VARIANSI: Journal of Statistics and Its Application on Teaching and Research **ISSN 2684-7590 (Online)**
Vol. 6 No. 3 (2024), 144-151

DOI: *10.35580/variansiunm20*

Algoritma K-Prototype dalam Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat Tahun 2020

Zulkifli Rais, Suwardi Annas*, Muhammad Refaldy

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Makassar, Indonesia

Keywords: Algoritma K-Prototype, Cluster, Metode *Elbow*, Indikator Kesejahteraan Rakyat

Abstract:

Clustering is something that is used to analyze data both in machine learning, data mining, pattern engineering, image analysis, and bioinformatics. To produce the information needed for a data analysis using the clustering process, this is because the data has a large variety and amount. Researchers will use the K-Prototype method, where this method becomes an efficient and effective algorithm in processing mixed-type data. The K-Prototype algorithm has problems in finding the best number of clusters. So, in this paper, researchers will conduct research by finding the best number of clusters in the K-Prototype method. There are many ways to determine this, one of which is the elbow method. The determination of this method is seen from the SSE (Sum Square Error) graph of several numbers of clusters. The results of the clustering formed 2 clusters, which were considered optimal based on the value of k that experienced the greatest decrease. The results showed that cluster 1 is a cluster that has characteristics of people's welfare, which is better than cluster 2.

1. Pendahuluan

Analisis *cluster* adalah salah satu dari metode dalam analisis *multivariat* yang memiliki tujuan utama untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Analisis *cluster* mengelompokkan individu atau objek penelitian, sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam *cluster* yang sama. *Cluster-cluster* yang terbentuk dalam satu cluster mempunyai ciri yang relatif sama *(homogen)*, sedangkan antar *cluster* mempunyari ciri yang berbeda *(heterogen)*. Pengelompokan ini dilakukan berdasarkan variabel-variabel yang diamati (Usman dan Sobari, 2013).

Menurut Mattjik & Sumertajaya (2011) algoritma *clustering* terbagi ke dalam dua jenis, yaitu metode *hierarki* dan *non-hierarki*. Algoritma *clustering hierarki* digunakan untuk mengelompokkan objek secara terstruktur berdasarkan kemiripan sifatnya dan *cluster* yang diinginkan belum diketahui banyaknya. Menurut Johnson & Wichern (2002) algoritma *clustering non-hierarki* digunakan untuk pengelompokan objek dimana banyaknya *cluster* yang akan dibentuk dapat ditentukan terlebih dahulu sebagai bagian dari prosedur pengelompokan.

Algoritma *K-Prototype* adalah salah satu metode *Clustering* yang berbasis *partitioning.* Algoritma ini adalah hasil pengembangan dari algoritma *K-Means* (Huang,1998) untuk menangani *clustering* pada data dengan atribut bertipe campuran numerik dan kategorikal. Pengembangan yang dilakukan oleh Huang mempertahankan efisiensi algoritma

* Corresponding author.
E-mail address: suwardi_annas@unm.ac.id

K-Means dalam menghadapi data berukuran besar dan dapat diterapkan pada data numerik dan kategorikal. Pengembangan yang mendasar pada algoritma *k-Prototype* terdapat pada pengukuran kesamaan *(similarity measure)* antara object dengan *centroid (prototype)*-nya. Berdasarkan pembahasan diatas maka peneliti menerapkan algoritma *k-prototype* karena adanya variabel kategorik dan numerik dalam indikator kesejahteraan rakyat untuk melakukan pengelompokkan kabupaten/kota di Provinsi Sulawesi Selatan.

2. Tinjauan Pustaka

2.1. Definisi *Clustering*

Clustering merupakan salah satu metode data mining yang bersifat tanpa arahan *(unsupervised)*, maksudnya metode ini diterapkan tanpa adanya latihan *(training)* dan tanpa ada guru *(teacher)* serta tidak memerlukan target output. Dimana sebuah *cluster* terdiri dari kumpulan benda-benda yang mirip antara satu dengan yang lainnya dan berbeda dengan benda yang terdapat pada *cluster* lainnya. Algoritma *clustering* terdiri dari dua bagian yaitu secara *hierarki* dan secara *non-hierarki*. Algoritma *hierarki* menemukan *cluster* secara berurutan dimana *cluster* ditetapkan sebelumnya, sedangkan algoritma *non-hierarki* menentukan semua kelompok pada waktu tertentu (Madhulatha, 2012). *Clustering* juga bisa dikatakan suatu proses dimana mengelompokan dan membagi pola data menjadi beberapa jumlah dataset sehingga akan membentuk pola yang serupa dan dikelompokan pada *cluster* yang sama dan memisahkan diri dengan membentuk pola yang berbeda ke *cluster* yang berbeda (Huang dkk., 2005). Analisis *cluster* mengklasifikasi objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam *cluster* yang sama. *Cluster-cluster* yang terbentuk memiliki homogenitas internal yang tinggi dan heterogenitas eksternal yang tinggi (Prasetyo, 2014).

2.2. Metode *Elbow*

Metode *Elbow* merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah *cluster* terbaik dengan cara melihat persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik. Metode ini memberikan ide/gagasan dengan cara memilih nilai *cluster* dan kemudian menambah nilai *cluster* tersebut untuk dijadikan model data dalam penentuan *cluster* terbaik. Dan selain itu persentase perhitungan yang dihasilkan menjadi pembanding antara jumlah *cluster* yang ditambah. Hasil persentase yang berbeda dari setiap nilai *cluster* dapat ditunjukan dengan menggunakan grafik sebagai sumber informasinya. Jika nilai *cluster* pertama dengan nilai *cluster* kedua memberikan sudut dalam grafik atau nilainya mengalami penurunan paling besar maka nilai *cluster* tersebut yang terbaik.Untuk mendapatkan perbandingannya adalah dengan menghitung SSE *(Sum of Square Error)* dari masing-masing nilai *cluster*. Karena semakin besar jumlah *cluster* k maka nilai SSE akan semakin kecil. Rumus SSE pada *k-prototype*

$$SSE = \sum_{K=l}^{K} \sum_{x_l \in S_K} \|X_i - C_k\|_2^2 \quad (1)$$

2.3. *Similarity Measure*

Bentuk umum ukuran kesamaan dinyatakan sebagai berikut

$$d(X_i\ ,\ Z_l) = \sum_{j=1}^{m} \delta(x_{ij}\ ,\ z_{lj}\ ) \quad (2)$$

$z_l = [z_{i1}, z_{i2}, \dots\ , z_{im}]^{\ T}$ adalah *prototype* untuk *cluster* l. Ukuran kesamaan untuk atribut numerik dikenal dengan jarak *euclidean* ditunjukkan dalam persamaan (3) berikut ini

$$d(X_i\ ,\ Z_l) = (\sum_{j=1}^{m_r} (x_{ij}^r - z_{lj}^r)^2)^{½} \quad (3)$$

x_{ij}^r adalah nilai pada atribut numerik j

z_{lj}^r adalah rata-rata atau *prototype* atribut numerik ke *j cluster l.*

m_r adalah jumlah atribut numerik.

Sedangkan ukuran kesamaan untuk data kategorikal adalah

$$d(X_i\ ,\ Z_l) = \gamma_l \sum_{j=l+1}^{m_c} \delta(x_{ij}^c - z_{lj}^c) \quad (4)$$

Dimana *simple matching similarity measure* untuk data kategorikal adalah

$$\delta(x_{ij}^c - z_{lj}^c) = \begin{cases} 0 & (x_{ij}^c = z_{lj}^c) \\ 1 & (x_{ij}^c \neq z_{lj}^c) \end{cases} \quad (5)$$

γ_l adalah bobot untuk atribut kategori pada *cluster* l yang nilainya merupakan nilai standar deviasi untuk atribut numerik pada masing-masing *cluster*

x_{ij}^c adalah nilai atribut kategorikal

z_{lj}^c adalah modus atribut ke j *cluster* l

m_c adalah jumlah atribut kategorikal.

He, memodifikasi *simple matching similarity measure* menjadi persamaan (6) untuk meningkatkan kemiripan objek dalam *cluster* dengan atribut kategorikal sehingga hasil pengclusteran menjadi lebih baik.

Jika

$$\delta(x_{ij}^c - z_{lj}^c) = \begin{cases} 1 - \omega(x_{ij}^c, l) & (x_{ij}^c = z_{lj}^c) \\ 1 & (x_{ij}^c \neq z_{lj}^c) \end{cases} \quad (6)$$

$\omega(x_{ij}^c, l)$ adalah nilai penimbang untuk x_{ij}^cdimana nilai $\omega(x_{ij}^c, l)$ adalah

$$\omega(x_{ij}^c, l) = \frac{f(x_{ij}^c|c_l)}{|c_l| f(x_{ij}^c|D)} \quad (7)$$

$f(x_{ij}^c|c_l)$ adalah frekuensi nilai x_{ij}^c dalam *cluster* l

$|c_l|$ adalah jumlah objek dalam *cluster* l

$f(x_{ij}^c|D)$ adalah frekuensi nilai x_{ij}^c pada keseluruhan dataset.

2.4. Algoritma *K-Prototype*

Algoritma *K-Prototype* adalah salah satu metode *clustering* yang berbasis *partitioning.* Algoritma ini adalah hasil pengembangan dari algoritma *k-means* (Huang,1998) untuk menangani *clustering* pada data dengan atribut bertipe campuran numerik dan kategorikal. Pengembangan yang dilakukan oleh Huang mempertahankan efisiensi algoritma *k-means* dalam menghadapi data berukuran besar dan dapat diterapkan pada data numerik dan kategorikal. Pengembangan yang mendasar pada algoritma *k-prototype* terdapat pada pengukuran kesamaan *(similarity measure)* antara object dengan *centroid (prototype)*-nya. Algoritma *k-prototype* data numerik dihitung dengan jarak *euclidean* sedangkan data kategori dihitung menggunakan jarak *k-modes*.

$$d(X_i, Z_l) = (\sum_{j=1}^{m_r}(x_{ij}^r - z_{lj}^r)^2 + \gamma_l \sum_{j=l+1}^{m_c} \delta(x_{ij}^c - z_{lj}^c))^{½} \quad (8)$$

Ada empat tahapan utama pada algoritma *K-Prototypes*. Sebelum menuju pada proses algoritma *k-prototypes* terlebih dahulu tentukan jumlah *cluster* k yang akan dibentuk batasannya minimal 2 dan maksimal √n atau n/2 dimana n adalah jumlah data *point*. Tahapan Algoritma *K-Prototype* adalah :

Tahap 1 : Tentukan k dengan inisial *cluster*

Tahap 2 : Hitung jarak seluruh data point pada dataset terhadap inisial *cluster* awal, alokasikan data point ke dalam *cluster* yang memiliki jarak *prototype* terdekat dengan objek yang diukur.

Tahap 3 : Hitung titik pusat *cluster* yang baru setelah semua objek dialokasikan. Lalu realokasikan semua data *point* pada dataset terhadap *prototype* yang baru.

Tahap 4 : Jika titik pusat *cluster* tidak berubah atau sudah konvergen maka proses algoritma berhenti tetapi jika titik pusat masih berubah-ubah secara signifikan maka proses kembali ke tahap 2 dan 3 hingga iterasi maksimum tercapai atau sudah tidak ada perpindahan objek.

2.5. Indikator Kesejahteraan Rakyat

Menurut kamus W.J.S Poerwadarminta (1990), sejahtera diartikan sebagai keadaan “aman, sentosa, dan makmur”. Sehingga arti kesejahteraan meliputi kemanan, keselamatan dan kemakmuran. Adapun istilah rakyat (sosial) dalam arti sempit berkait dengan sektor pembangunan sosial atau pembangunan kesejahteraan rakyat yang bertujuan untuk meningkatkan kualitas kehidupan manusia, terutama yang dikatagorikan sebagai kelompok yang tidak beruntung dan kelompok rentan (kelompok yang berpotensi untuk menjadi orang miskin). Dalam hal ini, kebijakan pembangunan kesejahteraan rakyat pada umumnya menyangkut program-program atau pelayanan-pelayanan sosial untuk mengatasi masalah-masalah sosial seperti, kemiskinan, keterlantaran, ketidakberfungsian fisik dan psikis, tuna sosial, tuna susila, dan kenakalan remaja. Sebagai konsekuesinya, pengertian kebijakan kesejahteran rakyat seringkali diartikan sebagail kegiatan amal atau bantuan publik yang dilakukan pemerintah bagi keluarga miskin dan anak-anak mereka; yang oleh

para pakar ilmu sosial dihubungkan dengan kondisi "Indeks Pembangunan Manusia/*Human Development Index*", yaitu: tinggi rendahnya tingkat hidup masyarakat yang dilihat dari tiga indikator utama: tingkat harapan hidup (expectation of life), tingkat pendidikan *(literacy education)*, dan tingkat pendapatan *(income)*.

Walaupun tidak ada batasan-batasan yang tegas tentang kesejahteraan, namun tingkat kesejahteraan berupa pangan, pendidikan, kesehatan, dan seringkali dikaitkan dengan perlindungan sosial lainya seperti kesempatan kerja, perlindungan hari tua, keterbebasan dari kemiskinan, dan sebagainya. Indikator yang digunakan untuk mengetahui tingkat kesejahteraan ada sepuluh, yaitu umur, jumlah tangungan, pendapatan, konsumsi atau pengeluaran kelaurga, keadaan tempat tinggal, fasilitas tempat tinggal, kesehatan anggota keluarga, kemudahan mendapatkan pelayanan kesehatan, kemudahan memasukkan anak ke jenjang pendidikan dan kemudahan mendapatkan fasilitas.

3. Metode Penelitian

3.1. Jenis Penelitian

Penelitian ini menggunakan metode penelitian deskriptif kualitatif yang bisa menjelasakan variabel indikator yang digunakan, serta dengan pendekatan atau eksplorasi data indikator kesejahteraan rakyat.

3.2. Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data sekunder yaitu data yang didapatkan secara tidak langsung atau diperoleh dari sumber yang sudah ada. Data tersebut diambil dari Badan Pusat Statistik Provinsi Sulawesi Selatan.

3.3. Teknik Analisis Data

Adapun teknik analisis data pada penelitian ini yaitu :

1. Melakukan analisis deskriptif untuk melihat karakteristik variabel-variabel indikator kesejahteraan rakyat setiap Kabupaten/Kota di Provinsi Sulawesi Selatan
2. Menentukan *cluster* optimum dengan menggunakan metode elbow
3. Menentukan pusat *cluster* awal
4. Menghitung jarak antara objek dengan pusat *cluster* menggunakan jarak *eucledian* pada varibel numerik dan jarak *k-modes* pada variabel kategori
5. Memperbarui pusat *cluster* setelah jarak antar *cluster* dihitung ulang perbedaannya
6. Jika pusat *cluster* berubah maka ulangi dari langkah 4, jika tidak maka akan dilanjutkan pada proses berikutnya
7. Menentukan *output* hasil pengelompokan dan mendeskripsikan masing-masing kelompok berdasarkan karakteristik variabelnya

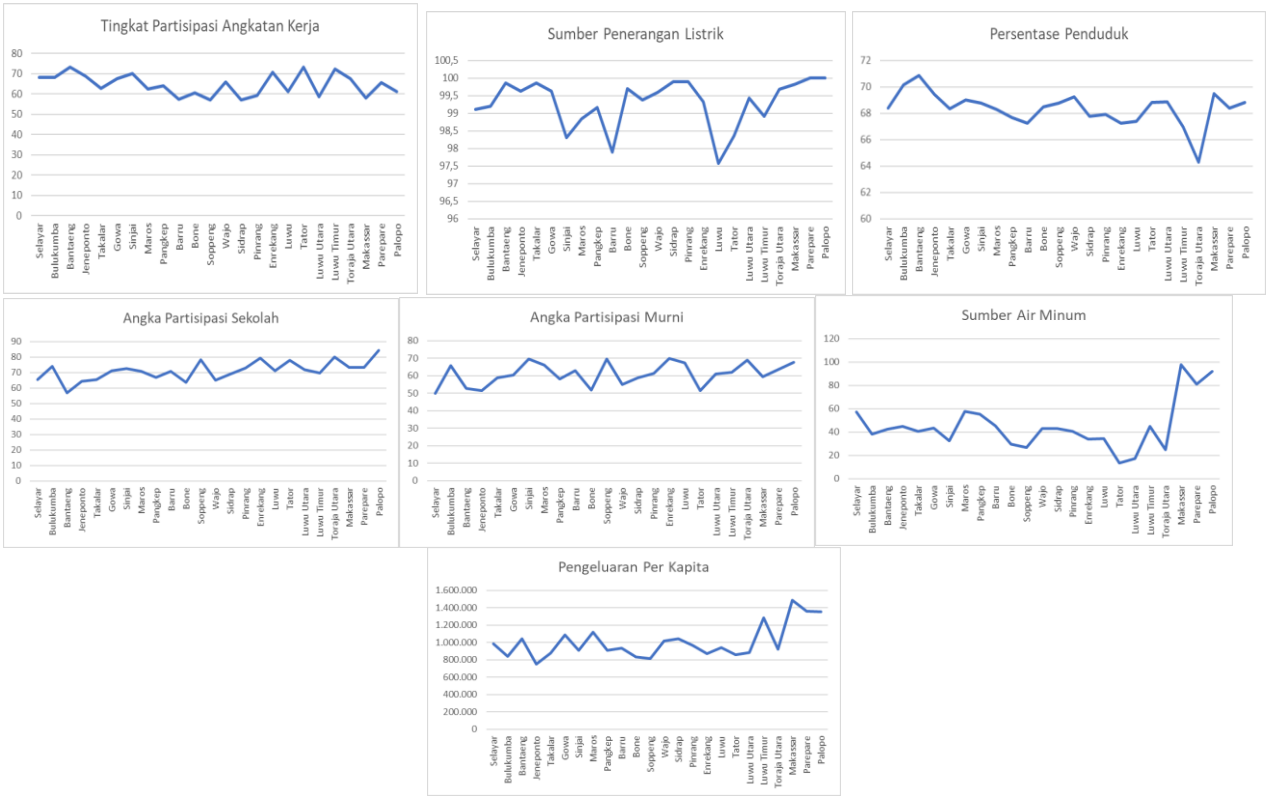
4. Hasil dan Pembahasan

4.1. Analisis Deskriptif

Berdasarkan variabel penelitian terdapat 24 sampel yang terdiri dari Kabupaten/Kota di Provinsi Sulawesi Selatan dengan menggunakan indikator kesejahteraan rakyat tahun 2020. Karakteristik data dibuat dalam bentuk deskriptif untuk menjelaskan setiap variabel. Adapun karakteristik data dapat dilihat pada Tabel 1. sebagai berikut.

Tabel 1. Karakteristik Data

Variabel	Minimum	Maksimum	Rata-rata	Standar Deviasi
Tingkat Partisipasi Angkatan Kerja (%)	56,92	73,25	64,59	5,38
Sumber Penerangan Listrik (%)	97,57	100,00	99,29	0,67
Persentase Penduduk (%)	64,32	70,86	68,36	1,26
Angka Partisipasi Sekolah (%)	56,85	84,31	71,14	6,16
Angka Partisipasi Murni (%)	49,87	69,80	60,94	6,39
Sumber Air Minum (%)	13,55	97,71	44,99	20,78
Pengeluaran Per Kapita (Rupiah)	752.620	1.489.084	1.003.736	191.877



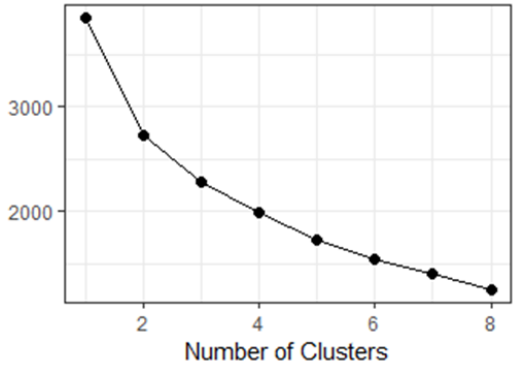
Gambar 1. Diagram Garis Masing-Masing Variabel

Rata-rata tingkat partisipasi angkatan kerja dari Kabupaten/Kota di Provinsi Sulawesi Selatan pada tahun 2020 sebesar 64,59% dengan keragaman 5,38%. Tingkat partisipasi angkatan kerja yang paling tinggi berada di Kabupaten Tanah Toraja yaitu sebesar 73,25% sedangkan yang paling rendah di Kabupaten Sidrap sebesar 56,92%. Rata-rata sumber penerangan listrik dari Kabupaten/Kota di Provinsi Sulawesi Selatan pada tahun 2020 sebesar 99,29% dengan keragaman 0,67%. Sumber penerangan listrik yang paling tinggi berada di Kota Palopo dan Parepare yaitu sebesar 100% sedangkan yang paling rendah di Kabupaten Luwu sebesar 97,57%. Rata-rata persentase penduduk dari Kabupaten/Kota di Provinsi Sulawesi Selatan pada tahun 2020 sebesar 68,36% dengan keragaman 1,26%. Persentase penduduk menurut umur 15-64 yang paling tinggi berada di Kabupaten Bantaeng yaitu sebesar 70,86% sedangkan yang paling rendah di Toraja Utara sebesar 64,32%. Rata-rata angka partisipasi sekolah dari Kabupaten/Kota di Provinsi Sulawesi Selatan pada tahun 2020 sebesar 71,14% dengan keragaman 6,16%. Angka partisipasi sekolah yang paling tinggi berada di Kota Palopo yaitu sebesar 84,31% sedangkan yang paling rendah di Kabupaten Bantaeng sebesar 56,85%. Rata-rata angka partisipasi murni dari Kabupaten/Kota di Provinsi Sulawesi Selatan pada tahun 2020 sebesar 60,94% dengan keragaman 6,39%. Angka partisipasi murni yang paling tinggi berada di Kabupaten Enrekang yaitu sebesar 69,8% sedangkan yang paling rendah di Kabupaten Selayar sebesar 49,87%. Rata-rata sumber air minum dari Kabupaten/Kota di Provinsi Sulawesi Selatan pada tahun 2020 sebesar 44,99% dengan keragaman 20,78%. Sumber air minum yang paling tinggi berada di Kota Makassar yaitu sebesar 97,71% sedangkan yang paling rendah di Kabupaten Tanah Toraja sebesar 13,55%. Rata-rata pengeluaran per kapita dari Kabupaten/Kota di Provinsi Sulawesi Selatan pada tahun 2020 sebesar Rp. 1.004.736 dengan keragaman Rp. 191.877. Pengeluaran per kapita yang paling tinggi berada di Kota Makassar yaitu sebesar Rp. 1.489.084 sedangkan yang paling rendah di Kabupaten Jeneponto sebesar Rp. 752.620.

4.2. Menentukan banyaknya *cluster* yang terbentuk

Algoritma *k-prototype* merupakan sebuah teknik untuk melakukan pengelompokan *non-hierarki* pada suatu objek. Sebelum proses algoritma dijalankan perlu dilakukan penentuan jumlah *cluster* terlebih dahulu. Menentukan jumlah *cluster* yang berbeda akan menghasilkan kesimpulan atau deskripsi *cluster* yang berbeda pula. Penentuan jumlah *cluster* diperoleh dari cara mengevaluasi perhitungan *elbow method* pada setiap *cluster* yang terbentuk sehingga dapat

diperoleh hasil *clustering* yang *homogen* dalam satu *cluster* dan *heterogen* antar *cluster*. Pemilihan jumlah *cluster* dilakukan secara bertahap dimulai dari jumlah *cluster* sebanyak 1 hingga 8 *cluster*. Hasil penentuan *cluster* optimal dengan menggunakan metode *elbow* dapat dilihat pada Gambar 1 sebagai berikut.



Gambar 2. Plot jumlah *cluster* yang optimal

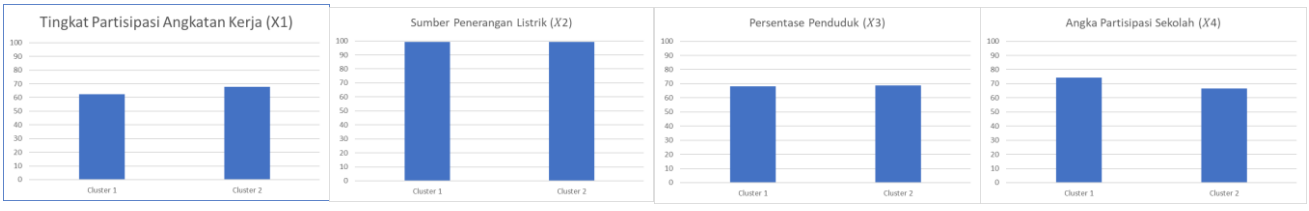
Gambar 2 menunjukkan bahwa ada beberapa nilai k yang mengalami penurunan paling besar dan selanjutnya hasil dari nilai k akan turun secara perlahan-lahan sampai hasil dari nilai k tersebut stabil. Misalnya nilai *cluster* k=1 ke k=2, kemudian dari k=2 ke k=3, terlihat penurunan drastis membentuk siku pada titik k=2 maka nilai *cluster* k yang ideal adalah k=2. Berikut adalah hasil keanggotan Kabupaten/Kota pada setiap *cluster* yang disajikan dalam Tabel 2.

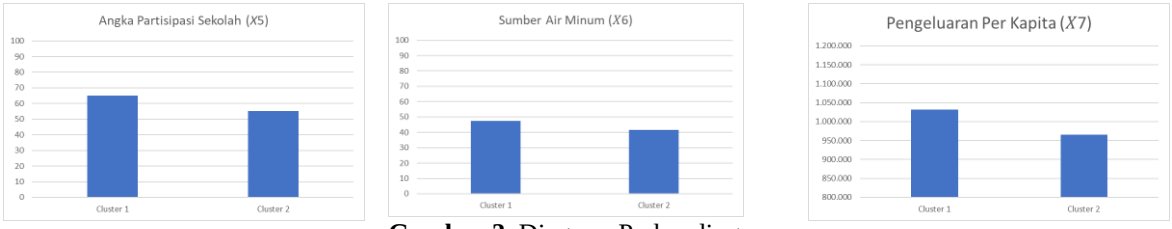
Tabel 2. Anggota-anggota *cluster*

Cluster	Anggota *Cluster*
1	Bulukumba, Sinjai, Maros, Barru, Soppeng, Sidrap, Pinrang, Enrekang, Luwu, Luwu Utara, Toraja Utara, Makassar, Parepare, Palopo
2	Selayar, Bantaeng, Jeneponto, Takalar, Gowa, Pangkep, Bone, Wajo, Tator, Luwu Timur

4.3. Interpretasi hasil *cluster*

Hasil pengolahan variabel menggunakan algoritma *k-prototype* dengan jumlah *cluster* sebanyak 2 dan banyak anggota tiap *cluster* yaitu 14 observasi pada *cluster* pertama dan 10 observasi pada *cluster* kedua. Variabel tingkat partisipasi angkatan kerja menunjukkan bahwa rata-rata *cluste*r 1 sebesar 62,39% lebih kecil dari *cluster* 2 yang memiliki rata-rata tingkat partisipasi angkatan kerja sebesar 67,66%. Variabel sumber penerangan listrik menunjukkan bahwa rata-rata *cluster* 1 sebesar 99,23% lebih kecil dari *cluster* 2 yang memiliki rata-rata sumber penerangan listrik sebesar 99,38%. Variabel persentase penduduk menunjukkan bahwa rata-rata *cluster* 1 sebesar 68,10% lebih kecil dari *cluster* 2 yang memiliki rata-rata persentase penduduk sebesar 68,72%. Variabel angka partisipasi sekolah menunjukkan bahwa rata-rata *cluster* 1 sebesar 74,38% lebih besar dari *cluster* 2 yang memiliki rata-rata angka partisipasi sekolah sebesar 66,61%. Variabel angka partisipasi murni menunjukkan bahwa rata-rata *cluster* 1 sebesar 65,06% lebih besar dari *cluster* 2 yang memiliki rata-rata angka partisipasi murni sebesar 55,19%. Variabel sumber air minum menunjukkan bahwa rata-rata *cluster* 1 sebesar 47,51% lebih besar dari *cluster* 2 yang memiliki rata-rata sumber air minum sebesar 41,46%. Variabel pengeluaran per kapita menunjukkan bahwa rata-rata *cluster* 1 sebesar Rp. 1.031.400 lebih besar dari *cluster* 2 yang memiliki rata-rata pengeluaran per kapita sebesar Rp. 965.007.





Gambar 3. Diagram Perbandingan

Berdasarkan pada diagram batang masing-masing variabel nilai rata-rata untuk semua indikator kesejahteraan rakyat pada *cluster* 1 yang terdiri dari Kabupaten Bulukumba, Kabupaten Sinjai, Maros, Kabupaten Barru, Kabupaten Soppeng, Kabupaten Sidrap, Kabupaten Pinrang, Kabupaten Enrekang, Kabupaten Luwu, Kabupaten Luwu Utara, Kabupaten Toraja Utara, Kota Makassar, Kota Parepare dan Kota Palopo memiliki 4 variabel yang nilai rata-ratanya lebih tinggi dibandingkan *cluster* 2 yaitu pada angka partisipasi sekolah, angka partisipasi murni, sumber air minum dan pengeluaran per kapita. *Cluster* 2 yang terdiri dari Kabupaten Selayar, Kabupaten Bantaeng, Kabupaten Jeneponto, Kabupaten Takalar, Kabupaten Gowa, Kabupaten Pangkep, Kabupaten Bone, Kabupaten Wajo, Kabupaten Tator, Kabupaten Luwu Timur memiliki 3 variabel yang nilai rata-ratanya lebih tinggi dibandingkan *cluster* 1 yaitu pada tingkat partisipasi angkatan kerja, sumber penerangan listrik dan persentase penduduk. Jika diurutkan berdasarkan besar kecilnya nilai rata-rata indikator kesejahteraan rakyat untuk masing-masing *cluster. Cluster* 1 merupakan *cluster* yang memiliki karakteristik kesejahteraan rakyat yang lebih baik dibandingkan *cluster* 2.

5. Kesimpulan

Hasil pengelompokan observasi dengan menggunakan algorima *k-prototypes* menunjukkan jumlah *cluster* yang terbentuk yaitu sebanyak 2 *cluster*. Penentuan *cluster* yang optimum yaitu dengan menggunakan metode *elbow*. Pemilihan *cluster* optimum didasari dengan melihat hasil *sum of square error* dari nilai k yang turun secara drastis. Hasil penelitian menunjukkan bahwa *cluster* 1 memiliki 4 variabel yang nilai rata-ratanya lebih tinggi dibandingkan *cluster* 2 yaitu pada angka partisipasi sekolah, angka partisipasi murni, sumber air minum dan pengeluaran per kapita. *Cluster* 2 memiliki 3 variabel yang nilai rata-ratanya lebih tinggi dibandingkan *cluster* 1 yaitu pada tingkat partisipasi angkatan kerja, sumber penerangan listrik dan persentase penduduk umur 15-64. Berdasarkan besar kecilnya nilai rata-rata indikator kesejahteraan rakyat untuk masing-masing *cluster*, *cluster* 1 merupakan *cluster* yang memiliki karakteristik kesejahteraan rakyat yang lebih baik dibandingkan *cluster* 2.

Daftar Pustaka

Amah, N., Wahyuningsih, S., Deny, F., & Amijaya, T. (2017). Analisis Cluster Non-Hirarki Dengan Menggunakan Metode K-Modes pada Mahasiswa Program Studi Statistika Angkatan 2015 FMIPA Universitas Mulawarman Non-Hierarchical Cluster Analysis Using K-Modes Method. *Eksponensial*, *8*, 9–16.

Annas, S., Rahmat, H. S., & Rais, Z. (2022). *K-Prototypes Algorithm For Clustering The Tectonic Earthquake In Sulawesi Island*. *5*(2), 191–198.

Azizah, R. N. (2013). Sistem Informasi Mengklasifikasi Pemilihan Jurusan Di Perguruan Tinggi Bagi Lulusan Sma Berbasis Web Menggunakan Algoritma K-Mean. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699.

Azwar, Saifuddin.Penyusunan Skala Psikologi / Azwar, Saifuddin .(2017)

Badruttamam, A., Sudarno, S., & Maruddani, D. A. I. (2020). Penerapan Analisis Klaster *K-Modes* dengan Validasi *Davies Bouldin* Index dalam menentukan Karakteristik Kanal Youtube di Indonesia (Studi Kasus: 250 Kanal YouTube Indonesia Teratas Menurut Socialblade). *Jurnal Gaussian*, *9*(3), 263–272. https://doi.org/10.14710/j.gauss.v9i3.28907

Badan Pusat Statistik. (2021) *Indikator Kesejahteraan Rakyat.* BPS: Sulawesi Selatan.

García Reyes, L. E. (2013). Kesejahteraan Masyarakat. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699.

Hair, J. F., Black, W. C., Babin, B. J. & Anderson, R. E., (2014). Multivariate Data Analysis 7th. USA: Pearson.

Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data mining and knowledge discovery*, *2*(3), 283-304.

Jia, Z., & Song, L. (2020). Weighted k-Prototypes Clustering Algorithm Based on the Hybrid Dissimilarity Coefficient. *Mathematical Problems in Engineering*, *2020*. https://doi.org/10.1155/2020/5143797

Johnson, R. A. & Wichern, D. W., (2002). Applied Multivariate Statistical Analysis 5th. New Jersey: Pearson.

Madhulatha, T. S. (2012). An overview on clustering methods. *arXiv preprint arXiv:1205.1117*.

Maiti, & Bidinger. (1981). Analisis Cluster. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699.

Mattjik, A. A. & Sumertajaya, I. M., 2011. Sidik Peubah Ganda dengan Menggunakan SAS. Bogor: IPB Press.

Nahak, M. (2017). Bab Ii Tinjauan Pustaka Dan Landasan Teori. *Journal of Chemical Information and Modeling*, *53*(9), 21–25. http://www.elsevier.com/locate/scp

Nooraeni, R., Supriadi, J., Si, S., & Sc, M. (2019). *K-Prototype Untuk Pengelompokan Data Campuran*. *February*, 1–6.

Nooraeni, R., Tinggi, S., & Statistik, I. (2015). Metode Cluster Menggunakan Kombinasi Algoritma Cluster K-Prototype Dan Algoritma Genetika Untuk Data Bertipe Campuran Cluster Method Using a Combination of Cluster K-Prototype Algorithm and Genetic Algorithm for Mixed Data. *Jurnal Aplikasi Statistika & Komputasi Statistik*, *7*(2), 17–17. https://jurnal.stis.ac.id/index.php/jurnalasks/article/view/23

Poerwadarminta, W. J. S. (1990). Kamus Besar Bahasa Indonesia Balai Pustaka, Jakarta p. 1158. *Go to reference in article*.

Prasetyo, E. (2010).Data Mining dan Aplikasi Menggunakan Matlab. Yogyakarta: Penerbit ANDI.

Rachmatin, D. (2014). Aplikasi Metode-Metode Agglomerative Dalam Analisis Klaster Pada Data Tingkat Polusi Udara. *Infinity Journal*, *3*(2), 133. https://doi.org/10.22460/infinity.v3i2.59

Rachmatin, D., & Sawitri, K. (2016). *Perbandingan Antara Metode Agglomeratif, Metode Divisif dan Metode K-Means Dalam Analisis Klaster*. *1*, 9–17.

Setyawan, A. H., & Pratiwi, N. (2019). Penerapan Metode Two Step Cluster Untuk Pengelompokan Potensi Desa. *Jurnal Statistika Industri Dan …*, *4*(2), 41–51. https://journal.akprind.ac.id/index.php/Statistika/article/view/1923

Sopha, B. M. (2018). *Analisis Klasterisasi Industri Kecil Menengah di Kabupaten Banyuasin, Provinsi Sumatera Selatan dengan Algoritma K-Prototypes ANDIKA YUSUF PUTRA, Bertha Maya Sopha, S.T., M.Sc., Ph.D.*

Sulastri, S., Usman, L., & Syafitri, U. D. (2021). K-prototypes Algorithm for Clustering Schools Based on The Student Admission Data in IPB University. *Indonesian Journal of Statistics and Its Applications*, *5*(2), 228–242. https://doi.org/10.29244/ijsa.v5i2p228-242

Suryono, A. (2018). Kebijakan Publik Untuk Kesejahteraan Rakyat. *Transparansi Jurnal Ilmiah Ilmu Administrasi*, *6*(2), 98–102. https://doi.org/10.31334/trans.v6i2.33

Usman, H., & Sobari, N. (2013). Aplikasi multivariate untuk riset pemasaran. *Jakarta: PT. RajaGrafindo*.