

Data Mining* Pengelompokan UMKM di Bidang Fashion dan Kerajinan Kota Yogyakarta Menggunakan *AHC Average Linkage

Muhammad Faishal¹, Rahmat Juniardi², Lisna Zahrotun³, Utaminingsih Linarti⁴, Anna Hendri Soleliza Jones⁵

^{2 3 5} Program Studi Informatika, Fakultas Teknologi Industri, Universitas Ahmad Dahlan

Jl. Ringroad Selatan, Kragilan, Tamanan, Kec. Banguntapan, Kab. Bantul, Daerah Istimewa Yogyakarta 55191

^{1 4} Program Studi Teknik Industri, Fakultas Teknologi Industri, Universitas Ahmad Dahlan

Jl. Ringroad Selatan, Kragilan, Tamanan, Kec. Banguntapan, Kab. Bantul, Daerah Istimewa Yogyakarta 55191

muhammad.faishal@ie.uad.ac.id,

ABSTRAK

UMKM di Kota Yogyakarta saat ini tumbuh sangat pesat dan menjadi salah satu penopang ketahanan ekonomi dengan tenaga kerja yang mendominasi. Pemerintah melalui Dinas Koperasi dan UMKM DIY sebagai pengelola memiliki kesulitan dalam mengolah data yang ada. Terdapat sebanyak 32.513 unit UMKM di berbagai bidang dalam database pemerintah yang terinput dari tahun 2021 hingga 2022 dan dalam kondisi masih tercampur. Pemerintah berniat untuk mengelompokkan data UMKM tersebut agar dapat dilakukan pengembangan dan pemberdayaan kedepannya. Diawali dari bidang fashion dan kerajinan sebagai salah satu komoditas penggaet wisatawan. Baru terdapat 195 unit UMKM bidang ini dari 2769 unit yang terverifikasi aktif pada database milik pemerintah. Metode yang digunakan dalam pengelompokan UMKM ini adalah AHC Average Linkage, dan pengujian yang dilakukan dari hasil pengelompokan menggunakan metode Silhouette Coefficient. Hasil penelitian ini berupa informasi pengelompokan UMKM bidang fashion dan kerajinan. Skor silhouette yang didapat pada UMKM fashion dan UMKM kerajinan sebesar 0.64 dan 0.65, didapatkan 2 cluster setiap UMKM-nya dan cluster yang dipilih untuk informasi kepada pemerintah adalah cluster 1. Kriteria UMKM fashion adalah kegiatan usaha penjualan, pemasaran dari DIY hingga ekspor, pendapatan kurang dari 10 juta hingga kurang dari 25 juta dan umur usaha 6,5 tahun. Kriteria UMKM kerajinan adalah kegiatan usaha penjualan

dan produksi, pemasaran dari DIY hingga ekspor, pendapatan kurang dari 10 juta dan umur usaha 7,1 tahun.

Kata kunci: *Agglomerative Hierarchical Clustering, Data Mining, Clustering UMKM*

ABSTRACT

MSMEs in the city of Yogyakarta are currently growing very rapidly and are one of the pillars of economic resilience with a dominating workforce. The government, through the DIY Cooperatives and MSMEs Service as the manager, has difficulty processing existing data. There are as many as 32.513 MSMEs units in various fields in the government database which were input from 2021 to 2022 and are still mixed. The government intends to group the MSMEs data so that development and empowerment can be carried out in the future. Starting from the fashion and crafts sector as one of the commodities that attracts tourists. There are only 195 MSMEs units in this sector out of 2769 units verified as active in the government database. The method used in grouping MSMEs is AHC Average Linkage, and testing carried out from the grouping results uses the Silhouette Coefficient method. The results of this research are information on the grouping of MSMEs in the fashion and crafts sector. The silhouette scores obtained for fashion MSMEs and craft MSMEs were 0.64 and 0.65, 2 clusters were obtained for each MSMEs and the cluster chosen for information to the government was cluster 1. The criteria for fashion MSMEs were sales and marketing business activities from DIY to export, low income from 10 million to less than 25 million and a business age of 6.5 years. The criteria for craft MSMEs are sales and production business activities, marketing from DIY to export, income of less than 10 million and business age of 7.1 years

Keywords: *Agglomerative Hierarchical Clustering, Data Mining, Clustering UMKM*

1. PENDAHULUAN

UMKM adalah istilah dalam roda perekonomian Indonesia yang berarti Usaha Mikro Kecil dan Menengah adalah model bisnis berskala kecil dijalankan oleh masyarakat secara perorangan, rumah tangga, hingga badan usaha. UMKM memiliki peranan yang sangat penting untuk perekonomian negara Indonesia, karena memiliki dampak baik bagi masyarakat yang ikut partisipasi, terciptanya lapangan kerja baru (Rudjito, 2003). Kota Yogyakarta merupakan wilayah dengan banyak UMKM dalam menopang pertumbuhan dan ketahanan ekonomi wilayah yang mendominasi jumlah tenaga kerja di Provinsi DIY (PERATURAN GUBERNUR DAERAH ISTIMEWA YOGYAKARTA TENTANG PEDOMAN PENGELOLAAN DATA USAHA MIKRO, KECIL, DAN MENENGAH, 2019). Data dari Pemkot Yogyakarta di Dinas Koperasi dan Usaha Kecil Menengah Daerah Istimewa Yogyakarta (DISKOP UMKM DIY) sebesar 32.513,00 unit UMKM di berbagai bidang pada bulan Juli tahun 2021 hingga tahun 2022, namun data ini belum banyak terverifikasi kebenaran UMKM yang masih aktif atau tidak aktif, sehingga pengelolaan data akan mengalami hasil tidak konsisten karena kurang akuratnya data yang diolah untuk kepentingan peningkatan pengelolaan data dan pemberdayaan seperti bantuan pendanaan, fasilitas, akses kredit perbankan. Kondisi UMKM di kota Yogyakarta dapat dipantau dengan aplikasi yang dikembangkan oleh Pemkot bernama JSS (*Jogja Smart Service*). Dalam aplikasi tersebut didapatkan data UMKM yang terverifikasi sebesar 2769 unit UMKM aktif dalam menjalankan usaha pada bulan Juli tahun 2022. Data yang terverifikasi tersebut terdapat beberapa bidang yaitu fashion sebanyak 195 unit, kerajinan sebanyak 145 unit, kuliner sebanyak 1336 unit, jasa sebanyak 399 unit dan campuran sebanyak 694 unit (Trisna, 2022).

Bidang *fashion* dan kerajinan menjadi salah satu komoditas untuk menggaet wisatawan berbelanja produk *fashion* dan kerajinan di kota Yogyakarta (Luthfiani, 2019). Contoh produk *fashion* kerajinan berupa baju batik, tas, baju adat Jawa, aksesoris, hingga blankon yang ditawarkan salah satu sentra UMKM di jalan Malioboro yang bernama Teras Malioboro, menjadi fasilitas dalam mendatangkan wisatawan yang berkualitas sebagai target calon pembeli UMKM dan mengembangkan berbagai event menarik (Danar, 2022). Contoh program pelatihan yang diadakan oleh Dinas Koperasi UMKM kota Yogyakarta melalui edaran pengumuman No. 893/091 tentang Diseminasi dan Pelatihan UMKM Tahun 2022 membuka kesempatan bagi para pelaku UMKM menumbuhkan wirausaha baru dalam upaya peningkatan jumlah wirausaha, menciptakan lapangan kerja, dan mendorong UMKM untuk naik kelas.

Hasil dari wawancara dengan Utaminingsih Linarti, S.T., M.T., selaku dosen Fakultas Teknik Industri di Universitas Ahmad Dahlan menjadi penghubung ke Pemkot Yogyakarta dalam pengumpulan data UMKM yang terverifikasi dalam bentuk data mentah yang belum dikelola, sehingga terdapat kelompok UMKM yang masih dalam tahap pengenalan usaha kepada pelanggan dengan UMKM yang menjalankan usaha kurun waktu yang cukup lama, diperlukan pengelolaan UMKM dengan pengelompokan unit usahanya berdasarkan kategori tertentu agar pemerintah lebih mudah dalam melakukan pengembangan dan pemberdayaan UMKM.

Berdasarkan persoalan di atas, pengelolaan data UMKM *fashion* dan bidang kerajinan diperlukan pengelompokan data menggunakan algoritma *Agglomerative Hierarchical Clustering Average Linkage*, karena algoritma ini mampu menangkap cluster dengan bentuk yang berbeda, mengakomodasi cluster yang berbentuk tidak beraturan dalam data dan juga algoritma *clustering* ini menggunakan pendekatan analisis hirarki *cluster* dengan mengelompokkan data yang dekat satu sama lain dan direpresentasikan secara grafis dalam bentuk *dendogram* (Maya et al., 2023). Hasil pengelompokan ini diharapkan dapat membantu

pemerintah kota yogyakarta dalam melakukan pengembangan dan pemberdayaan UMKM di bidang *fashion* dan kerajinan.

2. METODE

2.1. Agglomerative Hierarchical Clustering Average Linkage

Pengelompokan hirarki (*Hierarchical Clustering*) adalah metode analisis kelompok yang berusaha membangun sebuah hirarki kelompok data. Strategi pengelompokan umumnya ada 2 jenis yaitu *Agglomerative* metode pengelompokan hirarki dengan pendekatan bawah-atas dan *Divisive* metode pengelompokan hirarki dengan pendekatan atas-bawah (Sadewo et al., 2021).

Metode dalam penelitian ini adalah dengan mengimplementasikan algoritma *AHC Average Linkage*, dimana jarak antara dua kelompok didefinisikan sebagai jarak rata-rata antara semua pasangan data, masing-masing pasangan terdiri dari object kelompok (Marinova–Boncheva, 2008). Sedangkan dalam penelitian ini, tools yang digunakan adalah Phyton3, Framework Flask, Library Python untuk AHC Clustering.

Langkah-langkah penggunaan AHC Average Linkage secara umum :

1. Mencari jarak minimum dari dua objek.
2. Gabungkan dua objek dengan jarak minimum menjadi satu cluster.
3. Cari jarak antar cluster dengan menggunakan rata-rata.
4. Ulangi langkah no 1 dan 2 sampai semua objek bergabung menjadi satu cluster.

Untuk menghitung jarak antar cluster menggunakan *Euclidean Distance* :

$$D(a, b) = \sqrt{\sum_{i=1}^n (b_i - a_i)^2}$$

Keterangan :
D(a,b) = Kriteria
bi = nilai data latih
ai = nilai data uji

Berikut rumus Agglomerative Hierarchical Average Linkage:

$$d_{uv} = \text{average}\{d_{uv}\}, d_{uv} \in D$$

Keterangan :
D = Kriteria
d_{uv} = Data
∈ = Eliminasi

2.2. Silhouette Coefficient

Dalam pengujian cluster validation menjadi proses yang dapat menilai tingkat keberhasilan proses penambangan data yaitu proses clustering. Pengujian ditujukan untuk mengetahui kualitas hasil clustering yang didapat, maka pada penelitian ini digunakan metode silhouette coefficient untuk menguji kualitas clustering dan sebagai presentasi setiap kelompok pada sebuah cluster.

Silhouette coefficient adalah metode pengujian dari penggabungan 2 (dua) metode yaitu metode cohesion yang memiliki fungsi dalam mengukur jarak atau relasi antara objek dalam suatu cluster, kemudian ada metode separation yang memiliki fungsi dalam mengukur seberapa jauh suatu cluster terpisah dari cluster yang lain (Paembonan & Abduh, 2021).

- a. Menghitung rata-rata jarak objek dengan semua objek lain yang berada pada satu cluster, ambil nilai paling minimum dengan persamaan :

$a(i) = \frac{1}{[A] - 1} \sum_{j \in C} d(i, j)$	Keterangan :
	<p>a(i) = Perbedaan rata-rata objek (i) ke semua objek lain pada A.</p> <p>d(i,j) = Jarak antara data i hingga data j</p> <p>A = Cluster</p>

- b. Menghitung rata-rata jarak objek dengan objek lain yang berada pada cluster lain, ambil nilai paling minimum dengan persamaan :

$d(i, C) = \frac{1}{[A]} \sum_{j \in C} d(i, j)$	Keterangan :
	<p>d(i,C) = Perbedaan rata-rata objek (i) ke semua objek lain pada C.</p> <p>C = Cluster lain selain cluster A atau cluster yang tidak sama dengan cluster C.</p>

- c. Menghitung nilai Silhouette dengan persamaan :

$S(i) = \frac{b(i) - a(i)}{\max \{a(i); b(i)\}}$	Keterangan :
	<p>S(i) = Silhouette Coefficient</p> <p>a(i) = Rata-rata jarak objek dengan semua objek yang berada dalam satu cluster</p> <p>b(i) = Nilai minimum dari rata-rata jarak cluster pada cluster yang lain.</p>

2.3. One-Hot Encoding

Dalam proses data mining atau machine learning pemrosesan data yang diperlukan adalah data dalam bentuk angka atau integer, untuk mengatasi data dalam bentuk teks atau data jenis kategori perlu proses One-Hot-Encoding untuk mentransformasikan menjadi menjadi bentuk array 1 dimensi kemudian dipecah sebanyak class pada data teks dan memiliki nilai biner, jika data tersebut memiliki salah satu dari class yang dibuat maka akan memiliki nilai 1, sedangkan sisanya akan memiliki nilai 0 (Tamara, 2023).

2.4. Rank Transformation

Teknik dalam data transformasi dengan mengubah data ordinal menjadi angka untuk memberikan skala nilai yang sama dan mengurutkan datanya berdasarkan nilai datanya (Han et al., 2012).

2.5. Normalisasi Min-Max

Normalisasi data yang memiliki rentang nilai tertentu diubah ke bentuk yang lebih mudah untuk dipahami dengan skala 0 hingga 1 (Arhami et al., 2020).

Berikut rumus yang digunakan untuk normalisasi Min-Max :

$$Vn' = \frac{Vn - Min}{Max - Min}$$

Keterangan :

Vn' = data

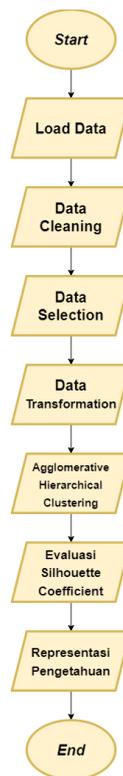
Vn = nilai data

Min= nilai minimum

Max = nilai maksimum

2.6. Tahapan Penelitian

Tahapan data mining dalam proses penambangan data dan proses penemuan informasi baru berupa hasil pengelompokan dengan metode *Agglomerative Hierarchical Clustering Average Linkage*.



Gambar 1. Tahapan Pengelompokan Data

2.6.1. Load Data

Proses load data adalah proses memindahkan data yang berupa file Microsoft Excel ke program pengolah data mining.

2.6.2. Data Cleaning

Pembersihan data dilakukan untuk melakukan pengisian data yang hilang, serta mengidentifikasi data yang akan menjadi outlier dan kemudian menghapusnya untuk mengatasi ketidakkonsistenan data yang akan diproses.

2.6.3. Data Selection

Proses pemilihan data atau atribut berdasarkan kebutuhan pengelompokan untuk proses data mining.

2.6.4. Data Transformation

Proses transformasi pada data yang sudah dipilih sebelumnya akan diubah dari suatu format atau struktur data ke format yang lain agar dapat diproses saat data mining.

2.6.5. Pengelompokan Agglomerative Hierarchical Clustering Average Linkage

Tahap AHC adalah pengelompokan setiap objek yang suatu cluster kemudian dicari objek lain yang mirip untuk dijadikan kelompok yang sama. Kemudian penggabungan objek ke dalam satu cluster dan ulangi langkah nya hingga tersisa satu cluster.

2.6.6. Pengujian Silhouette Coefficient

Setelah melakukan penambahan data dan pengelompokan cluster akan diuji berdasarkan data cluster-nya. Pengujian validasi juga menjadi penilaian untuk melihat kualitas hasil clustering apakah memiliki akurasi yang baik untuk dinyatakan hasil cluster memiliki kualitas yang baik.

2.6.7. Representasi Pengetahuan

Pada tahapan representasi dapat divisualisasikan dengan pengelompokan hirarki pada metode AHC yang sudah diolah akan dibuat ke bentuk dendrogram untuk dipresentasikan kepada pihak yang terkait sebagai pengetahuan baru.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan data

Setelah melakukan wawancara mengenai pengumpulan data dari Utaminingsih Linarti, S.T., M.T. sebagai narahubung Dinas Koperasi dan UMKM Daerah Istimewa Yogyakarta. Data yang didapat yakni data UMKM Kerajinan berjumlah 145 dan data pada UMKM Fashion berjumlah 194 pada tahun 2021 hingga 2022 yang telah diverifikasi dalam bentuk file excel.xls.

Format yang digunakan untuk Dataset UMKM tersebut sudah berupa file excel yang telah dipisah dengan nama "UKM Fashion.xlsx" dan "UKM Kerajinan.xlsx". Variabel atau atribut di dalam file tersebut terdapat 40 atribut yaitu No., Ref. OSS, Nomor Induk Kependudukan, Nama Lengkap, Tanggal Lahir, Usia, Jenis Kelamin, Pendidikan Terakhir, No. Telepon, e-Mail, Provinsi, Kab/Kota, Kecamatan, Desa/Kel, Nama Jalan, Nama Usaha, NIB, Tanggal Terbit NIB, Tanggal Pendirian Usaha, Provinsi, Kab/Kota, Kecamatan, Desa/Kel, Nama Jalan, Koordinat, Bidang Usaha, Sektor Usaha, Kegiatan Usaha, Produk Komoditas Ekspor, Tujuan Pemasaran, Status Kepemilikan Tanah/Bangunan, Sarana Media Elektronik, Modal Bantuan Pemerintah, Pinjaman Kredit Usaha Rakyat, Omset per-Tahun, Kepemilikan Asuransi Kesehatan, Laki-laki, Perempuan, Rerata Usia Pekerja, dan Status Formulir.

3.1.1. Kebutuhan data

Data sebagai objek yang akan diuji merupakan data Pendidikan Terakhir, Tanggal Pendirian Usaha, Kegiatan Usaha, Tujuan Pemasaran, Status Kepemilikan Tanah dan Bangunan,, Modal Bantuan Pemerintah, Pinjaman Kredit Usaha Rakyat, Omset per Tahun, Kepemilikan Asuransi Kesehatan, Tenaga Kerja Laki-laki, dan Tenaga Kerja Perempuan.

3.2. Data Mining

Berikut pengelompokan data atau pemrosesan data dengan data mining yang digunakan dalam penelitian menggunakan metode sebagai berikut :

3.2.1. Load Data

Berikut data yang ada di dalam dataset "UKMM Fashion" dan "UKMM Kerajinan" yang merupakan hasil dari proses Load data yang berisi atribut dan sampel datanya sebelum memasuki proses cleaning dan proses transformasi seperti pada Tabel 1.

Tabel 1. Load Data

No.	Ref. OSS	Nomor Induk Kependudukan	Nama Lengkap	...	Pendidikan Terakhir	...	Rerata Usia Pekerja	Status Formulir
20	5147	██████████	ACHMAD NURSOLEH, A.Md.	...	D3	...	-	Terverifikasi
24	107	██████████	ADEWANI DANGGORA LUBIS, S.E.	...	S1	...	25-35 tahun	Terverifikasi
32	8732	██████████	ADIK TRI AMINAH	...	SMK	...	-	Terverifikasi
151	3165	██████████	AMADEA PRISCILLIA DIANTY	...	SMA	...	-	Terverifikasi
220	15575	██████████	ANIS SUDARWATI	...	SMA	...	25-35 tahun	Terverifikasi
285	8804	██████████	ARI SANDI	...	SMA	...	35-50 tahun	Terverifikasi
324	4894	██████████	ASPRI ISMANTO	...	SMK	...	25-35 tahun	Terverifikasi
368	15969	██████████	BASIRIN	...	SMA	...	25-35 tahun	Terverifikasi
510	6973	██████████	DEVY LESTARI, S.H.	...	S1	...	35-50 tahun	Terverifikasi
620	17530	██████████	DWI PRIYANTO	...	SMA	...	35-50 tahun	Terverifikasi

3.2.2. Data Cleaning

Pembersihan data dilakukan untuk melakukan pengisian data yang hilang, serta mengidentifikasi data yang akan menjadi outlier dan kemudian menghapusnya untuk mengatasi ketidakkonsistenan data yang akan diproses. Pada data UMKM Fashion didapat sebanyak 187 dari 194 data, sedangkan data UMKM Kerajinan didapat sebanyak 136 dari 145 data setelah menghilangkan outlier yang ditemukan pada data yang memiliki jumlah tenaga kerja lebih dari 15 dan umur usaha yang terlalu tua lebih dari 30 tahun. Adapun hasil data cleaning seperti pada Tabel 2.

Tabel 2. Data Cleaning

ref_oss	nama_usaha	pendidikan	...	kepemilikan_tanah	modal_bantuan_pemerintah	...	omset_pertahun	asuransi	...
3619	SANDANG TRIAN	SMA	...	Sewa	-	...	Kurang dari 10 juta	BPJS	...
5147	TOKO FAY	D3	...	Milik sendiri	-	...	Kurang dari 10 juta	BPJS	...
...
8732	MH JEWELRY	SMK	...	Lainnya	Pemerintah Pusat	...	10 juta s/d 25 juta	BPJS	...
6408	ROARS ROCKINSTYLE	S1	...	Milik sendiri	-	...	Kurang dari 10 juta	BPJS	...
5846	DIVA CREATIVE	SMA	...	Milik sendiri	-	...	85 juta s/d 100 juta	BPJS	...
...
985	SYAHDU INSIDE	-	...	Magersari (adat)	-	...	Kurang dari 10 juta	-	...
3165	THRIFTHAUS	SMA	...	Milik sendiri	-	...	Kurang dari 10 juta	-	...

3.2.3. Data Selection

Proses seleksi pada tahapan ini berupa memilih atribut dalam dataset yang dapat diolah pada tahapan data mining. Berikut atribut yang akan digunakan dalam proses data mining serta pemberian label gaya penulisan dengan snake_case seperti pada Tabel 3.

Tabel 3. Data Selection

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

3.2.4. Data Transformation

Proses transformasi pada data yang sudah dipilih sebelumnya akan diubah dari suatu format atau struktur data ke format yang lain agar dapat diproses saat data mining.

Pada penelitian ini data yang akan diproses memiliki beberapa tipe data yaitu kategorikal, ordinal, dan numerik. Berikut pembahasan transformasi data berdasarkan tipe datanya :

a. Kategorikal

Transformasi pada tipe data kategorikal akan menggunakan metode One-Hot Encoding dengan memecah data menjadi atribut baru dan memiliki nilai biner dimana akan data yang memiliki nilai di atribut tersebut akan diberi nilai 1 sedangkan jika tidak akan bernilai 0. Berikut contoh atribut yang akan dirubah menggunakan One-Hot Encoding seperti terdapat pada Tabel 4.

Tabel 4. Data Transformation - Kategorikal

Penjualan	Penjualan & Produksi	Produksi	Dalam wilayah DIY	Dalam wilayah Kota Yogyakarta	Dalam wilayah Pulau Jawa	Dalam wilayah luar Pulau Jawa (tersebar)
0	0	1	1	0	0	0
0	1	0	1	1	1	1
0	1	0	1	0	0	0
0	1	0	0	1	0	0

0	1	0	1	1	1	1
---	---	---	---	---	---	---

b. Ordinal

Transformasi pada tipe data ordinal akan menggunakan metode rank transformation dan normalisasi min-max, dimana data akan diberi nilai berdasarkan peringkatnya dan prioritasnya kemudian diubah kembali ke bentuk skala dari 0 hingga 1. Berikut Tabel 5 berisi atribut yang akan diubah dengan metode rank transformation dan normalisasi min-max.

Tabel 5. Data Transformation - Ordinal

Atribut	Transformasi		
	Data	Rank Transformation	Normalisasi Min-Max
Pendidikan	Tidak ada pendidikan Formal	0	0
	SD	1	0,142
	SMP	2	0,285
	SMU	3	0,428
	Diploma	4	0,571
	S1	5	0,714
	S2	6	0,857
	S3	7	1
	kepemilikan_tanah	Lainnya	0
Magersari (adat)		1	0,33
Sewa		2	0,66
Milik Sendiri		3	1
modal_bantuan_pemerintah	Tidak ada modal bantuan pemerintah	0	0
	Pemkot Yogyakarta	1	0,33
	Pemda DIY	2	0,66
	Pemerintah Pusat	3	1
pinjaman	Tidak ada pinjaman	0	0
	Lainnya	1	0,5
	Koperasi	1	0,5
	Bank	1	0,5

Data Mining Pengelompokan UMKM di Bidang Fashion dan Kerajinan Kota Yogyakarta Menggunakan
AHC Average Linkage

	Pemerintah	2	1
omset_pertahun	Kurang dari 10 juta	0	0
	10 juta s/d 25 juta	1	0,1
	25 juta s/d 40 juta	2	0,2
	40 juta s/d 55 juta	3	0,3
	55 juta s/d 70 juta	4	0,4
	70 juta s/d 85 juta	5	0,5
	85 juta s/d 100 juta	6	0,6
	100 juta s/d 120 juta	7	0,7
	120 juta s/d 150 juta	8	0,8
	Lebih dari 150 juta	10	0,9
asuransi	Tidak ada asuransi	0	0
	Asuransi Swasta	1	0,5
	BPJS	2	1

c. Numerik

Data numerik pada penelitian ini tidak menggunakan metode dalam transformasi datanya, namun hanya menggunakan perhitungan matematik seperti data umur_usaha yang semula berisi data tanggal pendirian namun dikurangkan dengan waktu saat ini sehingga didapat data berupa umur dalam bentuk tahun, dan pada atribut tenaga kerja laki-laki dan tenaga kerja perempuan akan digabungkan menjadi atribut baru bernama tenaga_kerja. Berikut Tabel 6 berisi atribut yang akan diubah menggunakan perhitungan matematik.

Tabel 6. Data Transform - Numerik

Atribut	Keterangan	Transformasi		
		Tanggal Pendirian Usaha	Umur_usaha	
umur_usaha	'Tanggal Pendirian Usaha' dikurangi waktu saat ini	25 Juli 2019	4	
		12 Februari 2018	5	
		13 April 1975	48	
tenaga_kerja	Tenaga kerja 'laki-laki' ditambah 'perempuan'	Tenaga kerja Laki-laki	Tenaga kerja Perempuan	Tenaga kerja
		0	2	2

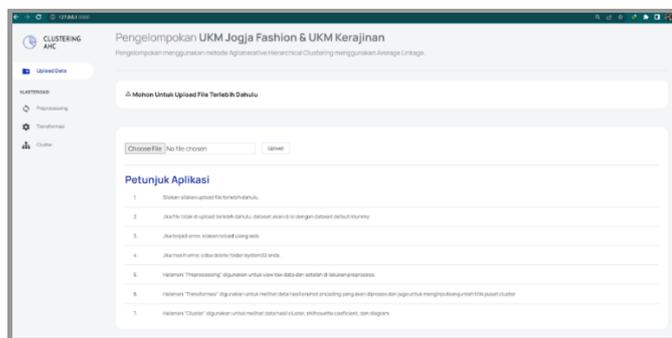
		3	4	7
		2	3	5

3.2.4. Implementasi

Dalam mengimplementasi Pengelompokan UMKM di bidang fashion dan Kerajinan menggunakan metode Agglomerative Hierarchical Clustering Average Linkage dapat dibuatkan aplikasi berbasis website dengan bahasa pemrograman yang dipakai berupa Python dan Framework yang bernama Flask untuk mengimplementasikan aplikasi data mining di Python menjadi web.

a. Halaman Dashboard

Halaman awal pada aplikasi berupa halaman Dashboard. Pada halaman ini terdapat informasi berupa petunjuk penggunaan aplikasi data mining, dan form untuk mengupload file UMKM seperti pada Gambar 2.



Gambar 2. Halaman Dashboard

b. Halaman Preprocessing

Pada halaman Preprocessing ini digunakan untuk menampilkan informasi berupa tabel 'Raw Data' untuk menampilkan isi tabel yang asli tanpa proses data mining dan informasi tabel 'Data Hasil Cleaning & Seleksi' yang merupakan tabel hasil proses data cleaning dan seleksi seperti pada Gambar 3 dan 4 berikut.

No.	Ref. OSS	Nomor Induk Kependudukan	Nama Lengkap	Tanggal Lahir	Ukura	Jenis Kelamin	Pendidikan Terakhir
4	5732	540213481030001	ADK TRI AURAH	07 November 1998	50	P	SMA
5	5409	5470441820002	ADUNG ISTANTO PUTRI, S.S.	02 November 1982	39	L	S1
6	5846	5470440860001	ADUS HARTONO	16 Agustus 1967	54	L	SMA
7	8703	5470820080001	ADUS SETIOWAN	28 Agustus 1981	40	L	SMA
8	988	5470201070004	ALIF SIBUCAP	20 Desember 1975	48	L	-
9	1951	5470105010001	SANDIKA PRICILLUS DANIY	14 Juli 2000	21	P	SMA
10	3633	5470805020001	JANNAH HUSNANDI	12 April 1978	44	P	SMP
11	176	550	ANASTASIA ANI BRP PURNAMA	21 Agustus 1973	48	P	S1
12	2405	180506470300001	ANASTASIA MARINA PURNAMA, S.P.	05 Februari 1993	29	P	S1
13	1362	760424210700004	ANDRIANASARI P	20 October 1977	44	L	SMA
14	13729	540312105820008	ANDRIAS WIRNANI	14 Juni 1982	39	L	SMP
15	202	550602009000002	ANGGA WIDHARTI	18 Maret 1996	27	P	SMA
16	15076	1807064020001	ANIS SUDHARTI	18 Februari 1963	62	P	SMA
17	3618	5471040080002	ANI PUTRI WULAN DARI	05 Juni 1989	32	P	SMA

Gambar 3. Halaman Preprocessing - Raw Data

Data Mining Pengelompokan UMKM di Bidang Fashion dan Kerajinan Kota Yogyakarta Menggunakan AHC Average Linkage

id	ref_usah	nama_usaha	pendidikan	tanggal_pendirian_usaha	kegiatan_usaha
1	8879	SINCEG TRISK	SMA	26 Juli 2018	Pengalutan
2	3347	TOUCH'FAY	SI	12 Februari 2018	Pengalutan
3	107	DONNA BATAK	SI	15 April 1975	Pengalutan, Produksi, Pany
4	8752	MAN JEREBAY	SMA	17 Maret 2016	Pengalutan
5	6038	ROARS ROCKNETTLE	SI	24 November 2016	Pengalutan, Produksi
6	1846	DWA CREATIVE	SMA	21 Juli 1998	Pengalutan, Produksi
7	8753	ADUNGSA	SMA	21 April 2016	Pengalutan
8	985	SHAKU INSEK	-	12 Desember 2018	Pengalutan, Produksi
9	3161	THEFTHUG	SMA	21 Desember 2017	Pengalutan
10	3533	ALPINO COLLECTION	SMP	27 Mei 1995	Pengalutan, Produksi
11	851	ALETTA COLLECTION	SI	28 April 2016	Pengalutan, Produksi
12	2426	PURIN SHIRAZ	SI	19 April 2016	Pengalutan, Produksi
13	13642	ABC ONLINE OFFLINE JEREBAYOLA	SMA	17 Desember 2018	Pengalutan
14	10229	CHELSA FASHION	SMP	10 Oktober 2020	Pengalutan
15	11832	EUNIKOL SHOP	SMA	30 Oktober 2020	Pengalutan, Produksi
16	10036	BUNYATA CLOTHES WORKSHOP	SMA	30 Maret 2016	Pengalutan

Gambar 4. Halaman Preprocessing - Cleaning & Selection

c. Halaman Transformasi

Pada halaman transformasi ini digunakan untuk menampilkan informasi berupa form untuk menginput titik pusat yang digunakan dalam clustering dan informasi berikutnya berupa tabel untuk menampilkan hasil proses transformasi setelah dilakukannya proses preprocessing seperti pada Gambar 5 dan 6 berikut.

Gambar 5. Halaman Transformasi - form input titik pusat

id	pendidikan	umur_usaha	Pengalutan	Pengalutan & Produksi	Produksi	Dalam wilayah DIF	Dalam wilayah Kota Yogyakarta	Dalam wilayah Pajera Jena	Dalam wilayah luar Pajera Jena (lora)
8	0.000000	-4.7154	0	1	0	0	0	0	0
9	0.428571	-1.7154	1	0	0	0	0	0	0
10	0.285714	18.28865	0	1	0	0	0	1	1
11	0.71286	-4.7154	0	1	0	1	0	0	0
12	1.000000	-9.7154	0	1	0	0	1	1	0
13	0.428571	9.28865	1	0	0	1	0	0	0
14	0.285714	-4.7154	1	0	0	1	0	0	0
15	0.428571	-6.7154	0	1	0	1	0	1	0
16	0.428571	-1.7154	1	0	0	1	1	0	0
17	0.428571	1.28865	0	1	0	0	0	0	1
18	0.428571	2.28865	0	1	0	1	1	0	0
19	0.571429	-3.7154	0	0	1	0	0	1	0
20	0.428571	-9.7154	1	0	0	0	0	1	0
21	0.000000	-4.7154	0	1	0	0	1	1	0
22	0.428571	-3.7154	0	0	1	1	1	1	1
23	0.71286	-1.7154	1	0	0	1	1	1	1
24	0.428571	-2.7154	1	0	0	1	1	0	0
25	0.285714	-6.7154	1	0	0	1	0	0	0

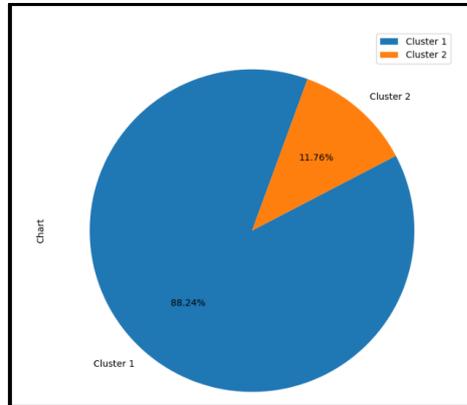
Gambar 6. Halaman Transformasi - tabel hasil transformasi

d. Halaman Transformasi

Pada halaman clustering ini digunakan untuk menampilkan informasi berupa 'Pie Chart' seperti pada Gambar 7 untuk memvisualisasikan jumlah kelompok yang dimiliki tiap cluster yang terbentuk dengan satuan persentase. Informasi berupa nilai 'Silhouette Coefficient' berdasarkan jumlah input titik pusat yang terdapat pada halaman transformasi, dan informasi berupa panduan ukuran nilai Silhouette Coefficient menurut 'Kaufman & Rousseeuw' seperti pada

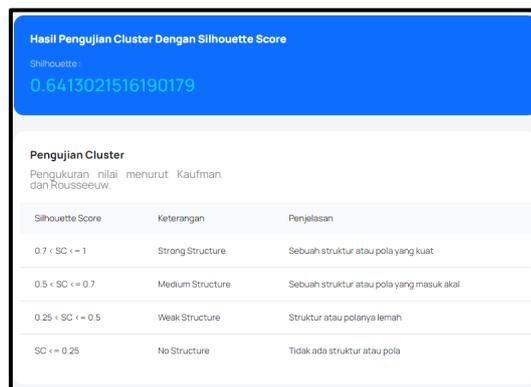
Gambar 8. Adapun halaman Clustering – Dendrogram dan hasil akhir seperti pada Gambar 9 dan 10.

1. Pie Chart



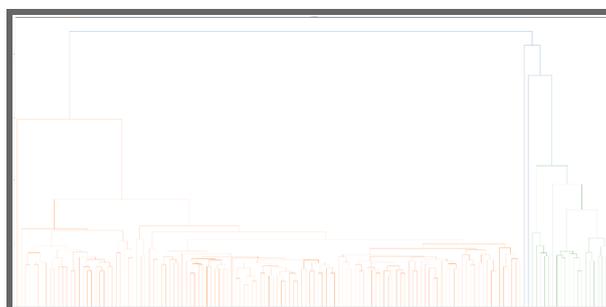
Gambar 7. Halaman Clustering - Pie Chart

2. Silhouette Coefficient



Gambar 8. Halaman Clustering – Silhouette Coefficient

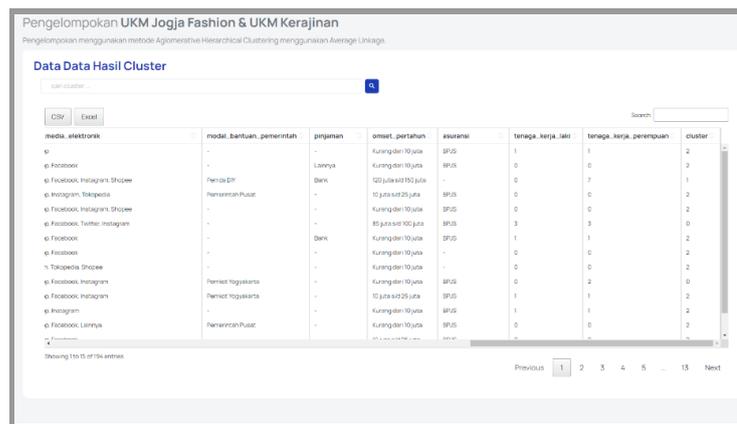
3. Dendrogram



Gambar 9. Halaman Clustering - Dendrogram

4. Hasil Pengelompokan

Data Mining Pengelompokan UMKM di Bidang Fashion dan Kerajinan Kota Yogyakarta Menggunakan *AHC Average Linkage*



Gambar 10. Halaman Clustering - Hasil Akhir

3.4. Hasil Pengujian

Dalam proses pengujian dapat disimpulkan dengan melihat nilai hasil Silhouette Coefficient untuk dapat melihat struktur pola berdasarkan input titik pusat yang diuji kemudian dapat dicari titik pusat dengan nilai yang baik sehingga dapat dirujuk bahwa pengelompokan memiliki kualitas yang baik seperti pada table 7.

Tabel 7. Nilai hasil Silhouette Coefficient

Titik Pusat	Silhouette Coefficient	
	UMKM Bidang Fashion	UMKM Bidang Kerajinan
2	0.6413021516190179	0.6503084709281498
3	0.49745283451340583	0.4788899592979065
4	0.47807233034542623	0.46277141751826567
5	0.4808892383578597	0.38326005266500074

Hasil pengelompokan berupa cluster yang akan digunakan sebagai hasil dari penelitian ini dapat dilihat pada tabel dibawah ini. Dalam cluster yang didapat bisa digunakan sebagai sumber informasi untuk dilakukannya layanan berupa program untuk pengembangan serta pemberdayaan untuk UMKM yang ada di Kota Yogyakarta. Pada UMKM bidang fashion dengan umur usaha yang berjalan kurang dari 10 tahun, belum meminjam pinjaman berupa kredit maupun bantuan berupa modal dari pemerintah, omset per tahun kurang dari 10 juta, dan tenaga kerja yang dimiliki kurang dari 2 SDM. Sedangkan untuk UMKM bidang kerajinan dengan umur usaha yang berjalan kurang dari 10 tahun, belum meminjam pinjaman kredit maupun bantuan berupa modal dari pemerintah, omset per tahun kurang dari 10 hingga yang memiliki omset kurang dari 25 juta, dan tenaga kerja yang dimiliki kurang dari 1.8 SDM.

Cluster	Bidang UMKM	Jumlah Anggota	Rata-rata									
			Pendidikan	Kegiatan Usaha	Pemasaran	Kepemilikan Tanah	Bantuan Pemerintah	Pinjaman Kredit	Omsset per Tahun	Asuransi	Tenaga Kerja	Umur Usaha
1	UMKM bidang Fashion	165	SMA/SMK	Penjualan	Dalam wilayah DIY, Dalam wilayah Pulau Jawa, Luar negeri (Import) ada 12 data)	Milik Sendiri	tidak memiliki bantuan pemerintah & Pemerintah Pusat	tidak ada pinjaman	Kurang dari 10 juta & 10 juta s/d 25 juta	BPJS	1 SDM	6,8 Tahun
2		22	SMA/SMK	Produksi	Dalam wilayah DIY	Milik Sendiri	tidak memiliki bantuan pemerintah & Pemkot Yogyakarta	tidak ada pinjaman & Bank	Kurang dari 10 juta	BPJS	2 SDM	23,4 Tahun
1	UMKM bidang Kerajinan	124	SMA/SMK	Penjualan, Produksi	Dalam wilayah DIY, Dalam wilayah Pulau Jawa, Luar negeri (Export) ada 13 data)	Milik Sendiri	tidak memiliki bantuan pemerintah & Pemerintah Pusat	tidak ada pinjaman	Kurang dari 10 juta	BPJS	2 SDM	7,1 Tahun
2		12	SMA/SMK	Produksi	Dalam wilayah DIY, Dalam wilayah Pulau Jawa, Luar negeri (Export) ada 2 data)	Milik Sendiri	tidak memiliki bantuan pemerintah	tidak ada pinjaman & Bank	10 juta s/d 25 juta	BPJS	2,5 SDM	24,8 Tahun

4. KESIMPULAN

Hasil pengelompokan menggunakan data UMKM bidang fashion sebanyak 187 unit dan UMKM bidang kerajinan sebanyak 136 unit, dengan cluster yang didapat pada penelitian ini sebanyak 2 cluster pada UMKM bidang fashion yang memiliki skor silhouette coefficient sebesar 0.64, cluster yang akan dipilih untuk menjadi informasi rekomendasi adalah cluster ke 1 sebanyak 165 data dengan rata-rata kriteria berupa UMKM dengan kegiatan usaha penjualan, pemasaran di wilayah DIY hingga Export, pendapatan pertahun kurang dari 10 juta hingga kurang dari 25 juta, tenaga kerja 1 SDM dan umur usaha sekitar 6,5.

Sedangkan pada UMKM bidang kerajinan yang memiliki skor silhouette coefficient sebesar 0.65, cluster yang akan dipilih untuk menjadi informasi rekomendasi adalah cluster ke 1 sebanyak 124 data dengan rata-rata kriteria berupa UMKM dengan kegiatan usaha penjualan dan produksi, pemasaran di wilayah DIY hingga Export, pendapatan pertahun kurang dari 10 juta, tenaga kerja 2 SDM dan umur usaha sekitar 7,1. Dengan demikian, hasil pengelompokan data UMKM DIY sudah terselesaikan dan semoga hasil pengelompokan ini bermanfaat untuk Dinas UMKM dan Koperasi DIY.

5. DAFTAR RUJUKAN

- Arhami, M., Kom, M., & Muhammad Nasir, S. T. (2020). *Data Mining-Algoritma dan Implementasi*. Penerbit Andi.
- Danar. (2022). Teras Malioboro Jadi Etalase UMKM. *Krjogja*. <https://www.krjogja.com/ekonomi/1242473741/teras-malioboro-jadi-etalase-umkm>
- PERATURAN GUBERNUR DAERAH ISTIMEWA YOGYAKARTA TENTANG PEDOMAN PENGELOLAAN DATA USAHA MIKRO, KECIL, DAN MENENGAH, Pub. L. No. 82, 1 (2019).
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*, Waltham: Morgan Kaufmann Publishers.
- Luthfiani, D. (2019). Dinas Pariwisata DIY Berupaya Menarik Kunjungan Wisatawan Lewat Busana. *TIMES Yogyakarta*. <https://timesindonesia.co.id/wisata/239368/dinas-pariwisata-diy-berupaya-menarik-kunjungan-wisatawan-lewat-busana>.
- Marinova–Boncheva, V. (2008). *Using the agglomerative method of hierarchical clustering as a data mining tool in capital market*.
- Maya, R., Lubis, F., Huang, J., Wang, P., Khoifin, K., & Elvina, Y. (2023). Agglomerative Hierarchical Clustering (AHC) Method for Data Mining Sales Product Clustering. *Building of Informatics, Technology and Science (BITS)*, 5(1), 285–294.

<https://doi.org/10.47065/bits.v5i1.3569>

- Paembonan, S., & Abduh, H. (2021). Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat. *PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik*, 6(2), 48. https://doi.org/10.51557/pt_jiit.v6i2.659
- Rudjito. (2003). Strategi Pengembangan UMKM Berbasis SInergi Bisnis. *Makalah Seminar Peran Perbankan Dalam Memperkokoh Ketahanan Nasional Kerja Sama Lemhanas RI Dengan BRI*.
- Sadewo, T. A., Kusuma, P. D., & Casi Setianingsih. (2021). Clustering Pada Data Sentimen Bpjs Kesehatan Menggunakan Algoritma Agglomerative Hierarchical Clustering Average Linkage. *E-Proceeding of Engineering*, 8(5), 1–23.
- Tamara, R. (2023). Data Mining Penentuan Jurusan Siswa Menggunakan Metode Agglomerative Hierarchical Clustering (AHC). *Jurnal Media Informatika Budidarma*, 07(02), 873–880. <https://doi.org/10.30865/mib.v7i2.6092>
- Trisna, Y. (2022). PELATIHAN UMKM BERBASIS DIGITAL DARI DINAS KOMUNIKASI DAN INFORMATIKA D.I.YOGYAKARTA. *Kemantren Mantirejojn*. <https://mantrijeronkec.jogjakota.go.id/detail/index/19785>.