

Penggunaan Analisis Cluster K-Means dan Analisis Diskriminan dalam Pengelompokan Desa Miskin di Kabupaten Pangkep

Febi Fajrianti*, Muhammad Nadjib Bustan, & Muhammad Arif Tiro

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

Keywords: K-Means Analysis, Discriminant Analysis, Poverty

Abstract:

This type of research is a study that wants to group the object by using grouping method k-means cluster as a preliminary grouping of research objects villages based on variables of the lighter. Variables used as indicator of grouping are education level (X1), population density (X2), crude birthrate (X3), crude mortality rate (X4), average number of family members (X5). Data from library unit of BPS and Civil Registry Office of Pangkep Regency of South Sulawesi with research result of villages in Pangkep Regency. The stages of data analysis were performed by preliminary clustering by using Cluster K-Means analysis, from some clusters that will be conducted discriminant analysis to see the accuracy of the grouping. Based on the analysis of discriminant obtained 98,06% (3 clusters), 94,17% (4 clusters), 91,26% (5 clusters), 94,17% (6 clusters), 86,41% (7 clusters), and 88,35% (8 clusters). From the above description is formed by forming clustering with 3 clusters, then the result of a better grouping deed of other clusters fotmed.

1. Pendahuluan

Analisis Multivariat merupakan salah satu metode statistika yang tepat untuk meringkas data dengan dengan peubah banyak dengan berbagai tujuan penelitian yang beragam (Sitepu, dkk., 2011). Salah satu metode analisis multivariat yang telah banyak digunakan dan mudah untuk diinterpretasikan secara statistik adalah analisis cluster. Menurut Webb (2002), Analisis Cluster digunakan untuk eksplorasi data dan untuk menyediakan data untuk digunakan dalam pengclusteran. Analisis Cluster adalah pengelompokan individu dalam suatu populasi untuk menemukan struktur dalam data.

Terkhusus di Sulawesi Selatan, salah satu daerah yang menjadi sorotan fokus penanggulangan kemiskinan yaitu di Kabupaten Pangkep, karena memiliki tingkat persentase kemiskinan yang cukup tinggi dibandingkan dengan kabupaten lainnya. Kebijakan kemiskinan oleh pemerintah Kabupaten Pangkep, selain harus mampu memperkecil jumlah penduduk miskin, juga sekaligus harus bisa mengurangi tingkat kedalaman dan keparahan dari kemiskinan. Pada penulisan Tugas Akhir ini, dilakukan suatu penelitian mengenai pengelompokan seluruh desa di Kabupaten Pangkep berdasarkan kriteria-kriteria kemiskinan, dengan menggunakan metode Analisis Cluster K-Means yang selanjutnya akan dilihat tingkat keakuratan pengelompokan tersebut dengan menggunakan metode analisis diskriminan.

* Corresponding author.

E-mail address: febifajrianti25@gmail.com



2. Tinjauan Pustaka

2.1. Analisis Multivariat

Analisis Multivariat merupakan analisis dimana masalah yang diteliti bersifat multidimensional dan menggunakan tiga atau lebih variabel (Kuncoro, 2003:211). Analisis multivariate dibagi menjadi dua kategori utama yaitu: Pertama metode ketergantungan, Ciri dari analisis ini adalah adanya satu atau beberapa variabel yang berfungsi sebagai dependen dan beberapa variabel lain variabel bebas.

Beberapa metode beserta tujuannya yang termasuk ke dalam golongan metode ini adalah: Analisis regresi Berganda, Analisis Diskriminan Berganda, Analisis Multivariat Varians, Analisis Korelasi Kanonikal. Dan ke dua metode saling ketergantungan, Ciri dari analisis ini adalah bahwa variabel saling berhubungan satu dengan yang lainnya sehingga tidak ada variabel dependen. Beberapa metode beserta tujuannya yang termasuk ke dalam golongan metode ini adalah: Analisis Faktor, Analisis kelompok, Skala Multidimensional.

2.2. Analisis Cluster

Analisis cluster merupakan teknik multivariate yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Proses cluster atau pengelompokan data bias dilakukan dengan dua metode yaitu metode hierarki dan metode non hierarki

2.3. Analisis Cluster K-Means

Metode K-Means pertama kali diperkenalkan oleh MacQueen JB pada tahun 1976. K-Means merupakan salah satu metode pengelompokan data nonhierarki (sekatan) yang beruaha mempartisi data yang ada kedalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama di masukkan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain. Prasetyo (2011).

Menurut (Santosa dalam Ong, 2013:14), langkah-langkah melakukan clustering dengan metode K-Means adalah sebagai berikut:

- a) Pilih jumlah cluster k.
- b) Inisialisasi k pusat cluster ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat cluster diberiduberi nilai awal dengan angka-angka random,
- c) Alokasikan semua data ke centroid terdekat dengan metric jarak yang sudah deitetapkan. Untuk menghiutng jarak semua data ke setiap tiitk pusat cluster dapat menggunakan beberapa teori jarak seperti Euclidean, SquaredEuclidean Distance, Chebychev dan Manhattan/City block.
- d) Hitung kembali centroid C berdasarkan data yang mengikuti cluster masing-masing
- e) Tugaskan lagi setiap objek memakai pusat cluster yang baru. Jika pusat cluster tidak berubah lagi maka proses clustering selesai. Atau, kembali ke langkah nomor 3 sampai pusat cluster tidak berubah lagi

2.4. Analisis Diskriminan

Analisis Diskriminan merupakan metode statistika untuk mengelompokkan sejumlah objek ke dalam sejumlah kelompok berdasarkan beberapa peubah yang membentuk fungsi diskriminan, sedemikian sehingga setiap objek menjadi anggota dari salah satu kelompok. Tujuan analisis diskriminan pada dasarnya untuk mengetahui peubah yang dapat membedakan anggota populasi. Analisis ini dilakukan dengan membentuk kombinasi linear dari sejumlah peubah, sehingga nilai diskriminan dari setiap objek yang terlihat merupakan petunjuk pada objek tersebut untuk menjadi anggota pada satu kelompok tertentu. Apabila banyaknya kelompok yang dilibatkan lebih dari dua, pada

umumnya diperlukan lebih dari satu fungsi diskriminan untuk pegelompokan objek-objek tersebut (Arif, Sukarna, Aswi, 2010).

Menurut Simamora dalam Irwan (2010), analisis Diskriminan terdiri dari lima tahap, yaitu: (1) merumuskan masalah, (2) mengestimasi koefisien fungsi Diskriminan, (3) memastikan signifikansi determinan, (4) menginterpretasi hasil, dan (5) menguji signifikansi analisis Diskriminan.

2.5. Kemiskinan dan Desa Miskin

Miskin adalah suatu keadaan seseorang yang mengalami kekurangan atau tidak mampu memenuhi tingkat hidup yang paling rendah serta tidak mampu mencapai tingkat minimal dari tujuan-tujuan yang telah ditetapkan. Pengertian Desa Miskin adalah daerah Kabupaten yang masyarakat serta wilayahnya relatif kurang berkembang dibandingkan daerah lain dalam skala nasional. Suatu daerah dikategorikan sebagai daerah tertinggal, karena beberapa faktor penyebab, antara lain.

3. Metode Penelitian

Data yang digunakan dalam penelitian adalah data sekunder berupa dokumentasi tertulis dan identifikasi variabel yang digunakan tentang wilayah/desa yang diperoleh dari unit Perpustakaan BPS dan Catatan Sipil Kabupaten Pangkep Sulawesi Selatan pada tahun 2017. Data yang digunakan dalam penelitian ini berupa dokumentasi tertulis dan identifikasi peubah yang digunakan tentang desa/kelurahan di kabupaten Pangkep Sulawesi Selatan Tahun 2017 yaitu: X1 (tingkat pendidikan), X2 (kepadatan penduduk per km²), X3 (angka kelahiran kasar), X4 (angka kematian kasar), X5 (rata-rata banyaknya anggota keluarga).

Langkah-langkah yang dilakukan pada penelitian ini adalah sebagai berikut:

- a) Melakukan pengumpulan data
- b) Melakukan pengujian asumsi terhadap variabel sesuai berdasarkan teori dari metode yang digunakan.
- c) Melakukan analisis statistika deskriptif untuk mengetahui karakteristik kondisi desa/kelurahan miskin.
- d) Melakukan pengclusteran desa/kelurahan miskin dengan menggunakan Analisis Cluster K-Means.
- e) Menghitung ketepatan/keakuratan cluster dengan menggunakan Analisis Diskriminan
- f) Membuat kesimpulan dan saran

4. Hasil dan Pembahasan

4.1. Analisis Deskriptif

Tabel 1. Statistik Deskriptif 103 Desa di Kabupaten Pangkep

	Minimum	Maksimum	Rerata	Simpangan Baku
SD (Banyaknya siswa)	109	1589	516,14	258,12
SMP (Banyaknya Siswa)	14	1395	387,84	263,60
SMA (Banyaknya Siswa)	14	3794	538,25	570,08
Kepadatan Penduduk (/km ²)	0,70	5576	491,73	691,45
Kelahiran (/tahun)	00	12	6,59	2,72
Kematian (tahun)	00	20	2,35	3,62
Anggota Rumah tangga	177	2568	720,90	359,70

4.2. Pembentukan Cluster

Dalam penelitian ini dilakukan peng-cluster-an dengan menggunakan metode analisis cluster K-means, dimana metode pada penelitian ini tidak ditentukan jumlah cluster yang akan dibentuk untuk menentukan pengelompokan awal desa berada dalam kategori miskin.

4.3. Akurasi Cluster Terbentuk

Hasil pengelompokan dari Analisis Cluster K-Means memberikan hasil kelompok yang terbentuk dari penelitian desa miskin dengan beberapa variabel. Akan tetapi pengelompokan analisis cluster k-means tidak memberikan hasil akurasi kelompok. Berdasarkan hasil pengelompokan awal dari Analisis Cluster K-Means maka digunakan analisis diskriminan untuk memperjelas batas-batas pengelompokan.

4.4. Penentuan Cluster Terbaik

Pada bagian ini akan ditentukan cluster yang memiliki akurasi ketepatan pengelompokan terbaik, yang selanjutnya cluster yang memiliki akurasi terbaik akan digunakan sebagai pengelompokan awal desa miskin di Kabupaten Pangkep. Penentuan akurasi cluster ini dengan menggunakan metode analisis diskriminan. Adapun rumus untuk menentukan Akurasi Analisis Diskriminan menurut Dillon (1984):

$$\text{Press's } Q = \frac{[N - (nk)]^2}{N(k-1)}$$

Ket:

N = ukuran sampel

n = jumlah kasus yang diklasifikasikan secara tepat

k = jumlah grup atau kelompok

Tabel 2 Akurasi masing-masing Cluster berdasarkan nilai Q

Cluster	Akurasi %
3 Cluster	98,06
4 Cluster	94,17
5 Cluster	91,26
6 Cluster	94,17
7 Cluster	86,41
8 Cluster	88,35

Berdasarkan hasil akurasi pada table 2 dapat dijelaskan bahwa pada 3 cluster terbentuk memiliki akurasi 98,06% ketepatan pengelompokan. Pada 4 cluster dan 6 cluster yang terbentuk memiliki akurasi 94,17% ketepatan pengelompokan. Adapun akurasi ketepatan pengelompokan dari 5 cluster yang terbentuk adalah 91,26%. Untuk 7 cluster yang terbentuk akurasi yang diperoleh adalah 86,41% dan pada 8 cluster yang terbentuk akurasi pengelompokan yang diperoleh adalah 88,35%. jadi, jumlah cluster yang memiliki akurasi terbaik adalah 3 cluster yang terbentuk dengan akurasi pengelompokan 98,06%. Berikut adalah wilayah-wilayah yang termasuk dalam anggota 3 cluster yang terbentuk:

Tabel 3 Anggota dari Cluster Terbaik

No.	Jumlah Anggota (Desa/Kelurahan)	Anggota
1	34	Pabbudukang, Anrong Appaka, Sibatua, Bonto Perak, Balleangin, Tonasa, Kel. Kassi, Kel. Bori Appaka, Kel. Sapanang, Bowong Cindea, Tabo-tabo, Mangilu, Bulu Cindea, Biring Ere, Labbakkang, Mangallekana,

		Pundata Baji, Borimasunggu, Taraweang, Bara Batu, Bonto-Bonto, Talaka, Ma'rang, Attang Salo, Padang Lampe, Bonto Matene, Segeri, Minasa Te'ne, Kalabbirang, Biraeng, Bonto Kio, Bontoa, Bonto Langkasa, Kabba
2	59	Sapuka, Sabaru, Balobaloang, Sabalana, Tampaang, Sailus, Kapoposang Bali, Poleonro, Kalu-Kalukuang, Pammas, Sabaru, Kanyurang, Doang-duangang Lompo, Marasende, Dewakang, Mattiro Sompe, Mattiro Bintang, Mattiro deceng, Mattiro langi, Mattiro Ujung, Mattiro Dolangeng, Mattiro Bone, Mattiro Adae, Tekolabbua, Balocci Baru, Pattallassang, Manakku, Gentung, Kanaungan, Bonto Manai, Kassiloe, Tamangapa, Pitusunggu, Pitue, Punranga, Alesipitto, Bone, Bawasolo, Baring, Parenreng, Panaikang, Benteng, Manggalung, Tamarupa, mandalle, Boddie, Coppo Tompong, Bantimurung, Tondongkura, Lanne, Bulu Tellue, Malaka, Bonto Birao, Mattiro Kanja, Mattiro Baji, Mattiro Bulu, Mattiro Labangeng, Mattiro Uleng Mattiro Walie
3	10	Satanger, Mattiro Matae, Jagong, Mappasaile, Paddoang-doangan, Tumampua, Tompo Bulu, Samalewa, Batara, Mattiro Bombang

5. Kesimpulan

Berdasarkan metode yang digunakan yakni cluster k-means, maka dibentuk beberapa cluster yang akan digunakan dalam membentuk kelompok awal yang mendiskripsikan miskin di kabupaten pangkep yakni 3 cluster, 4 cluster, 5 cluster, 6 cluster, 7 cluster dan 8 cluster. Dari beberapa cluster yang dibentuk selanjutnya akan dilakukan analisis diskriminan untuk melihat akurasi pengelompokan. Berdasarkan analisis diskriminan diperoleh 98,06% (3 cluster), 94,17% (4 cluster), 91,26% (5 cluster), 94,17% (6 cluster), 86,41% (7 cluster), dan 88,35% (8 cluster). Dari uraian di atas diperoleh bahwa dengan membentuk pengelompokan dengan 3 cluster yang terbentuk, maka hasil akurasi pengelompokan yang diperoleh lebih baik dari cluster lain yang dibentuk.

References

- Agusta, Y. (2007). K-Means – Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem dan Informatika*, 47-60
- Budhiantoro, S., Viktoria, F., Wiko, S., Ah, M. Dan Dwi R., P., A. (2013). Multidimensional Fuzzy Proverty Index (MPI): Konsep dan Pengukurannya di Indonesia. *PRAKARSA Economics Policy Working Paper*.
- BPS. (2016). *Statistik Kemiskinan di Indonesia*. Jakarta. Badan Pusat Statistik
- Dian, Podje, L., & Saleh. (2015). Analisis Diskriminan Pada Klasifikasi Nasabah Menunggak Dan Tidak Menunggak Dengan Cross Validation. *Jurnal Statistika*, 1-10.
- Ediyanto, Mara, M. N., & Hadewi, N.S. (201). Pengklasifikasian Karakteristik dengan Metode K-Means Cluster Analysis. *Buletin Ilmiah Mat.Stat. dan Terapannya*, 133-136.
- Febriyana. (2011). Analisis Diskriminan K-Means dan K-Median pada Data Indikator Kemiskinan (Studi Kasus data Indikator kemiskinan Kabupaten di Indonesia Tahun 2009). Jakarta. Skripsi. Universitas Islam Negeri Syarif Hidayatullah.
- Hastuti, N. F. (2013). Pemanfaatan Metode K-Means Clustering dalam Penentuan Penerimaan Beasiswa. Surakarta. Skripsi. FMIPA Universitas Sebelas Maret.
- Irwan. (2010). Penggunaan Analisis Diskriminan dan Self-Organizing Maps (SOM) dalam Pengelompokan Objek (Studi Pengelompokan Desa Miskin di Kabupaten Wajo). Makassar. Skripsi. FMIPA Universtas Negeri Makassar.
- Kristanto, N. H., L.A, A. C., & S, H. B. (2016). Implementasi K-Means Clustering untuk Pengelompokan Analisis Rasio Profitabilitas dalam working Capitaal. *Juisi*, 9-15.
- Kuncoro, M. 2003. *Metode Riset Untuk Bisnis dan Ekonomi*. Jakarta : Erlangga

- Laeli, S. (2014). Analisis Cluster dengan Average Linkage Method Linkage dan Ward's Method untuk Data Responden Nasabah Asuransi Jiwa Unit Link. *Statistika*
- Laraswati, T. F. (2014). Perbandingan Kinerja Metode Complete Linkage, Metode Average Linkage, dan Metode K-Means dalam Menentukan Hasil Analisis Cluster. *Statistika*.
- Ong, J. O. (2013). Implementasi Algoritma K-Means Clustering untuk Menentukan Strategi Marketing President University. *Jurnal Ilmiah Teknik Industri*, 10-20.
- Prasetyo, E. (n.d.). *Data Mining (Konsep dan Aplikasi Menggunakan MATLAB)*. Andi Yogyakarta.
- Sitepu, R., Irmeilyana, & Gultom, B. (2011). Analisis Cluster terhadap Tingkat Pencemaran Industri di Sumatra Selatan. *Jurnal Penelitian Sains*, 3.
- Tiro, M. A., Sukarna, & Aswi. (2010). *Statistika Deskriptif Peubah Banyak*. Makassar: Andira Publisher, Makassar
- Weeb, A.R., (2002). *Statistical Pattern Recognition Second Edition*, John Wiley and Sons Ltd England.
- Yulianto, S., & Hidayatullah, K. H. (2014). Analisis Klaster Untuk Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Indikator Kesejahteraan Rakyat. *Jurnal Statistika*, 56-63