada kota X.

Beberapa penelitian terkait UMKM pernah dilakukan sebelumnya diantaranya analisis persebaran UMKM di kota Malang menggunakan K-Means [3], klasifikasi pengkarakteristikan iklan UMKM [4], pemasaran online untuk meningkatkan UMKM [5], dan penentuan strategi promosi UMKM menggunakan metode pengelompokan menggunakan K-Means [6]. Selain K-Means teknik pengelompokan yang dapat digunakan adalah *Agglomerative Hierarchical Clustering (AHC)*. Teknik AHC ini dapat digunakan untuk data dalam bentuk kategori, beberapa penerapan teknik AHC adalah pendeteksian komunitas [7], pemodelan wilayah banjir [8], pengelompokan jumlah penumpang [9]. Pengelompokan UMKM juga dilakukan di Negara India [10].

Salah satu teknik pengelompokan adalah *Agglomerative Hierarchical Clustering (AHC)*. Dimana metode ini cocok untuk data yang bersifat kategorikal. Salah satu penelitian yang menggunakan AHC adalah pengelompokan data mahasiswa yang mendapatkan nilai akurasi 0.8 [11]. Dilihat dari peluang usaha UMKM bidang jasa yang sangat tinggi, maka dalam penelitian ini dilakukan analisis persebaran UMKM bidang jasa menggunakan metode AHC *single linked*. Penelitian ini diharapkan dapat menerapkan keunggulan dari metode dan memenuhi kebutuhan akan pemrosesan data.

## 2. Metode

Tahapan yang dilakukan dalam penelitian ini ditunjukkan pada Gambar 1.

### 2.1. Load Data

Pada proses ini dilakukan *load data* atau *input dataset* berbentuk file excel dengan format xlsx ke dalam aplikasi agar dapat diproses.

### 2.2. Data Cleaning & Selection

Pada tahap ini, dilakukan pembersihan data untuk dijadikan data penelitian terbebas dari data yang tidak relevan dan *noise*.

### 2.3. Data Transformation

Pada tahap ini adalah mengubah dataset yang berupa string menjadi integer berniali 0 dan 1 agar datanya dapat digunakan atau diproses oleh sistem.

### 2.4. AHC Complete Linkage

Pada tahap ini adalah memasukkan data hasil transformasi ke dalam metode AHC. Dalam statistik, pengelompokan hirarki (*hierarchical clustering*) adalah metode analisis kelompok yang berusaha untuk membangun sebuah hierarki kelompok [12]. *Agglomerative Hierarchical Clustering (AHC)* merupakan metode pengelompokan hierarki dengan pendekatan bawah – atas (*bottom – up*) [13].

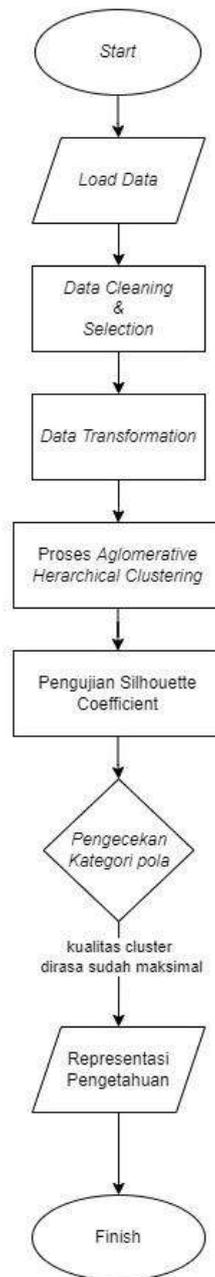
Berikut merupakan langkah-langkah dalam pengelompokan hierarki:

- a) Menghitung matriks kedekatan berdasarkan jenis jarak yang digunakan, untuk penelitian ini menggunakan jarak *euclidian*,

$$||x - y||^2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Keterangan:

$x_i$  = value data 1;       $y_i$  = value data 2



**Gambar 1.** Tahap Penelitian

- b) Mengulangi Langkah 3 sampai 4, hingga hanya 1 kelompok yang tersisa,
- c) Menggabungkan dua kelompok yang memiliki jarak terdekat pada matriks jarak.
- d) Memperbarui matriks kedekatan untuk merefleksikan kedekatan di antara kelompok baru dan kelompok asli yang sudah digabung menggunakan *complete linkage*.

$$d_{UV} = \max\{d_{UV}\}, d_{UV} \in D$$

$d_{uv}$  = jarak antara tetangga terjauh dari kelompok (U dan V);

D = matriks kedekatan.

### 2.5. Silhouette Coefficient

*Cluster validation* adalah cara untuk mengetahui seberapa baik hasil yang diperoleh dari proses *clustering* [14]. Pada penelitian ini menggunakan metode *silhouette coefficient*. *Silhouette coefficient* merupakan validasi yang menggabungkan dua unsur yaitu *cohesion* dan *separation*. Nilai dari *silhouette coefficient* ini mempunyai *range* dari -1 sampai 1. Nilai 1 dikategorikan dalam *cluster* baik, atau *cluster* yang terbentuk merupakan hasil *cluster* yang baik, sedangkan nilai -1 menunjukkan bahwa hasil *cluster* yang terbentuk adalah kurang baik, semakin mendekati 0 menunjukkan semakin sedikit dokumen yang benar pengelompokannya.

### 2.6. Representasi Pengetahuan

Setelah dilakukan perhitungan *silhouette* kemudian dilakukan evaluasi untuk penentuan klaster terbaik.

## 3. Hasil dan Pembahasan

### 3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data UMKM bidang jasa di kota X dengan jumlah 399 data. *Dataset* berupa file excel dengan 40 variabel.

### 3.2. Proses Cleaning dan Selection

Data set yang diperoleh akan dilakukan proses *preprocessing* yaitu *cleaning* data, dalam proses ini data akan dicek satu persatu untuk melihat apakah ada data *outlier* atau *noise*. Dari 399 data UMKM setelah dilakukan proses *cleaning* tidak ditemukan adanya *noise* sehingga jumlah data tetap 399 data UMKM. Langkah selanjutnya adalah *selection* atau seleksi *variable*. Dalam proses ini 40 variabel di seleksi menjadi 12 variabel. 12 Variabel hasil proses *selection* ditunjukkan pada Tabel 1.

**Tabel 1.** Tabel Hasil Seleksi

Hasil Seleksi	Hasil Pelabelan
Pendidikan Terakhir	pendidikan
Tanggal Pendirian Usaha	tanggal_pendirian_usaha
Kegiatan Usaha	kegiatan_usaha
Tujuan Pemasaran	tujuan_pemasaran
Status Kepemilikan Tanah/Bangunan	kepemilikan_tanah
Sarana Media Elektronik	sarana_media_elektronik
Modal Bantuan Pemerintah	modal_bantuan_pemerintah
Pinjaman Kredit Usaha Rakyat	pinjaman
Omset per-Tahun	omset_pertahun
Kepemilikan Asuransi Kesehatan	asuransi
Tenaga Kerja Laki-laki	tenaga_kerja_laki
Tenaga Kerja Perempuan	tenaga_kerja_perempuan

### 3.3. Proses Data Transformation

Tahap selanjutnya selanjutnya setelah yaitu data transformation. Pada tahap ini *variable* yang sudah terpilih akan dilakukan pengkategorian menggunakan proses One Hot Encoding. Hasil Variabel Hot Encoding ditunjukkan pada Tabel 2. Sedangkan hasil data transformation ditunjukkan pada Tabel 3.

**Tabel 2.** Hasil Atribut Baru

No.	Variabel	Hasil Variabel dari One Hot Encoding
1	Pendidikan Terakhir	-, 0, D1, D2, D3, S1, S2, S3, SD, SMA, SMK, SMP
2	Kegiatan Usaha	Penjualan, Produksi
3	Tujuan Pemasaran	Dalam wilayah DIY,

No.	Variabel	Hasil Variabel dari One Hot Encoding
		Dalam wilayah Kota Yogyakarta,
		Dalam wilayah Pulau Jawa,
		Dalam wilayah luar Pulau Jawa (tersebar),
		Luar Negeri (ekspor),
		Luar Negeri (jenis pengiriman biasa),
4	Status Kepemilikan Tanah/Bangunan	Lainnya, Magersari (adaf), Milik sendiri, Sewa
5	Sarana Media Elektronik	-, Facebook, Gojek, Grab, Instagram, Shopee, Tokopedia, Twitter, WhatsApp, Lainnya
6	Modal Bantuan Pemerintah	-, Pemda DIY, Pemerintah Pusat, Pemkot Yogyakarta
7	Pinjaman Kredit Usaha Rakyat	-, Bank, Koperasi, Lainnya, Pemerintah
8	Omset per-Tahun	Kurang dari 10 juta,
		10 juta s/d 25 juta,
		25 juta s/d 40 juta,
		40 juta s/d 55 juta,
		55 juta s/d 70 juta,
		70 juta s/d 85 juta,
		85 juta s/d 100 juta,
		100 juta s/d 120 juta,
120 juta s/d 150 juta,		
Lebih dari 150 juta		
9	Kepemilikan Asuransi Kesehatan	-, Asuransi Swasta, BPJS

**Tabel 3.** Tabel Hasil Transformasi

-	0	D1	D2	D3	S1	S2	.....	tenaga kerja_laki	tenaga kerja_perempuan
0	0	0	0	1	0	0	.....	2	2
1	0	0	0	0	0	0	.....	1	0
0	0	0	0	0	0	0	.....	0	0
0	0	0	0	0	1	0	.....	1	0
0	0	0	0	0	1	0	.....	5	3
0	0	0	0	0	0	0	.....	2	0
0	0	0	0	0	0	0	.....	0	0
0	0	0	0	0	1	0	.....	0	0
0	0	0	0	0	1	0	.....	4	0
0	0	0	0	0	0	0	.....	0	0
0	0	0	0	0	0	0	.....	0	1
0	0	0	0	0	0	0	.....	0	1
0	0	0	0	0	0	0	.....	0	0
0	0	0	0	0	1	0	.....	0	0
0	0	0	0	0	1	0	.....	1	0
0	1	0	0	0	0	0	.....	0	0
1	0	0	0	0	0	0	.....	3	2
0	0	0	0	0	0	0	.....	0	0
0	0	0	0	0	1	0	.....	0	1
0	0	0	0	0	1	0	.....	0	1
1	0	0	0	0	0	0	.....	0	0
....	....	....	....	....	....	....	.....	....	....
....	....	....	....	....	....	....	.....	....	....
0	0	0	0	0	0	0	.....	0	0

### 3.4. Proses AHC *Completed Lingked*

Tahap selanjutnya dengan menghitung AHC menggunakan *Completed lingked*. Dalam penghitungan ini akan dilakukan penghitungan jarak antar data menggunakan *Euclidian Distance* untuk menghitung jarak terdekat pada setiap umkm. Dikatakan jarak terdekat jika data memiliki jarak mendekati 0 atau merupakan jarak terkecil.

$$\|x - y\|^2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{2}$$

Keterangan:

$x_i$  = value data 1

$y_i$  = value data 2

Menghitung jarak *euclidean* pada setiap data umkm. Seperti pada 2 sampel berikut:

a) Menghitung jarak umkm 1 dan 2

$$\begin{aligned} & \sqrt{\left( \begin{aligned} & (0-1)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + \\ & (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (2-3)^2 + (0-1)^2 + (1-0)^2 + (1-1)^2 + \\ & (1-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-1)^2 + (0-0)^2 + (1-0)^2 + \\ & (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-1)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + \\ & (0-0)^2 + (0-0)^2 + (1-1)^2 + (1-1)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (1-1)^2 + \\ & (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + \\ & (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (1-1)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + \\ & (0-1)^2 + (2-1)^2 + (2-0)^2 \end{aligned} \right)} \\ & = \sqrt{\left( \begin{aligned} & 1 + 0 + 0 + 0 + 1 + 0 + 0 + 0 + 0 + \\ & 0 + 0 + 0 + 1 + 1 + 1 + 0 + 1 + 0 + \\ & 0 + 0 + 0 + 1 + 0 + 1 + 0 + 0 + 0 + \\ & 1 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + \\ & 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + \\ & 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + \\ & 1 + 0 + 1 + 1 + 4 \end{aligned} \right)} \\ & = \sqrt{16} \\ & = 4 \end{aligned}$$

b) Menghitung jarak UMKM 1 dan 3

$$\begin{aligned} \|U - V\|_2 &= \sqrt{\sum_i (U_i - V_i)^2} \\ &= 4,69041576 \end{aligned}$$

*Euclidean distance* dihitung untuk seluruh data hasil transformasi sehingga menghasilkan matrik kedekatan, hasil matrik kedekatan ditunjukkan pada Tabel 4.

**Tabel 4.** Matrik Kedekatan

	UMKM 1	UMKM 2			UMKM 6	UMKM 399
UMKM 1	0					
UMKM 2	4	0				
...	...	...				
...	...	...				
...	...	...				
UMKM 6	3,605551275	3,605551175			0	
UMKM 399	7,549834435	6,08276253			7,483314774	0

Selanjutnya adalah menentukan jarak terkecil. Dari matrik kedekatan terdapat umkm 6, 1, dan 2 yang memiliki jarak terkecil maka dilakukan penggabungan dan menghasilkan matrik baru pada Tabel 5.

**Tabel 5.** Hasil Penggabungan Jarak Terdekat

	UMKM 1 UMKM 2 UMKM 6				UMKM 399
UMKM 1 UMKM 2 UMKM 6	0				
...		0			
...		...	0		
...		...	...	0	
UMKM 399		4,582575695	5,830951895	7,681145748	0

Pada data tabel 5 terdapat kekosongan jarak, untuk mengisi kekosongan jarak tersebut diisi menggunakan *complete linkage*. Yaitu dengan mengecek setiap nilai jarak pada setiap data. Untuk mengisi data pada jarak UMKM(1,2,6) dengan UMKM(3), dilakukan dengan perbandingan jarak dari matriks kedekatan UMKM 1 dan 3, UMKM 2 dan 3, UMKM 6 dan 3. Dari ketiga data didapatkan jarak dari matriks daintaranya 4,69041576; 3,741657387; dan 5. Lalu dari ketiga data itu dipilih jarak yang paling jauh atau besar sehingga didapatkan 5 sebagai jarak antara UMKM(1,2,6) dengan UMKM(3). Lalu dilanjutkan dengan menentukan jarak terkecil pada matriks baru dan menggabungkan lagi, kemudian mengisi kekosongan jarak dengan *complete linkage*. Tahap ini dilakukan terus menerus hingga tersisa 2 kelompok. Perulangan ini menghasilkan iterasi pada Tabel 6 hanya menggunakan 7 data UMKM.

**Tabel 6.** Hasil Pengelompokkan

Tahap	Jumlah hasil cluster	Hasil Anggota Cluster	Jarak Terkecil
1	5	UMKM 1,	3,605551275
		UMKM 2,	
		UMKM 6	
2	4	UMKM 3,	4,582575695
		UMKM 7	
3	3	UMKM 1,	5
		UMKM 2,	
		UMKM 6,	
		UMKM 7	
4	2	UMKM 4	5,916079783
		UMKM 5	

Selanjutnya adalah menguji hasil klaster menggunakan *silhouette coefficient*. Pada pengujian ini akan dilakukan dengan menguji *silhouette 2* klaster.

- a) Menghitung rata – rata jarak setiap umkm dengan umkm lain yang masih dalam satu klaster dengan persamaan. Hasil menghitung rata-rata ditunjukkan pada Tabel 7.

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in Cd(i,j)} \quad (3)$$

Keterangan :

a(i) = Perbedaan rata – rata objek (i) ke semua objek lain pada A.

d(I,j) = Jarak antara i ke j

A = Cluster

**Tabel 7.** Hasil Perhitungan Rata Rata Jarak dengan satu kluster

Data	a(i)	Cluster
UMKM 1	4,961450368	0
UMKM 2	4,357492798	0
UMKM 3	4,503662211	0
UMKM 4	5,916079783	1
UMKM 5	5,916079783	1
UMKM 6	4,923604331	0
UMKM 7	6,424621859	0

b) Lalu dilanjutkan dengan menghitung rata – rata jarak umkm dengan umkm lain yang berbeda kluster. Hasil penghitungan jarak dengan cluster ditunjukkan pada Tabel 8.

$$d(i, C) = \frac{1}{|A|} \sum_{j \in C} d(i, j) \quad (4)$$

$$b(i) = \min_{C \neq A} d(i, C) \quad (5)$$

Keterangan:

d(i,C) = perbedaan rata - rata objek (i) ke semua objek lain pada C.

C = Cluster lain selain cluster A atau cluster C tidak sama dengan A.

b(i) = Rata – rata jarak terkecil umkm dengan setiap umkm pada cluster lain

**Tabel 8.** Hasil Perhitungan Jarak dengan Cluster lain

Data	d(I, 0)	d(I, 1)	b(i)	Cluster
UMKM 1	-	10,41165512	10,41165512	0
UMKM 2	-	9,924428901	9,924428901	0
UMKM 3	-	8,368265443	8,368265443	0
UMKM 4	8,971018069	-	8,971018069	1
UMKM 5	9,455268467	-	9,455268467	1
UMKM 6	-	7,885164805	7,885164805	0
UMKM 7	-	6,756048822	6,756048822	0

c) Lalu dilanjutkan dengan menghitung nilai silhouette. Hasil penghitungan silhouette ditunjukkan pada Tabel 9.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (6)$$

Keterangan:

S(i) = Silhouette Coefficient

a(i) = Perbedaan rata – rata objek (i) ke semua objek lain pada A

b(i) = Rata – rata jarak terkecil umkm dengan setiap umkm pada cluster lain

**Tabel 9.** Hasil Perhitungan Silhouette

Data	d(I, 0)	d(I, 1)	b(i)	S(i)	Cluster
UMKM 1	-	10,41165512	10,41165512	0,523471503	0
UMKM 2	-	9,924428901	9,924428901	0,56093264	0
UMKM 3	-	8,368265443	8,368265443	0,461816521	0
UMKM 4	8,971018069	-	8,971018069	0,34053418	1
UMKM 5	9,455268467	-	9,455268467	0,37430864	1
UMKM 6	-	7,885164805	7,885164805	0,375586376	0
UMKM 7	-	6,756048822	6,756048822	0,04905633	0
Rata – rata s(i)				0,383672313	

Lalu dilanjutkan dengan melakukan evaluasi pola. Berdasarkan dari percobaan menggunakan sampel data didapatkan nilai *silhouette* sebesar 0,383672313. Yang artinya nilainya mendekati 0 dan 2 *cluster* bukan merupakan *cluster* yang baik.

### 3.5. Pengujian Hasil dari Sistem

Pada penelitian ini menggunakan pengujian *Silhouette coefficient*. Pada system telah dilakukan 9 kali pengujian terhadap pengujian 10 klaster Tabel 10 merupakan hasil *silhouette coefficient* setiap klaster. Hasil pengujian menunjukkan bahwa cluster 2 memiliki nilai kaurasi tertinggi yaitu 0,729.

**Tabel 10.** Hasil Pengujian

Jumlah Cluster	Hasil <i>Silhouette Coefficient</i>
2	0,729
3	0,709
4	0,488
5	0,473
6	0,471
7	0,423
8	0,423
9	0,420
10	0,264

### 3.6. Representasi pengetahuan

Dari hasil pengujian dari sistem didapatkan 2 kelompok sebaran UMKM bidang jasa yang memiliki nilai *silhouette* 0,729. Hasil analisis persebaran UMKM bidang jasa di Kota X pada 2 kelompok tersebut adalah::

a) Kelompok 1

Pada Kelompok 1 ini terdiri dari 16 data UMKM yang memiliki rata-rata lama berdiri 41 tahun, Pendidikan dari pemilik UMKM Sekolah Menengah ke bawah, jumlah karyawan rata rata 1 orang dan omset pertahun rata rata di bawah 10 juta.

b) Kemompok 2

Pada kelompok 2 terdiri dari 383 data UMKM yang memiliki rata-rata lama berdiri 7 tahun, Pendidikan Pemilik UMKM Sekolah Menengah ke atas, jumlah karyawan rata-rata 1 dan ada yang tidak memiliki karyawan serta rata-rata omset pertahunnya di atas 10 juta.

Dari hasil analisis persebaran UMKM 2 kelompok tersebut dapat dijadikan acuan bagi Dinas Perindustrian Koperasi Usah Kecil dan Menengah di kota X dalam meningkatkan UMKM -UMKM yang belum maju yang terlihat pada UMKM di kelompok 1. Salah satu usulan yang bisa dilakukan disini adalah peningkatan pengetahuan untuk menignkatkan omset pertahun dengan memberikan pelatihan-pelatihan kepada UMKM pada kelompok1.

## 4. Kesimpulan

Kesimpulan dalam penelitian ini adalah bahwa penerapan algoritma AHC *complete linkage* dapat dilakukan pada UMKM bidang jasa menghasilkan 2 cluster yang direkomendasikan dengan nilai akurasi 0,729, dimana kelompok 1 merupakan kelompok UMKM yang perlu mendapatkan perhatian khusus dari Dinas Perindustrian Koperasi Usah Kecil dan Menengah kota X karena usia berdiri sudah lama namun omsetnya masih dibawah 10 juta.

## Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Lembaga Penelitian dan Pengabdian Masyarakat Universitas Ahmad Dahlan yang telah memberikan hibah pada penelitian ini melalui skem Penelitian Terapan tahun 2022.

## Daftar Pustaka

- [1] I. R. Maksum, A. Y. Sri Rahayu, and D. Kusumawardhani, "A social enterprise approach to empowering micro, small and medium enterprises (SMEs) in Indonesia," *J. Open Innov. Technol. Mark. Complex.*, vol. 6, no. 3, 2020, doi: 10.3390/JOITMC6030050.
- [2] A. H. Jankowiak, "Industrial Clusters As an Important Factor for Development of Micro, Small and Medium-Sized Enterprises (Msmes)," *12Th Int. Days Stat. Econ.*, no. November, pp. 708–718, 2018.
- [3] P. Puntoriza and C. Fibriani, "Analisis Persebaran UMKM Kota Malang Menggunakan Cluster K-means," *JOINS (Journal Inf. Syst.*, vol. 5, no. 1, pp. 86–94, 2020, doi: 10.33633/joins.v5i1.3469.
- [4] M. T. Satya, N. K. Rahayu, and A. Fidayan, "Pengklasifikasian Karakteristik Iklan UMKM Menggunakan Analisis Cluster," *Jesya (Jurnal Ekon. Ekon. Syariah)*, vol. 3, no. 2, pp. 218–231, 2020, doi: 10.36778/jesya.v3i2.169.
- [5] D. Solihin, A. Ahyani, K. Karolina, L. Pricilla, and I. S. Octaviani, "Pelatihan Pemasaran Online Berbasis Digital Untuk Meningkatkan Penjualan Bisnis Online Pada Umkm Di Desa Cicalengka Kecamatan Pagedangan Kabupaten Tangerang," *Dedik. Pkm*, vol. 2, no. 3, pp. 307–311, 2021, doi: 10.32493/dedikasipkm.v2i3.10726.
- [6] D. Astuti, A. R. Iskandar, and A. Febrianti, "Penentuan Strategi Promosi Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 1, no. 2, pp. 60–72, 2019, doi: 10.20895/inista.v1i2.71.
- [7] I. W. Rahayu *et al.*, "Hierarchical Clustering Untuk Deteksi Komunitas Pada Media Sosial Facebook Analysis and Implementation of Agglomerative Hierarchical Clustering Algorithm for Community Detection in Social Media Facebook," *e-Proceeding Eng.*, vol. 5, no. 1, pp. 1460–1468, 2018.
- [8] R. O. Pratikto and N. Damastuti, "Klasterisasi Menggunakan Agglomerative Hierarchical Clustering Untuk Memodelkan Wilayah Banjir," *JOINTECS (Journal Inf. Technol. Comput. Sci.*, vol. 6, no. 1, p. 13, 2021, doi: 10.31328/jointecs.v6i1.1473.
- [9] L. Zahrotun, "Analisis Pengelompokan Jumlah Penumpang Bus Trans Jogja Menggunakan metode Clustering K-Means dan Agglomerative Hierarchical Clustering (AHC)," *J. Inform.*, vol. 9, no. 1, pp. 1039–1047, 2015.
- [10] D. Rawat, R. K. Mittal, and V. S. Aggarwal, "Cluster Development Approach in India: an Antidote for Micro, Small, and Medium Enterprises," *Indian J. Commer. Manag. Stud.*, vol. VIII, no. 2, pp. 19–29, 2017, doi: 10.18843/ijcms/v8i2/03.
- [11] B. Harli Trimulya Suandi As and L. Zahrotun, "Penerapan Data Mining Dalam Mengelompokkan Data Riwayat Akademik Sebelum Kuliah dan data Kelulusan Mahasiswa Menggunakan Metode Agglomerative Hierarchical Clustering(Implementation Of Data Mining In Grouping Academic History Data Before Students And Stud," *J. Teknol. Informasi, Komput. dan Apl.*, vol. 3, no. 1, pp. 62–71, 2021, [Online]. Available: <http://jtika.if.unram.ac.id/index.php/JTIKA/>
- [12] K. E. T. Luthfi, *Algoritma Data Mining*, 1st ed. Yogyakarta: C.V Andi Offset, 2019.
- [13] E. Prasetyo, *DATA MINING-Konsep dan Aplikasi Menggunakan MATLAB*, 1st

- ed. Yogyakarta: Andi Offset, 2012.
- [14] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data An Introduction to Cluster Analysis*. Canada: John Wiley & Sons, Inc, Hoboken, New Jersey.