

KOMPARASI ANALISIS GEROMBOL (*CLUSTER*) DAN BILOT DALAM PENGELOMPOKAN

I MADE ANOM ARIAWAN¹, I PUTU EKA NILA KENCANA²,
NI LUH PUTU SUCIPTAWATI³

^{1,2,3}Jurusan Matematika FMIPA Universitas Udayana, Bukit Jimbaran-Bali
e-mail: ¹anom.ariawan41@gmail.com, ²I.Putu.ENK@gmail.com, ³putusuciptawati@yahoo.co.id

Abstract

One of functions of multivariate analysis is to group data. Multivariate analysis often used in grouping data are cluster analysis and biplot analysis. In this paper, a comparative analysis will be made between clusters analysis and biplot analysis for grouping the data. Technique used in the cluster analysis is k-mean method and biplot analysis used two-dimensional display. The results ware that biplot analysis produces are better in grouping accuracy than clusters analysis. But in general, biplot analysis can not be said to be better than clusters analysis in grouping the data and vice versa.

Keywords: *comparative, multivariate analysis, clusters analysis, biplot analysis*

1. Pendahuluan

Analisis multivariat merupakan analisis statistika yang berkaitan dengan pendiskripsian ataupun interpretasi data yang melibatkan banyak peubah dan objek secara bersama-sama [1]. Salah satu interpretasi yang dimaksud adalah pengelompokan data [4]. Pengelompokan data merupakan pengorganisasian sekumpulan data yang besar dengan cara membagi data tersebut kedalam beberapa gerombol (kelompok). Gerombol-gerombol yang terbentuk akan mampu menjelaskan adanya persamaan maupun perbedaan dari keseluruhan data yang diteliti [2]. Tujuan dari pengelompokan data adalah untuk mempermudah proses analisis dan interpretasi dari data besar dengan membagi data tersebut menjadi beberapa gerombol.

Terdapat beberapa teknik dalam pengelompokan suatu data. Namun dalam tulisan ini, pengelompokan data yang dimaksudkan adalah pengelompokan data yang didasarkan pada objek dari data tersebut. Salah satu analisis yang sering digunakan untuk pengelompokan ini adalah analisis gerombol (*cluster*). Analisis gerombol merupakan analisis multivariat yang bertujuan untuk mengelompokkan objek-objek dari data yang diteliti berdasarkan kesamaan karakteristik yang dimilikinya [7]. Kesamaan karakteristik ini biasanya diukur menggunakan ukuran kedekatan antarobjek yang dapat berupa ukuran kemiripan atau ketakmiripannya.

¹ Mahasiswa Jurusan Matematika FMIPA Universitas Udayana

^{2,3} Staf Pengajar Jurusan Matematika FMIPA Universitas Udayana

Lebih jauh mengenai pengelompokan data berdasarkan objek, analisis multivariat lain yang juga sering digunakan adalah analisis biplot. Tujuan analisis ini adalah untuk menggambarkan baris (objek) dan kolom (peubah) yang ada pada matriks data secara bersama-sama dalam sebuah grafik berdimensi rendah (biasanya dua atau tiga) [3]. Penggambaran ini meliputi keragaman dan korelasi antar peubah, serta kedekatan antarobjek yang nantinya akan mampu mengidentifikasi pengelompokan objek.

Kedua analisis multivariat tersebut memiliki keunggulan masing-masing, analisis gerombol mampu digunakan dalam mengelompokkan data dalam jumlah objek yang besar serta dengan skala pengukuran peubah yang berbeda mulai dari nominal sampai interval [2]. Analisis biplot mampu menampilkan secara langsung peubah penciri atau peubah yang paling dominan dari suatu kelompok objek yang terbentuk pada hasil tampilan analisis biplot [7]. Pada tulisan ini akan dilakukan komparasi pada kedua analisis tersebut untuk melihat analisis terbaik dalam pengelompokan objek dari suatu data. Komparasi ini dilakukan dengan melihat ketepatan setiap objek dalam kelompok-kelompok yang dihasilkan dari kedua analisis itu terhadap posisi objek tersebut dalam kelompok pada data kontrol. Suatu objek dikatakan memiliki ketepatan jika berada dalam kelompok yang sama baik itu pada kelompok hasil analisis maupun kelompok pada data kontrol. Sebagai contoh, objek i dikatakan memiliki ketepatan apabila objek tersebut berada dalam kelompok 1 pada data kontrol dan kelompok 1 pada hasil pengelompokan analisis gerombol.

2. Metode Penelitian

Data yang digunakan dalam tulisan ini adalah data sekunder yang diambil dari tulisan tentang kredit usaha rakyat (KUR) suatu bank swasta di Kota Denpasar [6]. Dalam data ini terdapat seratus objek yang telah terbagi menjadi tiga kelompok dan dengan sembilan peubah yang memiliki skala pengukuran yang beragam mulai dari nominal sampai interval. Tiga kelompok tersebut digunakan sebagai kontrol dalam proses perbandingan yang akan dilakukan. Teknik pengelompokan data yang digunakan dalam analisis gerombol adalah teknik tak berhirarki yaitu menggunakan *k-mean methods*, sementara itu dalam tampilan analisis biplot hanya akan menggunakan dua dimensi saja, sehingga lebih mudah dalam penafsiran kelompok yang terbentuk.

Langkah-langkah analisis data dalam tulisan ini sebagai berikut.

1. Mempersiapkan data yang akan dianalisis dengan tahapan sebagai berikut.
 - a. Menghilangkan peubah dengan skala nominal dan mengubah peubah dengan skala ordinal menjadi kontinu dengan *Method of Successive Interval* (MSI) agar semua peubah yang akan digunakan nantinya akan berskala kontinu, sehingga mampu dianalisis dengan analisis gerombol dan biplot bersama-sama.

- b. Melakukan standarisasi data pada hasil langkah *a* sebelumnya, agar satuan pada setiap peubah dalam data yang digunakan sama.
 2. Melakukan analisis gerombol dengan tahapan sebagai berikut.
 - a. Pengelompokan data dengan teknik tak berhirarki (*k-means method*), memperhatikan jumlah kelompok yang nantinya akan dibentuk sudah ditentukan.
 - b. Melakukan uji hipotesis beda rata-rata untuk dua populasi saling bebas, yaitu pada kelompok-kelompok hasil analisis gerombol terhadap kelompok-kelompok pada data kontrol.
 - c. Membandingkan hasil pengelompokan analisis gerombol dengan pengelompokan pada data kontrol untuk mendapatkan persentase ketepatan untuk semua objek pada data yang dianalisis.
 3. Melakukan analisis biplot dengan tahapan sebagai berikut.
 - a. Pemilihan parameter $\beta = 1$ pada dekomposisi matriks $S = UL^\beta L^{\beta-1}A'$ agar memberikan tampilan biplot yang menekankan pada gambaran jarak antara pasangan baris sehingga baik digunakan untuk melihat kedekatan objek-objeknya^[5].
 - b. Pembentukan kelompok pada hasil analisis biplot yaitu dengan melakukan proyeksi ortogonal setiap objek terhadap semua vektor (peubah) yang ada dalam tampilan analisis biplot. Semua objek yang berada dalam satu kelompok yang sama akan memiliki jarak yang relatif dekat terhadap suatu peubah yang sama.
 - c. Melakukan analisis deskriptif pada kelompok dalam data kontrol untuk melihat variansi setiap peubah dalam kelompok tersebut.
 - d. Membandingkan hasil pengelompokan analisis biplot dengan pengelompokan pada data kontrol untuk mendapatkan persentase ketepatan untuk semua objek pada data yang dianalisis.
 4. Mengkomparasikan dua hasil analisis yang telah dilakukan.

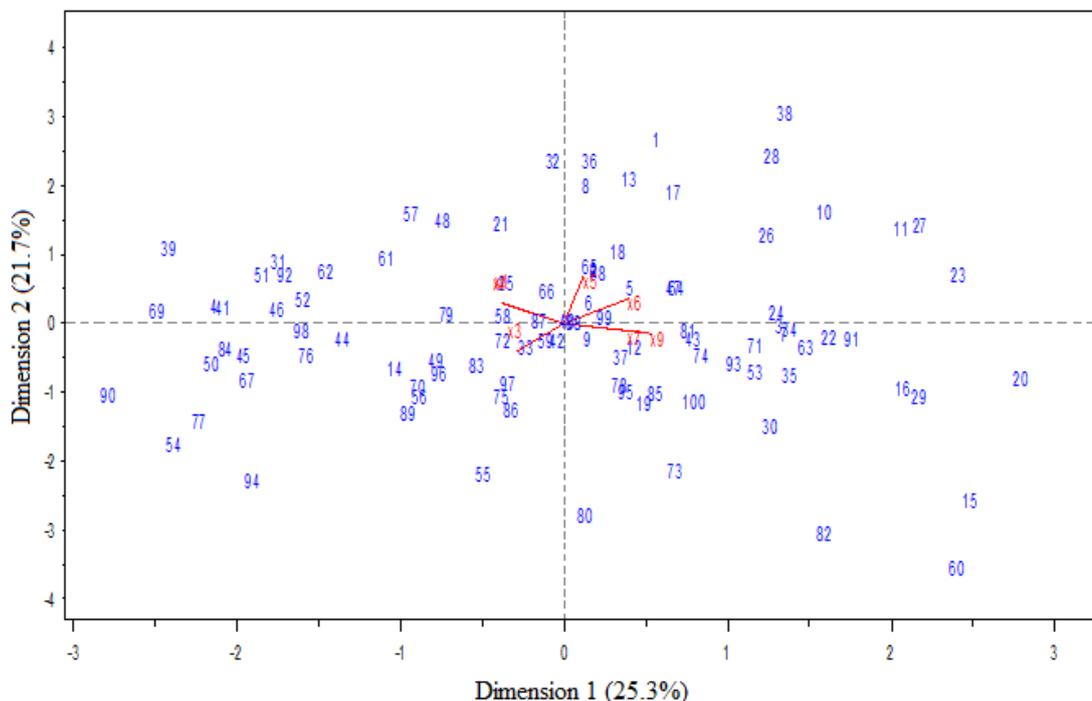
3. Hasil dan Pembahasan

Pada langkah awal, dari sembilan peubah yang ada pada data KUR hanya digunakan enam peubah dengan skala pengukuran kontinu dan satu peubah dengan skala pengukuran ordinal. Peubah dengan skala ordinal ini selanjutnya diubah menjadi skala pengukuran kontinu dengan *Method of Successive Interval* (MSI). Sehingga pada analisis berikutnya digunakan data dengan tujuh peubah kontinu. Berikutnya data tersebut distandarasi untuk menyamakan satuannya, data hasil standarisasi selanjutnya dikelompokkan menjadi tiga kelompok yaitu sesuai dengan jumlah kelompok pada data kontrol. Pengelompokan dilakukan dengan analisis gerombol yaitu dengan *k-mean methods*. Selanjutnya dilakukan uji hipotesis beda rata-rata dua populasi saling bebas pada kelompok hasil analisis gerombol dengan kelompok pada data kontrol. Hasil uji tersebut menyimpulkan bahwa kelompok 1, 2 dan 3 pada pengelompokan hasil analisis gerombol masing-

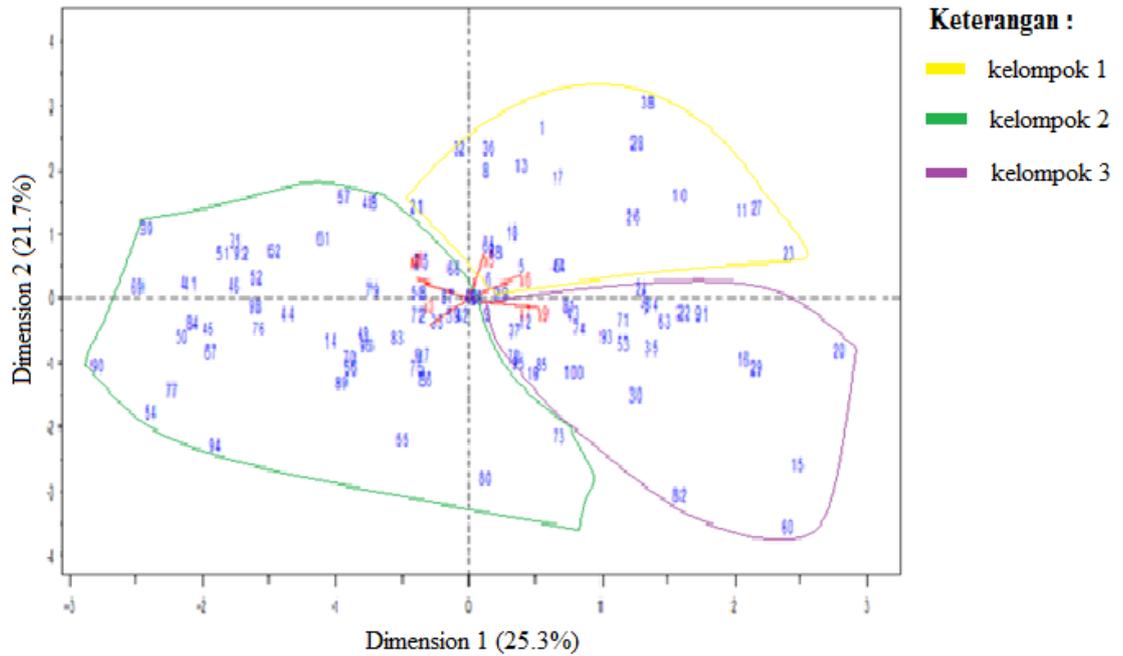
masing sama dengan kelompok 1, 2 dan 3 pada data kontrol, sehingga proses perbandingan dapat dilakukan. Berdasarkan hasil perbandingan antara kelompok hasil analisis gerombol dengan kelompok pada data kontrol, didapat 44 objek berada dalam kelompok yang sama baik itu pada kelompok hasil analisis gerombol maupun kelompok pada data kontrol, dengan kata lain didapatkan persentase ketepatan sebesar 44%.

Setelah mendapatkan persentase ketepatan pada hasil pengelompokan analisis gerombol, selanjutnya dilakukan analisis biplot. Hasil analisis biplot ditampilkan pada Gambar 1. Hasil tersebut menunjukkan bahwa informasi yang ditampilkan dari dimensi satu dan dua pada tampilan analisis biplot adalah sebesar 47% dari keseluruhan informasi dari data yang digunakan. Tampilan biplot pada Gambar 1 juga memperlihatkan adanya korelasi antarpeubah yang ditunjukkan oleh sudut yang terbentuk antarpeubah tersebut, yaitu semakin kecil sudut yang terbentuk antarpeubah maka korelasi antarpeubah tersebut makin besar. Peubah-peubah yang saling berkorelasi yaitu X_1 dengan X_4 , X_5 dengan X_6 dan X_7 dengan X_9 .

Selanjutnya dilakukan proyeksi ortogonal pada setiap titik koordinat objek terhadap semua vektor peubah untuk mencari jarak setiap objek terhadap vektor peubah dalam tampilan biplot yang dihasilkan. Berdasarkan hasil proyeksi ortogonal dan korelasi antarpeubah pada Gambar 1, diperoleh hasil pengelompokan objek pada analisis biplot dan ditampilkan pada Gambar 2.



Gambar 1. Tampilan Hasil Analisis Biplot



Gambar 2. Tampilan Biplot Dengan Pengelompokan Objek

Gambar 2 menunjukkan bahwa, pada setiap kelompok yang terbentuk terdapat peubah penciri atau peubah yang paling dominan dalam kelompok tersebut. Peubah penciri dapat dilihat dari panjang vektor dari peubah tersebut yang lebih panjang dibandingkan vektor yang dibentuk oleh peubah lain dalam satu kelompok yang sama. Pada kelompok 1, 2 dan 3 mempunyai peubah penciri yaitu berturut-turut adalah X_5 , X_3 dan X_0 .

Berikutnya dilakukan analisis deskriptif pada data kontrol untuk melihat peubah dengan variansi terbesar pada masing-masing kelompok yang ada pada data kontrol. Hasil analisis deskriptif diperoleh bahwa peubah dengan variansi terbesar untuk kelompok 1, 2 dan 3 pada data kontrol masing-masing adalah X_5 , X_3 dan X_0 . Berdasarkan hasil analisis deskriptif dan pengelompokan objek pada Gambar 2 dapat disimpulkan bahwa, kelompok 1 pada hasil pengelompokan analisis biplot dan kelompok 1 pada data kontrol memiliki peubah dominan yang sama yaitu X_5 . Selanjutnya, kelompok 2 pada hasil pengelompokan analisis biplot juga memiliki peubah dominan yang sama dengan kelompok 2 pada data kontrol yaitu X_3 . Pada kelompok 3 dari hasil pengelompokan analisis biplot dan kelompok 3 pada data kontrol juga memiliki peubah dominan yang sama yaitu X_0 . Selanjutnya dilakukan proses perbandingan antara kelompok pada data kontrol dengan kelompok hasil analisis biplot. Hasilnya diperoleh persentase ketepatan sebesar 52%.

Berdasarkan hasil perbandingan antara kelompok data kontrol dengan kelompok hasil analisis gerombol dan analisis biplot, didapatkan bahwa persentase ketepatan yang diperoleh dari hasil analisis gerombol dan analisis biplot secara berturut-turut adalah 44% dan 51%. Hal ini berarti, pada perbandingan yang dilakukan terlihat bahwa pengelompokan analisis biplot

menghasilkan persentase ketepatan yang lebih baik dari pada pengelompokan analisis gerombol.

4. Kesimpulan

Hasil penelitian diperoleh bahwa pengelompokan analisis biplot menghasilkan persentase ketepatan yang lebih baik dari pada pengelompokan analisis gerombol. Namun secara umum belum dapat dikatakan bahwa analisis biplot lebih baik dari analisis gerombol dalam pengelompokan data dan begitu juga sebaliknya.

Daftar Pustaka

- [1] Bilodeau, Martin and Brenner. D. 1999. *Theory of Multivariate Statistics*. New York: Springer.
- [2] Everitt, B. S., Landau. S., Leese. M., dan Stahl, D. 2011. *Cluster Analysis*. Fifth Edition. London: John Wiley & Sons, Ltd.
- [3] Gabriel, K.R. 1971. The Biplot Graphic Display of Matrices with Application to Principal Componen Analysis. *Biometrika* 58(3): 453-467
- [4] Johnson, R. A. and D. W. Wichern. 2007. *Applied Multivariat Statistical Analysis*. Sixth Edition. New Jersey: Printice Hall.
- [5] Kohler, Ulrich dan Magdalena Luniak. 2005. *Data Inspection Using Biplot*. Berlin: Stata Corp LP.
- [6] Kusumawardani, N. M. D., 2011. Pengklasifikasian Debitur Kredit Usaha Rakyat (KUR) Dengan Analisis Deskriminan Dan Regresi Logistik Ordinal. *Skripsi*. Jurusan Matematika, Fakultas Mipa dan Ilmu Pengetahuan Alam, Universitas Udayana.
- [7] Mattjik, A. A., dan I. M. Sumertajaya. 2011. *Sidik Peubah Ganda*. Bogor: Institut Pertanian Bogor Press.