

Perbandingan Metode *Clustering* Dengan Menggunakan Metode *Average Linkage* Dan Metode *K-Means* Pada Industri Kecil Dan Menengah Di Kabupaten Wajo

Ahmad Zikir

Program Studi Matematika, Universitas Islam Negeri Alauddin Makassar, 60600116011@uin-alauddin.ac.id

Khalilah Nurfadilah

Program Studi Matematika, Universitas Islam Negeri Alauddin Makassar, khalilah.nurfadilah@uin-alauddin.ac.id

Irwan

Program Studi Matematika, Universitas Islam Negeri Alauddin Makassar, irwan.msi@uin-alauddin.ac.id

Adiatma

Program Studi Matematika, Universitas Islam Negeri Alauddin Makassar, adiatma.rasyid@uin-alauddin.ac.id

ABSTRAK, Dalam upaya mengembangkan perekonomian di Indonesia, tentunya aspek ekonomi masyarakat menjadi hal yang perlu diperhatikan. Hal ini bertujuan untuk meningkatkan taraf hidup seluruh masyarakat Indonesia menjadi lebih baik. Salah satu upaya pemerintah dalam pembangunan ekonomi di Indonesia adalah Pengembangan Industri Kecil Menengah (IKM). Penelitian ini akan mengelompokkan IKM salah satu daerah di Sulawesi Selatan yakni Kabupaten Wajo dengan menggunakan Metode *Average Linkage Clustering* dan Metode *K-means*. Hasil penelitian menunjukkan bahwa penggunaan metode *Average Linkage*, diperoleh 1 industri di kategorikan sebagai anggota dari Industri Menengah, sementara 29 lainnya dikategorikan sebagai anggota Industri Kecil. Untuk metode *K-means*, sebanyak 6 industri di kategorikan sebagai anggota dari Industri Menengah, sementara 24 lainnya dikategorikan sebagai anggota Industri Kecil. Dengan melihat nilai centroid-nya, metode *K-means* dinilai lebih baik untuk merepresentasikan IKM di Kabupaten Wajo.

Kata Kunci: *Industri, Clustering, K-means, Average, Wajo*

1. PENDAHULUAN

Dalam upaya mengembangkan perekonomian di Indonesia, tentunya aspek ekonomi masyarakat menjadi hal yang perlu diperhatikan. Hal ini bertujuan untuk meningkatkan taraf hidup seluruh masyarakat Indonesia menjadi lebih baik. Salah satu upaya pemerintah dalam pembangunan ekonomi di Indonesia adalah pengembangan Industri Kecil Menengah (IKM).

IKM adalah industri yang memiliki jangkauan industri berskala kecil dan menengah. IKM dinilai berperan penting dalam upaya pemerintah membangun ekonomi nasional, ini dibuktikan dengan tingginya penyerapan jumlah

tenaga kerja pada tahun 2016, sekitar 97,22%. Salah satu daerah di Sulawesi Selatan yang memiliki perkembangan yang cukup baik dalam sektor industri adalah kabupaten Wajo.

Berdasarkan data Badan Pusat Statistik (BPS) yang terdapat dalam buku “Kabupaten Wajo dalam Angka 2017”, Kelompok IKM yang tercatat dengan jumlah unit usaha paling banyak di Kabupaten Wajo adalah industri tekstil, pakaian jadi, dan kulit dengan jumlah usaha sebanyak 109 usaha.

Melihat peranan dari IKM ini, maka perlunya dilakukan pengelompokan untuk mempermudah dalam identifikasi karakteristik sehingga pemerintah dapat menentukan kebijakan-kebijakan yang sesuai dengan kondisi IKM di setiap daerah. Tentu saja untuk mengelompokkan jenis industri di IKM, pemerintah memiliki kriteria yang menjadi landasan untuk menentukan bagian mana yang termasuk industri kecil dan bagian mana yang termasuk industri menengah. Namun yang menjadi masalah adalah, kita tidak tau metode apa yang sesuai untuk mengelompokkan hal tersebut.

Jika berbicara tentang pengelompokan, tentunya Analisis *Cluster* adalah analisis yang paling cocok untuk digunakan. Analisis *Cluster* merupakan analisis multivariat yang bertujuan untuk memecahkan masalah yang berkaitan dengan pengelompokan objek dengan mengamati karakteristiknya. Analisis *Cluster* memiliki dua metode dalam pengerjaannya yaitu Metode Hierarki dan Non-hierarki. Beberapa Metode Hierarki yang dikenal yaitu *Single Linkage*, *Average Linkage*, *Complete Linkage*,

dan Metode *Ward*. Sedangkan pada Metode Non-hierarchical salah satunya adalah Metode *K-means*.

Pada penelitian ini, peneliti membatasi dengan hanya menggunakan metode *Average Linkage* berdasarkan penelitian Vicenzo Tola (2008), yang menyatakan bahwa hasil Metode *Average Linkage* dianggap lebih baik dibanding Metode *Single Linkage*, Metode *Complete Linkage*, dan Metode *Ward*. Sementara untuk metode *K-means* dinilai sebagai metode yang sederhana dan umum digunakan. Berdasarkan penelitian Nanda (2010), yang menyatakan hasil yang diperoleh dengan menggunakan Metode *K-means* lebih stabil dibanding dua metode lain yang dibandingkan yaitu *Self Organizing Maps* dan *Fuzzy C-means*.

Berdasarkan uraian diatas, penelitian ini dimaksudkan membandingkan metode *Average Linkage* dengan *K-means* pada Industri Kecil Menengah Kabupaten Wajo untuk melihat metode mana yang sesuai berdasarkan kriteria-kriteria yang telah ditentukan oleh pemerintah.

2. TINJAUAN PUSTAKA

Analisis *Cluster* adalah teknik multivariat yang tujuan utamanya adalah untuk mengelompokkan objek berdasarkan karakteristiknya. Dalam proses *clustering* dapat digunakan dua metode untuk prosesnya, yaitu Metode *Hierarki* dan Metode *Non-hierarki*.

Metode *Hierarki* adalah metode yang pengelompokannya dimulai dengan dua atau lebih amatan yang memiliki kesamaan terdekat. Kelebihan dari metode *hierarki* adalah cepat dalam pemrosesan sehingga dapat meminimalkan waktu dalam proses. Namun kelemahan dari metode ini adalah dapat menimbulkan kesalahan, selain itu tidak baik diterapkan untuk menganalisis sampel yang besar. Tipe dasar dalam metode ini adalah *Agglomerative* dan *Devisive*. Jenis-jenis observasi pada metode *Agglomerative* pada awalnya dianggap sebagai *cluster* yang terpisah, sehingga jumlah *cluster* yang diperoleh sama dengan jumlah observasi. Kemudian menggabungkan dua *cluster* dengan kesamaan terdekat menjadi *cluster* baru, mengurangi jumlah *cluster* sebanyak satu pada setiap langkah. Sebaliknya,

Metode *Devisive* dimulai dengan *cluster* besar yang merupakan hasil dari semua pengamatan. Ini kemudian memisahkan pengamatan yang lebih berbeda untuk membentuk *cluster* yang lebih kecil. Proses ini berlanjut sampai setiap pengamatan merupakan *cluster* yang terpisah menurut persamaan karakteristik. Yang penting dalam metode *Hierarchy* adalah hasil dari langkah sebelumnya selalu disertakan dalam hasil langkah selanjutnya membentuk pohon.

K-means adalah metode analisis data atau metode data mining yang melakukan proses pemodelan tanpa pengawasan, dan merupakan metode pengelompokan data ke dalam sistem partisi. Metode *K-means* mencoba untuk mengelompokkan data yang ada menjadi beberapa kelompok. Di sini, data dalam satu kelompok memiliki karakteristik yang sama satu sama lain dan karakteristik yang berbeda dari data pada kelompok lainnya. Dengan kata lain, metode ini mencoba meminimalkan varians antara data dalam satu *cluster* dan memaksimalkan varians dengan data di *cluster* lainnya

3. METODOLOGI

Data Penelitian

Data penelitian yang digunakan adalah data tahun 2017 industri kecil menengah di Kabupaten Wajo dengan jumlah data sebanyak 30 data yang diperoleh secara sekunder di Dinas Perindustrian Provinsi Sulawesi Selatan. Penelitian ini dilakukan mulai 15 Agustus – 20 Oktober 2021.

Prosedur Analisis

Langkah-langkah analisis dalam penelitian ini adalah sebagai berikut

1. Mendesain Penelitian dalam Analisis *Cluster*
2. Penentuan Asumsi-asumsi dalam Analisis *Cluster*
3. Menentukan Ukuran Kemiripan Objek Menggunakan jarak *Euclidean*.
4. Analisis *Clustering*
5. Interpretasi *Cluster*

4. PEMBAHASAN

Data

Dalam penelitian ini data yang digunakan adalah data Industri Kecil dan Menengah (IKM) kabupaten wajo 2017. Data sekunder merupakan jenis data yang digunakan dan diambil dari Dinas Perindustrian Sulawesi Selatan. Untuk melakukan pengelompokan industri kecil dan industri menengah, penelitian ini menggunakan indikator yang merupakan variabel bebas yaitu Nilai Investasi (X_1) dan Total Tenaga Kerja (X_2) dan untuk mengolahnya, penelitian ini menggunakan Program R.

Standarisasi Data

Standarisasi dilakukan jika format data yang dimiliki tidak seragam antar variabel yang satu dengan variabel yang lainnya. Dan ini terjadi pada penelitian yang akan diteliti. Indikator Nilai Investasi (X_1) menggunakan satuan mata uang sedangkan untuk indikator Tenaga Kerja (X_2) menggunakan satuan per-orang. Untuk itu perlunya dilakukan standarisasi.

Tabel 4.1 Hasil Standarisasi Data

USAHA INDUSTRI	(X_1)	(X_2)
Industri Produk Roti Dan Kue	-0,09	2,02
Industri Produk Makanan lainnya	-0,20	-0,34
Industri Gula Merah	-0,23	-0,63
Industri Tempe Kedelai	-0,22	-0,59
Industri Pengolahan Lainnya YTDL	-0,21	-0,53

Outlier

Data standar yang kita peroleh akan menjadi tolak ukur untuk menentukan data apa saja yang outlier. Data dikatakan outlier jika nilai z tidak berada di rentang ± 2.5 . Sehingga data-data yang kami anggap outlier di tampilkan pada tabel dibawah ini.

Tabel 4.2 Data Outlier

USAHA INDUSTRI	Variabel	Nilai
Industri Penggilingan Padi dan Penyosohan Beras	Z_{X_1}	5,28
Industri Penjahitan dan Pembuatan Pakaian sesuai Pesanan	Z_{X_2}	2,89

Terdapat 2 data outlier pada data penelitian ini yaitu anggota variabel X_1 pada industri penggilingan padi dan penyosohan beras dan anggota variabel X_2 pada industri penjahitan dan pembuatan pakaian sesuai pesanan. Pada kasus ini peneliti memilih mempertahankan data outlier, karena data outlier pada penelitian ini merupakan data yang sebenarnya atau adanya sesuai dengan apa yang didapat dari sumber, pertimbangan lainnya agar kedua metode yang digunakan memiliki hasil yang dapat dibandingkan. Ini karena jika data outlier dikeluarkan, hasil yang diperoleh dari kedua metode adalah sama atau dengan kata lain hasil yang diperoleh tidak dapat dibandingkan.

Uji Kecukupan Sampel

Penelitian ini menggunakan *Kaiser Mayer Olkin (KMO)* sebagai tolak ukur untuk menguji kecukupan sampel. Dengan menggunakan program R, sampel penelitian ini mendapat skor KMO sebesar 0.5 yang artinya data layak untuk diteliti karena nilai KMO pada sampel tidak kurang dari 0.

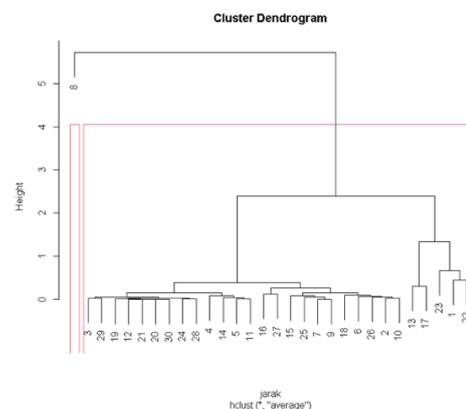
Uji Multikolinearitas

Selanjutnya melihat apakah terjadi multikolinearitas di sampel penelitian. Nilai korelasi menjadi tolak ukurnya. Nilai korelasi antar variabel yang diperoleh adalah 0,29. Yang artinya tidak terjadi kasus multikolinearitas pada data penelitian ini.

Analisis Cluster

Average Linkage

Dibawah ini adalah dendrogram hasil dari pada pengelompokan metode *Average Linkage* dengan menggunakan program R.

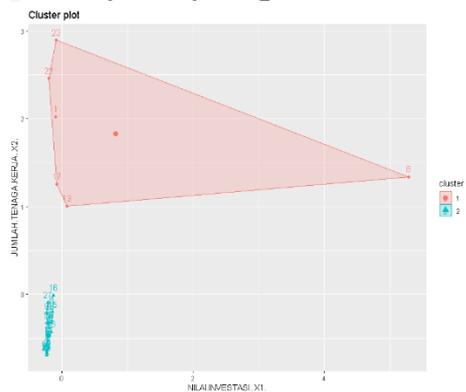


Gambar 4.1 Dendrogram Average Linkage

Berdasarkan tabel dendrogram diatas, pengelompokan dengan menggunakan Metode *Average Linkage* membentuk 2 *cluster*, yang dimana untuk cluster pertama memiliki 1 anggota dan cluster lainnya memiliki 29 anggota.

K-means

Selanjutnya adalah metode *K-means*. Dengan menggunakan program *R*, hasilnya ditampilkan dalam bentuk plot seperti gambar dibawah ini.



Gambar 4.2 Plot *K-means*

Pada Metode *K-means* juga terbentuk 2 *cluster*, dimana kedua *cluster* dibedakan berdasarkan warna plotnya yakni plot merah untuk *Cluster 1* dan plot biru untuk *Cluster 2*, bisa dilihat pula *Cluster 1* memiliki 6 anggota, sedangkan *Cluster 2* memiliki 24 anggota.

Menentukan Anggota Setiap Cluster

Dibawah ini merupakan cluster beserta anggota-anggota yang termasuk dalam cluster tersebut.

Tabel 4.3 Anggota Cluster *Average Linkage*

Cluster	Anggota
Cluster 1	Industri Produk Makanan lainnya, Industri Gula Merah, Industri Tempe Kedelai, Industri Pengolahan Lainnya YTDL, Industri Air Minum dan Air Mineral, Industri Penggaraman/Pengeringan Ikan, Industri Penggilingan Tepung Beras dan Tepung Jagung, Industri Barang Dari Semen, Industri Batu Bata dari Tanah Liat / Keramik, Industri Sabun dan Bahan Pembersih dan Keperluan

Rumah Tangga, Jasa Industri untuk Berbagai Pengerjaan, Industri Barang dari Logam bukan Khusus Logam dan Barang dari Logam, Reparasi Produk Logam Pabrikasi Lainnya, Reparasi Mobil, Industri Furnitur dari Logam, Industri Alat Potong dan Perkakas Tangan untuk Pertanian, Industri Oven, Perapian dan Tungku Pembakar, Industri Kain Sulaman/Bordir, Industri Furnitur Lainnya, Industri Percetakan Umum, Industri Kerajinan YTDL, Reparasi Furnitur dan Perlengkapan Rumah, Reparasi Peralatan Komunikasi dan Reparasi Alat-Alat Elektronik Konsumen, Industri Produk Roti Dan Kue, Industri Wadah dari Kayu, Reparasi dan Perawatan Sepeda Motor, Industri Kain Tenun Ikat, Penjahitan dan Industri Pembuatan Pakaian sesuai Pesanan.

Cluster 2 Industri Penggilingan Padi dan Penyosohan Beras

Tabel 4.4 Anggota Cluster *K-means*

Cluster	Usaha Industri
Cluster 1	Industri Produk Makanan lainnya, Industri Gula Merah, Industri Tempe Kedelai, Industri Pengolahan Lainnya YTDL, Industri Air Minum dan Air Mineral, Industri Penggaraman/Pengeringan Ikan, Industri Penggilingan Tepung Beras dan Tepung Jagung, Industri Barang Dari Semen, Industri Batu Bata dari Tanah Liat / Keramik, Industri Sabun dan Bahan

	Pembersih dan Keperluan Rumah Tangga, Jasa Industri untuk Berbagai Pengerjaan, Industri Barang dari Logam bukan Khusus Logam dan Barang dari Logam, Reparasi Produk Logam Pabrikasi Lainnya, Reparasi Mobil, Industri Furnitur dari Logam, Industri Alat Potong dan Perkakas Tangan untuk Pertanian, Industri Oven, Perapian dan Tungku Pembakar, Industri Kain Sulaman/Bordir, Industri Furnitur Lainnya, Industri Percetakan Umum, Industri Kerajinan YTDL, Reparasi Furnitur dan Perlengkapan Rumah, Reparasi Peralatan Komunikasi, dan Reparasi Alat-Alat Elektronik Konsumen
Cluster 2	Industri Produk Roti Dan Kue, Industri Penggilingan Padi dan Penyosohan Beras, Industri Wadah dari Kayu, Reparasi dan Perawatan Sepeda Motor, Industri Kain Tenun Ikat, Penjahitan dan Industri Pembuatan Pakaian sesuai Pesanan

Interpretasi

Berikut adalah nilai *centroid* setiap *cluster* dengan menggunakan metode *Average Linkage* dan *K-means*.

Tabel 4.5 Nilai *Centroid Average Linkage*

Variabel	X_1	X_2
Cluster 1	319.452.000	32,2
Cluster 2	36.625.000.000	99

Tabel 4.6 Nilai *Centroid K-means*

Variabel	X_1	X_2
----------	-------	-------

Cluster 1	6.977.036.000	123
Cluster 2	167.788.000	12,3

Berdasarkan tabel diatas, interpretasi yang diperoleh adalah sebagai berikut.

Metode *Average Linkage*

Nilai *centroid* variabel Nilai Investasi (X_1) dan varabel Tenaga Kerja (X_2) pada *Cluster 1* lebih kecil dibanding *Cluster 2*. Hal ini menjelaskan bahwa *Cluster 1* adalah Industri Kecil. Untuk *Cluster 2*, nilai *centroid* pada kedua variabel lebih besar dibanding nilai *centroid* kedua variabel pada *Cluster 1*. Hal ini menjelaskan bahwa *Cluster 2* adalah Industri Menengah

Metode *K-means*

Nilai *centroid* variabel Nilai Investasi (X_1) dan varabel Tenaga Kerja (X_2) pada *Cluster 1* lebih besar dibanding *Cluster 2*. Hal ini menjelaskan bahwa *Cluster 1* adalah Industri Menengah. Untuk *Cluster 2*, nilai *centroid* pada kedua variabel lebih kecil dibanding nilai *centroid* kedua variabel pada *Cluster 1*. Hal ini menjelaskan bahwa *Cluster 2* adalah Industri Kecil

5. KESIMPULAN

Berdasarkan hasil pembahasan pada bab sebelumnya, diperoleh hasil untuk penggunaan metode *Average Linkage*, 1 industri di kategorikan sebagai anggota dari Industri Menengah, sementara 29 lainnya dikategorikan sebagai anggota Industri Kecil. Selanjutnya, hasil yang diperoleh dengan metode *K-means*, sebanyak 6 industri di kategorikan sebagai anggota dari Industri Menengah, sementara 24 lainnya dikategorikan sebagai anggota Industri Kecil.

Berdasarkan nilai *centroidnya*, metode *K-means* dinilai lebih cocok untuk merepresentasikan pengelompokan Industri Kecil dan Menengah.

6. DAFTAR PUSTAKA

- [1] Alwi, Wahidah, & Muh Hasrul, 2018. Analisis Klaster Untuk Pengelompokan Kabupaten/Kota Di Provinsi Sulawesi

- Selatan Berdasarkan Indikator Kesejahteraan Rakyat. *Jurnal MSA (Matematika Dan Statistika Serta Aplikasinya)*. Vol. 6, no. 1, 2018, p. 35.
- [2] Badan Pusat Statistik (BPS), 2017. Kabupaten Wajo Dalam Angka Tahun 2017. Badan Pusat Statistik. Wajo.
- [3] Chaturvedi, A.D., Green, P.E. and Carroll, J.D, 2001. K-Modes Clustering. *Journal of Classification*. Vol.18. pp. 35-56.
- [4] Gunawan, Imam, 2017. “*Pengantar Statistika Inferensial*”. PT. Rajagrafindo Persada. Jakarta h. 340 & 333-334.
- [5] Handoyo, Rendy dkk., 2014. “*Perbandingan Metode lustering Menggunakan Metode Single Linkage dan K-Means pada Pengelompokan Dokumen*”. Fakultas Teknik Elektro Universitas Telkom. Bandung.
- [6] Nanda S.R, Mahanty B., Tiwari M.K., 2010. Clustering Indian stock market data for portfolio management. *Expert Systems with Applications*. Vol.37 No. 12, pp. 8793.
- [7] Nafisah, Qonitatin & Novita Eka C., 2017. “*Analisis Cluster Average Linkage Berdasarkan Faktor-Faktor Kemiskinan di Provinsi Jawa Timur*”. Jurusan Matematika Fakultas MIPA Universitas Islam Darul Ulum Lamongan. Lamongan. h. 32
- [8] Tola, Vicenzo, Lilo, Fabrizio., Gallegati, Mario., Mategna, & Rosario N., 2008. Cluster Analysis for Portfolio Optimization. *Journal of Economics Dynamics and Control*. Vol. 32, issue 1, pp. 235-258.
- [9] Yuliato, Safa’at & Kishera Hilya Hidayatullah, 2014. “*Analisis Klaster Untuk Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Indikator Kesejahteraan Rakyat*”. Akd. Statistika Muhammadiyah Semarang. Semarang.