

PENENTUAN AWAL KEANGGOTAAN ANALISIS KLASTER NON HIRARKI (*K-MEANS*)

Monika Afriana¹, Sigit Nugroho² dan Fachri Faisal²

¹Mahasiswa Jurusan Statistika, Prodi Matematika, FMIPA, UNIB, Bengkulu

²Dosen Pembimbing Jurusan Statistika, Prodi Matematika, FMIPA, UNIB, Bengkulu
monikaafriana027@gmail.com

ABSTRAK

Penelitian ini bertujuan untuk menentukan awal keanggotaan analisis klaster non hirarki *K-Means* dan memberikan contoh penerapan penentuan awal keanggotaan analisis klaster non hirarki *K-Means*. Metode penelitian yang digunakan adalah studi literatur dan teladan penerapan yang diambil dari buku berjudul “Applied Multivariate Statistical Second Edition” karangan Johnson dan Winchern (2002: 741). Hasil penelitian menunjukkan bahwa dalam menentukan awal keanggotaan (jumlah klaster K) pada klaster *K-Means* secara signifikan nilai K harus lebih kecil dari jumlah objek dalam himpunan data. Namun, perlu diperhatikan bahwa tidak terdapat aturan khusus dalam menentukan jumlah klaster K , terkadang jumlah klaster yang diinginkan tergantung pada subjektif seseorang. Algoritma *K-Means* tidak selalu menghasilkan hasil pengklasteran yang sama, salah satu penyebabnya karena ada unsur pengacakan pada titik pusat klaster .

Kata kunci: *K-Means*, Algoritma *K-Means*, Penentuan nilai K .

1. PENDAHULUAN

Analisis klaster merupakan salah satu teknik dalam analisis multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek pengamatan menjadi beberapa kelompok berdasarkan karakteristik yang dimilikinya. (Nugroho, 2008). Proses pengelompokan data pada analisis klaster dapat dilakukan dengan dua metode yaitu metode hirarki dan metode non hirarki. Metode hirarki memulai pengelompokkan dengan dua atau lebih objek yang memiliki kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang mempunyai kedekatan kedua. Demikian seterusnya sehingga klaster akan membentuk semacam ‘pohon’ dimana ada hirarki (tingkatan) yang jelas antar objek, dari yang paling mirip sampai paling tidak mirip. Beberapa metode untuk proses klastering secara hirarki yaitu *single linkage*, *complete linkage*, *average linkage*, *ward’s method*, dan *centroid method* (Santoso, 2010).

Berbeda dengan metode hirarki, analisis klaster dengan metode non hirarki merupakan metode klaster yang menentukan jumlah klaster secara manual. Kekurangan metode ini adalah tidak disediakan gambar dendrogram bentuk dari klaster untuk masing-masing sampel (Baroroh, 2012). Metode ini dimulai dengan menentukan terlebih dahulu jumlah klaster yang diinginkan. Setelah jumlah klaster diketahui, baru proses klaster dilakukan tanpa mengikuti proses hirarki. Metode ini disebut dengan Klaster *K-Means* (Santoso, 2010).

Pemecahan permasalahan pengklasteran dengan menggunakan metode *K-Means* ini secara umum dijalankan dengan menggunakan algoritma yang dikenal dengan algoritma *K-Means*. Algoritma *K-Means* adalah metode klastering berbasis jarak yang

membagi data ke dalam sejumlah klaster dan hanya bekerja pada atribut numerik. Algoritma ini pada awalnya mengambil sebagian dari banyaknya komponen dari populasi yang akan dijadikan sebagai pusat awal klaster (*centroid*). Pada langkah ini *centroid* dipilih secara acak dari sekumpulan populasi data (Martiana, dkk, 2011). Selanjutnya menghitung ukuran jarak dari masing-masing objek ke *centroid*. Setelah objek masuk pada *centroid* terdekat dan membentuk klaster baru, *centroid* baru ditentukan kembali dengan menghitung rata-rata objek pada *centroid* yang sama. Jika masih ada perbedaan dengan *centroid* yang sudah dibentuk, maka dilakukan perhitungan kembali *centroid* baru (Lathifaturrahmah, 2010).

Beberapa penelitian yang menggunakan analisis klaster non hirarki *K-means* telah dilakukan. Wakhidah (2010) membahas tentang pengklasteran menggunakan algoritma klaster *K-Means*. Kumar dkk (2013) membahas tentang algoritma klaster *K-Means* terbaru dan efisien. Usman dkk (2013) meneliti tentang peningkatan algoritma *K-Means* untuk mendapatkan *centroid* awal. Hasil pengklasteran dengan metode *K-Means* sangat bergantung pada nilai pusat awal klaster (*centroid*) yang diberikan. Pemberian nilai awal yang berbeda bisa menghasilkan klaster yang berbeda (Lathifaturrahmah, 2010). Berdasarkan uraian singkat diatas, penulis tertarik melakukan penelitian mengenai penentuan awal keanggotaan pada analisis klaster non hirarki *K-means*.

2. ANALISIS KLASTER NON HIRARKI (*K-MEANS*)

A. Analisa Klaster

Analisis Klaster merupakan salah satu teknik dalam analisis multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek pengamatan menjadi beberapa kelompok berdasarkan karakteristik yang dimilikinya. Analisis klaster mengelompokkan objek-objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam kelompok yang sama, serta mempunyai kemiripan satu dengan yang lain (Johnson & Wichern, 2002). Analisis klaster dapat dibagi menjadi dua langkah dasar yaitu :

1. Pemilihan ukuran kedekatan
Memeriksa setiap pasangan pengamatan (objek) untuk kemiripan nilainya. Sebuah kemiripan (*proximity*) ukuran didefinisikan untuk mengukur "kedekatan" dari objek. Semakin dekat maka semakin homogen.
2. Pemilihan algoritma kelompok yang membangun
Berdasarkan kedekatan mengukur obyek ditugaskan untuk kelompok sehingga perbedaan antara kelompok menjadi besar dan pengamatan dalam kelompok menjadi sedekat mungkin (Härdle & Simar, 2007).

Secara logika sebuah klaster dikatakan baik jika mempunyai homogenitas (kesamaan) yang tinggi antar anggota dalam satu klaster (*within cluster*) dan heterogenitas (perbedaan) yang tinggi antar klaster yang satu dengan klaster lainnya (*between cluster*). Analisis klaster harus memenuhi dua asumsi yaitu:

1. Sampel yang diambil harus benar-benar dapat mewakili populasi.
2. Tidak ada multikolinieritas (korelasi antar obyek), jika ada maka nilai multikolinieritasnya kurang dari 0,5 (Nugroho, 2008).

B. Pengukuran Konsep Kemiripan Antar Jarak Objek

Kemiripan antar objek diukur dengan menggunakan ukuran jarak. Beberapa ukuran jarak yang sering digunakan antara lain jarak *Euclid*, jarak kuadrat *Euclid*, jarak City-block (Manhattan), dan jarak *Chebychev*.

1. Jarak *Euclid*

Jarak *Euclid* biasanya dihitung dari data mentah dan tidak dari data standar. Metode ini memiliki beberapa keuntungan, antara lain jarak dari dua objek tidak dipengaruhi oleh penambahan dari objek baru untuk di

analisis, yang mungkin merupakan pencilan.

Jika $x' = (x_1, x_2, \dots, x_n)$ dan $y' = (y_1, y_2, \dots, y_n)$ maka $d(x, y)$ adalah jarak antara objek x dengan objek y (Johnson & Wichern, 2002).

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$
$$d(x, y) = \sqrt{(x - y)'(x - y)} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

2. Jarak kuadrat *Euclid*

Jarak kuadrat *Euclid* merupakan jumlah kuadrat yang berbeda dari nilai antara dua objek pada seluruh variable. $d(x, y)$ adalah Jarak antara objek x dengan objek y (Fiedling, 2007).

jika $x = (x_1, x_2, \dots, x_n)$ dan $y = (y_1, y_2, \dots, y_n)$ maka

$$d(x, y) = \sum_{i=1}^n (x_i - y_i)^2 \quad (2)$$

untuk $i = 1, 2, 3, \dots, n$

3. Jarak *city-block* atau *Manhattan*

Jarak ini sering juga disebut jarak *Minkowski*, diberikan oleh jarak antara objek x dengan objek y (Johnson & Wichern, 2002).

$$d(x, y) = [\sum_{i=1}^n |x_i - y_i|^m]^{\frac{1}{m}} \quad (3)$$

untuk $i = 1, 2, 3, \dots, n$

4. Jarak *Chebychev*

Jarak *Chebychev* merupakan jarak yang didasarkan pada nilai perbedaan mutlak maksimum antara dua objek. Pengukuran dengan jarak ini tidak dapat dilakukan untuk data yang bersifat *outlier*. $d(x, y)$ adalah Jarak antara objek x dengan objek y (Mattjik & Sumertajaya, 2011).

$$d(x, y) = \max |x_i - y_i| \quad (4)$$

untuk $i = 1, 2, 3, \dots, n$

C. Analisis Klaster Non Hirarki

Analisis klaster dengan metode non hirarki merupakan metode klaster yang menentukan jumlah klaster secara manual (Baroroh, 2012). Teknik analisis klaster non hirarki dirancang untuk mengelompokkan item, bukan variabel, menjadi kumpulan K klaster. Banyaknya klaster, K , ditentukan terlebih dahulu untuk memulai prosedur pengklasteran. Metode analisis klaster non hirarki berhubungan dengan *K-means*, ada tiga pendekatan yang digunakan untuk menempatkan masing-masing observasi dalam satu klaster yaitu:

1. ***Sequential Threshold***, Metode *Sequential Threshold* memulai dengan pemilihan satu klaster dan menempatkan semua objek yang berada pada jarak tertentu ke dalamnya sebagai nilai awal klaster. Jika semua objek yang berada pada jarak tertentu telah dimasukkan, klaster yang kedua dipilih dan ditempatkan semua objek yang berjarak tertentu ke dalamnya.

Setelah itu, kluster ketiga dipilih dan proses dilanjutkan seperti yang sebelumnya.

2. **Parallel Threshold**, Metode *Parallel Threshold* berbanding terbalik dengan pendekatan *Sequential Threshold*, metode ini memilih sejumlah kluster secara bersamaan dan menempatkan objek-objek kedalam kluster yang memiliki jarak terdekat kluster awal. Pada saat proses berlangsung, jarak antar kluster awal dapat ditentukan untuk memasukkan beberapa objek ke dalam kluster-kluster. Beberapa variasi pada metode ini, yaitu sisa objek-objek tidak dikelompokkan jika berada di luar jarak tertentu dari sejumlah kluster.
3. **Optimization**, Metode ketiga serupa dengan kedua metode sebelumnya, perbedaannya metode ini memungkinkan untuk menempatkan kembali objek-objek ke dalam kluster yang lebih dekat (Prayudho, 2008).

D. K-Means

K-Means merupakan metode analisis kluster non hirarki yang dimulai dengan menentukan terlebih dahulu jumlah kluster yang diinginkan. Setelah jumlah kluster diketahui, baru proses kluster dilakukan tanpa mengikuti proses hirarki (Santoso, 2004). *K-Means* banyak digunakan di berbagai bidang karena sederhana dan mudah diimplementasikan. Pengklasteran *K-Means* sangat cocok untuk data dengan ukuran yang besar karena memiliki kecepatan yang lebih tinggi dibandingkan metode hirarki. Hasil pengklasteran mungkin tergantung pada urutan observasi data (Nugroho, 2008).

Algoritma *K-Means* adalah metode klastering berbasis jarak yang membagi data ke dalam sejumlah kluster dan hanya bekerja pada atribut numerik. Secara umum metode analisis kluster *K-means* menggunakan algoritma sebagai berikut:

1. Tentukan k sebagai jumlah kluster yang di bentuk.
Untuk menentukan banyaknya kluster K dilakukan dengan beberapa pertimbangan seperti pertimbangan teoritis dan konseptual yang mungkin diusulkan untuk menentukan berapa banyak kluster.
2. Bangkitkan K titik pusat kluster (*centroid*) awal secara *random*.
Penentuan *centroid* awal dilakukan secara *random*/acak dari objek-objek yang tersedia sebanyak K kluster, kemudian untuk menghitung *centroid* kluster ke- i berikutnya, digunakan rumus sebagai berikut :

$$v = \frac{\sum_{i=1}^n x_i}{n} ; i = 1,2,3,\dots,n \quad (5)$$

Dimana:

v = *centroid* pada kluster
 x_i = objek ke- i
 n = banyaknya objek

3. Hitung jarak setiap objek ke masing-masing *centroid* dari masing-masing kluster. Untuk menghitung jarak antara objek dengan *centroid* digunakan jarak *euclidian* yang dirumuskan sebagai berikut:

$$d(i, l) = (\sum_{j=1}^p [x(i, j) - \bar{x}(j, l)]^2)^{1/2} \quad (6)$$

Dimana :

$d(i, l)$ = jarak *euclid*
 $x(i, j)$ = nilai objek ke- i pada variabel ke- j
 $\bar{x}(j, l)$ = rata-rata objek ke- i pada variabel ke- j
 i = banyaknya objek
 j = banyaknya variabel
 l = banyaknya kluster

4. Alokasikan masing-masing objek ke dalam *centroid* yang paling terdekat. Untuk melakukan pengalokasian objek kedalam masing-masing kluster pada saat iterasi secara umum dapat dilakukan dengan dua cara yaitu dengan *hard k-means*, dimana secara tegas setiap objek dinyatakan sebagai anggota kluster dengan mengukur jarak kedekatan terhadap titik pusat kluster tersebut, cara lain dapat dilakukan dengan *fuzzy C-Means*.
5. Lakukan iterasi, kemudian tentukan posisi *centroid* baru dengan menggunakan persamaan (2.7).
6. Ulangi langkah 3 jika posisi *centroid* baru tidak sama.
Pengecekan konvergensi dilakukan dengan membandingkan matriks *group assignment* (matrik kelompok yang disimbolkan dengan G^m dimana m merupakan banyaknya iterasi) pada iterasi sebelumnya dengan matrik *group assignment* pada iterasi yang sedang berjalan. Jika hasilnya sama maka algoritma *K-means* sudah konvergen, tetapi jika berbeda maka belum konvergen sehingga perlu dilakukan iterasi berikutnya.

Berdasarkan cara kerjanya, algoritma *K-Means* memiliki kelebihan dan kekurangan antara lain sebagai berikut:

A. Kelebihan *K-Means*

1. Algoritma *K-Means* adalah algoritma klasik untuk menyelesaikan masalah pengklasteran dikarenakan algoritma ini relatif sederhana dan cepat.
2. Algoritma *K-Means* memberikan hasil yang relatif baik pada kluster konveks (Zhang & Fang, 2013).
3. Algoritma *K-Means* memiliki ketelitian yang cukup tinggi terhadap ukuran objek, sehingga algoritma ini relatif lebih terukur dan efisien untuk mengolah data dalam jumlah besar.
4. Algoritma *K-Means* tidak berpengaruh pada urutan objek (Simamora, 2005).

B. Kekurangan *K-Means*

1. Sangat sensitif pada pembangkitan titik pusat awal secara *random* (acak).
2. Memungkinkan suatu kluster tidak memiliki anggota
3. Hasil pengklasteran bersifat tidak unik (selalu berubah-ubah) terkadang bagus terkadang tidak.
4. Sulit untuk mencapai *global optimum*.
5. Menentukan banyaknya jumlah kluster sebelum mengetahui jumlah kluster yang optimal (Lathifaturrahmah, 2010).
6. Tidak jelas berapa banyak kluster *K* yang terbaik.
7. Hanya bekerja pada atribut numerik.
8. Hanya dapat digunakan dalam data yang *mean*-nya dapat ditentukan
9. Tidak mampu menangani data yang mempunyai penyimpangan-penyimpangan (*noisy data* dan *outlier*) (Andayani, 2007).

E. Indeks Validitas

Indeks validitas merupakan suatu ukuran yang digunakan untuk menentukan jumlah kelompok yang optimal. Untuk menentukan jumlah kluster yang paling optimal dan dapat memvalidasi apakah partisi yang diterapkan dalam proses klasterisasi sesuai dengan data, maka digunakan indeks pengukuran kluster validasi. Validitas dalam *K-Means* ditentukan dengan banyaknya kelompok optimum melalui perhitungan indeks validitas. Salah satu indeks validitas yang digunakan dalam *K-Means* adalah indeks *XB*.

Formula indeks *XB* sebagai berikut:

$$XB(c) = \frac{\sum_{i=1}^n \sum_{k=1}^c [\sum_{j=1}^m (x_{ij} - v_{kj})^2]^{(U_{ik})^w}}{N \min \|v_{kj}^1, v_{kj}^t\|^2} \quad (7)$$

Dimana :

- c = jumlah kluster
- N = jumlah objek yang akan dikelompokkan
- U_{ik} = derajat keanggotaan data ke-*i* pada kluster ke-*k*
- $\|v_{kj}^1, v_{kj}^t\|^2$ = jarak antara pusat kluster v_{kj}^1 dan v_{kj}^t
- t = banyaknya iterasi atau perulangan
- $\min \|v_{kj}^1, v_{kj}^t\|^2$ = jarak minimum antara pusat kluster v_{kj}^1 dan v_{kj}^t

3. PENENTUAN AWAL KEANGGOTAAN ANALISIS KLASTER NON HIRARKI (*K-MEANS*)

A. Komponen Utama

Dalam metode non hirarki (*K-Means*) data dibagi menjadi *K* partisi dimana setiap partisi mewakili sebuah kluster. Dasar pengklasteran dalam metode ini adalah menempatkan objek berdasarkan rata-rata (mean) kluster terdekat. Oleh karena itu, metode ini bertujuan untuk meminimumkan *error* akibat partisi *n* objek ke dalam *k* kluster (Nuningsih, 2010). Algoritma *K-Means* dalam proses pengklasterannya memerlukan 3 komponen yaitu:

1. Jumlah Kluster

Untuk menentukan banyaknya kluster *K* dilakukan dengan beberapa pertimbangan seperti pertimbangan teoritis dan konseptual yang mungkin diusulkan untuk menentukan berapa banyak kluster. Jumlah kluster *K* juga dapat ditentukan melalui pendekatan hirarki. Namun, perlu diperhatikan bahwa tidak terdapat aturan khusus dalam menentukan jumlah kluster *K*, terkadang jumlah kluster yang diinginkan tergantung pada subjektif seseorang.

2. Kluster Awal

Kluster awal yang dipilih berkaitan dengan penentuan pusat kluster awal (*centroid* awal). Dalam hal ini, terdapat beberapa pendapat dalam memilih kluster awal untuk metode *K-Means* sebagai berikut:

- a. Berdasarkan Hartigan (1975), pemilihan kluster awal dapat

ditentukan berdasarkan interval dari jumlah setiap observasi.

- b. Berdasarkan Rencher (2002), pemilihan kluster awal dapat ditentukan melalui pendekatan salah satu metode hirarki
- c. Berdasarkan Teknomo (2007), pemilihan kluster awal dapat secara acak dari semua observasi.

Oleh karena adanya pemilihan kluster awal yang berbeda ini maka kemungkinan besar solusi kluster yang dihasilkan akan berbeda pula.

3. Ukuran Jarak

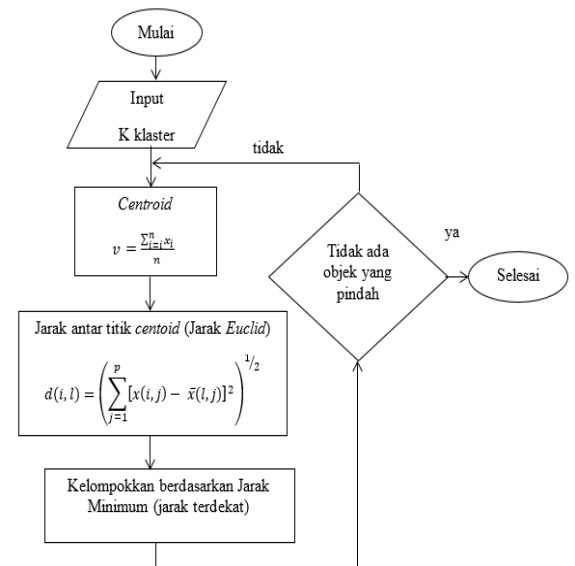
Dalam hal ini, ukuran jarak digunakan untuk menempatkan observasi ke dalam kluster berdasarkan centroid terdekat. Ukuran jarak yang digunakan dalam metode *K-Means* adalah jarak *euclid* (Nuningsih, 2010).

B. Penentuan Awal Keanggotaan Analisis Kluster Non Hirarki (*K-Means*)

Kinerja algoritma *K-Means* dapat dipengaruhi oleh nilai yang dipilih dari *K*. Oleh karena itu, ditetapkan sebuah nilai *K*. Hal ini penting untuk nilai *K* yang dianggap cukup besar, untuk mencerminkan karakteristik khusus dari himpunan data. Nilai *K* yang dipilih harus secara signifikan lebih kecil dari jumlah objek dalam himpunan data, yang merupakan tujuan utama untuk melakukan pengelompokan data (Pham dkk, 2004). Pada penelitian ini, kita akan menentukan penentuan awal keanggotaan analisis kluster non hirarki *K-Means* yaitu *K* kluster yang dapat dibentuk dari *n* data dan membagi objek menjadi *K* himpunan dengan memperhatikan semua kemungkinan. Berikut langkah-langkah penentuan awal keanggotaan pada analisis kluster non hirarki *K-Means*:

1. Tentukan nilai *K* sebagai jumlah kluster yang dibentuk. Pada penelitian ini, semua kemungkinan kluster akan digunakan dimana nilai harus *K* lebih kecil dari jumlah objek dalam himpunan data.
2. Bangkitkan *K* titik pusat kluster (*centroid*) awal secara acak, dimana kita bisa mencoba semua kemungkinan centroid dari *n* data.
3. Hitung jarak setiap objek ke masing-masing *centroid* dari masing-masing kluster dengan menggunakan jarak *euclidian* pada persamaan (2.1). Kemudian buat matriks jarak yang disimbolkan dengan D^m , dimana *m* adalah iterasi proses pengklasteran.

4. Pengelompokan objek, untuk menentukan anggota kluster kita dapat memperhitungkan jarak minimum objek. Nilai yang diperoleh dalam keanggotaan data pada matriks jarak adalah 0 atau 1, dimana nilai 1 untuk data yang dialokasikan ke kluster dan nilai 0 untuk data yang dialokasikan ke kluster lain. Sama halnya seperti langkah 3, untuk mempermudah pengelompokan maka dibuat sebuah matriks yaitu matriks kelompok atau *group* yang dilambangkan dengan G^m , dimana *m* adalah iterasi proses pengklasteran.
5. Kembali ke langkah 2, lakukan pengulangan hingga nilai *centroid* yang dihasilkan tetap dan anggota kluster tidak berpindah ke kluster lain. Secara flowchart algoritma *K-Means* digambar sebagai berikut (Wakhidah, 2010):



Gambar 1. Flowchart Algoritma *K-Means*

4. TELADAN PENERAPAN DAN PEMBAHASAN

A. Teladan Penerapan

Untuk lebih memahami materi mengenai penentuan awal keanggotaan pada analisis kluster non hirarki *K-Means* diberikan teladan penerapan yang diperoleh dari buku berjudul "Applied Multivariate Statistical Second Edition" karangan Johnson dan Winchern (2002: 741) berupa data 2 variabel X_1 dan X_2 dengan empat item *A, B, C, D*. Adapun data tersebut yaitu:

Tabel 1 Data Teladan Penerapan

Item	Observasi	
	x_1	x_2
A	5	4
B	1	-2
C	-1	1
D	3	1

B. Pembahasan

Dalam menyelesaikan persoalan pada teladan penerapan, hal pertama yang kita lakukan adalah menentukan nilai K sebagai jumlah kluster yang dibentuk. Dengan memperhatikan semua kemungkinan pada data Tabel.1 kita dapat mengelompokkan item menjadi 2 dan 3 kluster.

Setelah penentuan nilai K , kita tentukan nilai kluster awal secara acak dengan memperhatikan semua kemungkinan dari empat item pada Tabel.1 dimana untuk $K=2$ yaitu (AB dan CD), (AC dan BD), (AD dan BC), (A dan BCD), (B dan ACD), (C dan ABD) dan (D dan ABC). Untuk $K=3$ yaitu (A, B dan CD), (A, BC dan D), (AB, C dan D), (AC, B dan D), (AD, B dan C) dan (A, BD dan C).

Selanjutnya, setelah penentuan kluster awal maka dapat dihitung nilai *centroid*, jarak antar titik terhadap *centroid*, mengelompokkan data sesuai jarak minimum jika tidak terjadi perpindahan kluster proses berhenti dan jika terjadi perpindahan maka dihitung kembali nilai *centroid* baru dari kemungkinan kluster awal yang telah diperoleh.

Dengan melakukan proses pengklasteran seperti diatas diperoleh hasil akhir sebagai berikut:

Untuk $K=2$

Tabel 1. Hasil Akhir Pengklasteran Untuk $K=2$

Kluster Awal	Kluster Akhir
AB dan CD	AD dan BC
AC dan BD	ACD dan B
	A dan BCD
	AD dan BC
AD dan BC	AD dan BC
A dan BCD	A dan BCD
B dan ACD	B dan ACD
C dan ABD	BC dan AD
D dan ABC	A dan BCD

Untuk $K=3$

Tabel 2. Hasil Akhir Pengklasteran Untuk $K=2$

Kluster Awal	Kluster Akhir
A, B dan CD	A, B dan CD
A, BC dan D	A, BC dan D
AB, C dan D	A, BC dan D
	AD, C dan B
	A, C dan BD
AC, B dan D	AC, B dan D
AD, B dan C	AD, B dan C
A, BD dan C	A, BD dan C

5. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan penelitian yang dilakukan tentang penentuan awal keanggotaan analisis kluster non hirarki *K-Means* dapat diambil kesimpulan sebagai berikut:

1. Dalam menentukan awal keanggotaan pada analisis kluster non- hirarki *K-Means* yaitu jumlah K , secara signifikan nilai K yang dipilih harus lebih kecil dari jumlah objek dalam himpunan data. Jumlah kluster K juga dapat ditentukan melalui pendekatan hirarki. Namun, perlu diperhatikan bahwa tidak terdapat aturan khusus dalam menentukan jumlah kluster K , terkadang jumlah kluster yang diinginkan tergantung pada subjektif seseorang.
2. Dari teladan penerapan pada penentuan awal keanggotaan analisis kluster non hirarki *K-Means* dapat disimpulkan bahwa algoritma *K-Means* tidak selalu menghasilkan hasil pengklasteran yang sama. Hal ini dikarenakan adanya unsur pengacakan pada titik pusat kluster dan jumlah sampel yang sedikit yaitu $n < 30$.

B. Saran

Dalam penulisan skripsi ini, penulis hanya membahas masalah penentuan awal keanggotaan analisis kluster non hirarki *K-Means* pada data 2 dimensi dengan mencoba semua kemungkinan K . Untuk penelitian lebih lanjut, sebaiknya penentuan awal keanggotaan analisis kluster non hirarki *K-Means* diaplikasikan pada data lebih dari 2 dimensi dengan mencoba semua kemungkinan K , namun dalam pengukuran kemiripan antar objek menggunakan 2 jarak yaitu jarak *city-block* (*manhattan*) dan jarak *euclid* sebagai perbandingan.

6. DAFTAR PUSTAKA

- Agusta, Y. 2007. *K-Means – Penerapan, Permasalahan dan Metode Terkait*. Jurnal Sistem dan Informatika Vol.3 (Februari 2007), 47-60.
- Andayani, S. 2007. Pembentukan *cluster* dalam *Knowledge Discovery in Database* dengan Algoritma *K-Means*. SEMNAS Matematika dan Pendidikan Matematika. Jurdik Matematika FMIPA UNY Yogyakarta pada tanggal 24 Nopember 2007. <http://staff.uny.ac.id/sites.pdf>. (Diakses tanggal 20 Mei 2013).
- Baroroh, A. 2012. *Analisis Multivariat dan Time Series dengan SPSS 21*. Tangerang Selatan: PT. ELEK MEDIA KOMPUTINDO.
- Ediyanto, Sari, M.N., dan Satyahadewi, N. 2013. Pengklasifikasian Karakteristik Dengan Metode *K-Means Cluster Analysis*. Buletin Ilmiah Mat. Stat. dan Terapannya(Bimaster) Volume 02, No.2 (2013), hal 133-136. <http://download.portalgaruda.org/article.php?article=79314&val=2316>. (Diakses tanggal 19 April 2013).
- Härdle, W and Simar, L. 2007. *Applied Multivariate Statistical Analysis Second Edition*. Springer: Berlin and Louvain la Neuve.
- Heryandi, A. 2009. Kajian Analisis Clustering Dan Penerapannya Pada Sistem Pendukung Keputusan. *Tesis*. Program Studi Magister Informatika. ITB.
- Johnson, R. A and D, W. Wichern. 2002. *Applied Multivariate Statistical Analysis, Fifth Edition*. Pearson Education. United States Of America.
- Kumar, G. K., Chary, T. B., and Premchand, P. 2013. A New and Efficient K-Means Clustering Algorithm. *Volume 3, Issue 11, November 2013*. <http://www.cs.umd.edu/~mount/Projects/KMeans/pami02.pdf>. (Diakses tanggal 19 April 2013).
- Lathifaturrahmah. 2010. Perbandingan Hasil Penggerombolan Metode *K-Means*, *Fuzzy K-Means*, dan *Two Step Cluster*. Tesis, Departemen Matematika Sekolah Pasca Sarjana Institut Pertanian Bogor: Bogor. Tidak Dipublikasikan. <http://repository.ipb.ac.id/handle/123456789/26950>. (Diakses tanggal 22 April 2013)
- Martiana, E., Mubtada'i N.R., dan Purnomo.E. 2012. Penggunaan Metode Pengklasteran Untuk Menentukan Bidang Tugas Akhir Mahasiswa Teknik Informatika. <http://repo.eepis-its.edu/130/>. (Diakses tanggal 22 April 2013)
- Mattjik, A. A dan I. M. Sumertajaya. 2011. Sidik Peubah Ganda Dengan menggunakan SAS. IPB: Bandung.
- Nugroho, S. 2008. *Statistika Multivariat Terapan*. Unib Press: Bengkulu.
- Nuningsih, Sri. 2010. *K-Means Clustering: Studi Kasus Pada Data Pengujian Kualitas Susu Di Koperasi Peternakan Bandung Selatan*. *Skripsi*. Fakultas Matematika Dan Ilmu Pengetahuan Alam. UPI. <http://a-research.upi.edu/skripsiview.php?start=8112>. (Diakses tanggal 1 Oktober 2013)
- Pham. D. T., Dimov. S. S., and Nguyen. C.D. 2004. Selection of *K* in *K-means Clustering*. *Manufacturing Engineering Centre, Cardiff University, Cardiff, UK*. <http://www.ee.columbia.edu/~dpwe/papers/PhamDN05-kmeans.pdf>. (Diakses tanggal 1 Mei 2013).
- Prayudho, B.J., 2008. Analisis Cluster. <http://prayudho.wordpress.com>. (Diakses tanggal 19 April 2013).
- Rencher, A.C., 2002. *Methods Of Multivariate Analysis Second Edition*. Wiley Series In Probability and Mathematical Statistics: Canada.
- Santoso, S. 2010. *Statistika Multivariat Konsep & Aplikasi dengan SPSS*. Jakarta: PT. Elex Media Komputindo.
- Simamora B. 2005. *Analisis Multivariat Pemasaran*. PT. Gramedia Pustaka Utama: Jakarta.
- Supranto, J. 2004. *Analisis Multivariat Arti dan Interpretasi*. Rineka Cipta: Jakarta.
- Usman, G., Ahmad, U., dan Ahmad, M. 2013. Improved K-Means Clustering Algorithm by Getting Initial Cenroids. *World Applied Sciences Journal 27 (4): 543-551, 2013*. [http://www.idosi.org/wasj/wasj27\(4\)13/21.pdf](http://www.idosi.org/wasj/wasj27(4)13/21.pdf). (Diakses tanggal 1 Mei 2013)
- Wakhidah, N. 2010. Clustering Menggunakan *K-Means Algorithm (K-Means Algorithm Clustering)*. *Tr@nsForMat!ka Vol 8 No 1*.
- Zhang, C and Fang, Z. 2013. An Improved *K-Means Clustering Algorithm*. *Journal of Information & Computational Science 10: 1 (2013) 193–199*.