

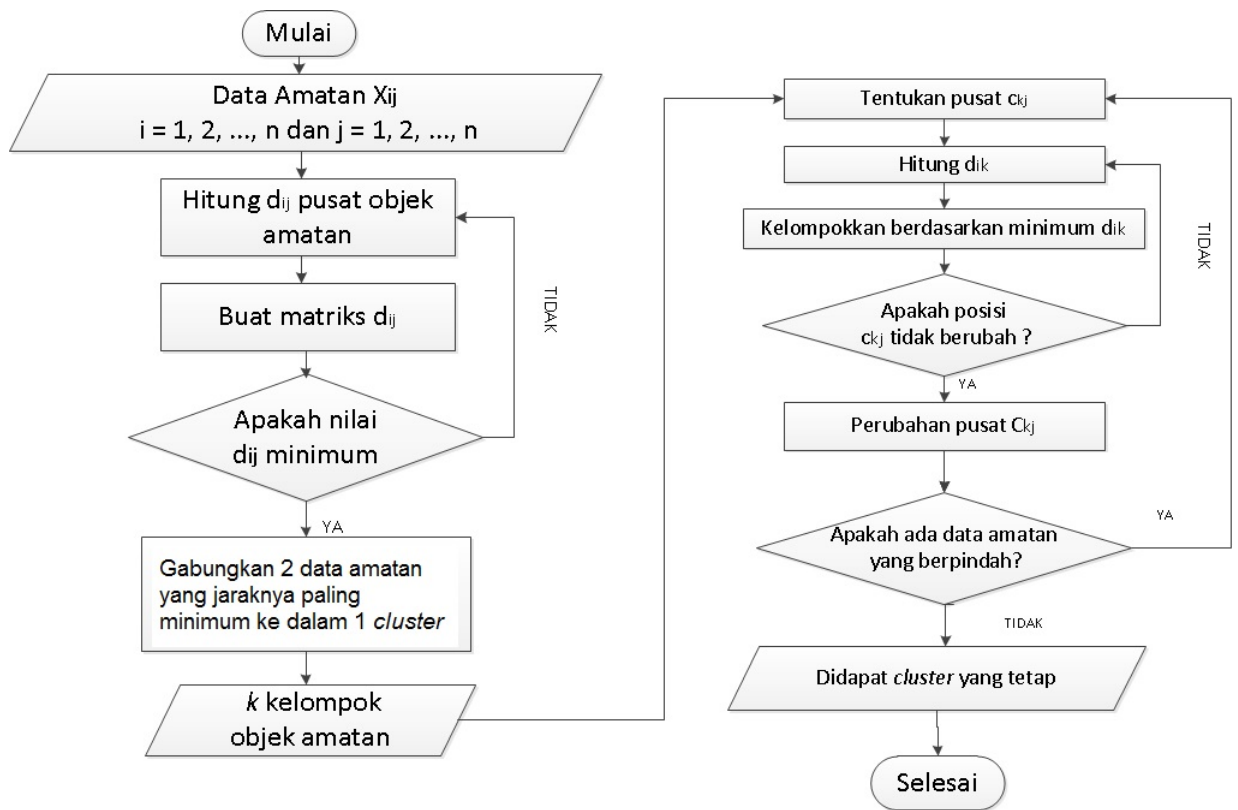
## BAB III

### PEMBAHASAN

Pada bab ini akan dibahas secara lebih terperinci mengenai pengelompokan menggunakan metode *Hybrid Mutual Clustering* dengan jarak *Square Euclidean* dan juga penerapan metode *Hybrid Mutual Clustering* ke dalam contoh kasus. Sebelumnya, akan dibahas terlebih dahulu mengenai pengelompokan metode *Hybrid Mutual Clustering* menggunakan jarak *Square Euclidean*.

#### 3.1 Pengelompokan Menggunakan Metode *Hybrid Mutual Clustering* dengan Jarak *Square Euclidean*

Metode *Hybrid Mutual Clustering* merupakan metode penggabungan antara metode *bottom-up* dan *top-down*. Pengelompokan *Hybrid Mutual Clustering* ini menggunakan jarak *Square Euclidean*. *Hybrid mutual* disini mengkombinasikan kelebihan metode *bottom-up clustering (agglomerative)* dan *top-down clustering (k-means)*. Metode *Hybrid* terdiri dari dua metode yaitu *bottom-up* yaitu metode pengelompokan dimulai dari kelompok kecil menjadi kelompok yang lebih besar (*agglomerative*) dan *top-down* yaitu metode pengelompokan dengan memecah kelompok besar menjadi kelompok lebih kecil seperti metode *k-means* membagi sebanyak  $k$  kelompok. Berikut adalah diagram alir pengelompokan *Hybrid Mutual Clustering* menggunakan jarak *Square Euclidean*:



Gambar 3.1: Diagram alir pengelompokan menggunakan metode *Hybrid Mutual Clustering* dengan jarak *Square Euclidean*.

Berikut akan dijelaskan algoritma pengelompokan *Hybrid Mutual Clustering* menggunakan jarak *Square Euclidean*

1. Masukkan data amatan  $X_{ij}$  dengan  $i = 1, \dots, n$  dan  $j = 1, \dots, n$ .
2. Hitung  $d_{ij}$  untuk mencari jarak antar objek amatan, dengan persamaan berikut:

$$d_{ij} = \sum_{q=1}^p (x_{iq} - x_{jq})^2.$$

3. Bentuk matriks  $d_{ij}$ .
4. Tentukan nilai minimum  $d_{ij}$ . Jika  $d_{ij}$  minimum maka proses berlanjut ke tahap 4 dari diagram alir. Namun, jika  $d_{ij}$  tidak minimum, maka proses

mengulang dari tahap 2 diagram alir.

5. Gabungkan 2 data amatan yang jaraknya paling minimum ke dalam satu *cluster* sampai diperoleh  $k$  kelompok, dimana  $d_{ij}=0$ .
6. Bentuk sebanyak  $k$  kelompok objek amatan.
7. Tentukan pusat  $c_{kj} = \frac{\sum_{j=1}^p x_{ij}}{p}$  dengan peubah sebanyak  $p$ .
8. Hitung  $d_{ik} = \sum_{j=1}^p (x_{ij} - c_{kj})^2$ .
9. Kelompokkan berdasarkan minimum  $d_{ik}$ .
10. Tentukan nilai  $c_{kj}$ . Jika  $c_{kj}$  berubah maka proses berlanjut ke tahap 11. Jika  $c_{kj}$  tidak berubah, maka proses mengulang dari tahap 8.
11. Pusat  $c_{kj}$  berubah.
12. Identifikasi data amatan. Jika data amatan berubah posisi maka proses mengulang dari tahap 7. Jika data amatan tidak berubah, maka proses berlanjut ke tahap 13.
13. Didapat *Cluster* tetap yang memiliki proporsi terbaik.

Dari pengelompokkan *Hybrid Mutual Clustering* menggunakan jarak *Square Euclidean* didapatkan kelompok sebanyak  $k$  dan memiliki jarak berdekatan yg menunjukkan bahwa karakteristik setiap objek amatan pada kelompok adalah sama.

Pada subbab selanjutnya akan dibahas mengenai penerapan metode *Hybrid Mutual Clustering* dengan jarak *Square Euclidean* ke dalam contoh kasus menggunakan program aplikasi SPSS

### 3.2 Penerapan Metode *Hybrid Mutual Clustering* dengan Jarak *Square Euclidean* ke dalam contoh kasus

Sumber data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) dengan judul Persentase Penduduk Buta Aksara Menurut Provinsi di Indonesia. Pada situs tersebut data yang tertera ada dalam periode 2003-2015, tetapi yang diambil oleh penulis hanya tahun 2015 saja. Unit pengamatan yang dipakai ada 34 provinsi di Indonesia, meliputi Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, dan Papua. Data yang dipakai hanya mencakup 3 variabel, yaitu penduduk dengan umur di bawah 14 tahun, penduduk dengan umur 15 sampai 44 tahun, dan penduduk dengan umur di atas 45 tahun. Variabel tersebut digunakan karna tersediaan data.

Metode *Hybrid Mutual Clustering* ini akan diaplikasikan ke dalam pengelompokan Penduduk Buta Aksara Menurut Provinsi di Indonesia tahun 2015 menggunakan *software SPSS*. Data yang berhasil dikumpulkan ditampilkan di Lampiran 4.1.

Setelah melakukan *Hybrid Mutual Clustering*, dihasilkan output dengan penjelasan sebagai berikut:

- Tabel *output Proximity Matrix* yang terdapat pada lampiran. Pada tabel tersebut menunjukkan bahwa semua data sejumlah 34 obyek telah diproses tanpa ada data yang hilang. Tabel tersebut menunjukkan matriks jarak antara variabel satu dengan variabel yang lain. Semakin kecil jarak *Euclidean*, maka semakin mirip kedua variabel tersebut sehingga akan membentuk kelompok (*cluster*).
- Selain tabel *Proximity Matrix*, terdapat pula tabel *outpyt* yang berjudul *Agglomeration Schedule*. Tabel tersebut merupakan hasil proses clustering dengan metode Between Group Linkage. Setelah jarak antar variabel diukur dengan jarak *Square Euclidean*, maka dilakukan pengelompokan, yang dilakukan secara bertingkat. Berikut adalah tabel *Agglomeration Schedule*:

Penjelasan tabel *Agglomeration Schedule*:

1. Stage 1: terbentuk 1 *cluster* yang beranggotakan Kota 4 (Riau) dan Kota 21 (Kalimantan Tengah) dengan jarak hampir mendekati nol (perhatikan pada kolom *Coefficients*). Karena proses aglomerasi dimulai dari 2 obyek yang terdekat, maka jarak tersebut adalah yang terdekat dari sekian kombinasi jarak 34 obyek yang ada. Selanjutnya lihat kolom terakhir (*Next Stage*), terlihat angka 5. Hal ini berarti *clustering* selanjutnya dilakukan dengan melihat *Stage 5*, dengan penjelasan berikut.
2. Baris ke-5 (*stage 5*) terlihat obyek ke-3 (Kota Sumatera Barat) membentuk *cluster* dengan kota 4 (Riau). Dengan demikian, sekarang *cluster* terdiri dari 3 obyek yaitu Kota Sumatera Barat, Riau, dan Kalimantan Tengah. Sedangkan jarak sebesar 0,004 merupakan jarak rata-rata obyek terakhir yang bergabung dengan 2 obyek

Agglomeration Schedule						
Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	4	21	,000	0	0	5
2	10	23	,002	0	0	11
3	12	22	,002	0	0	15
4	6	29	,003	0	0	6
5	3	4	,004	0	1	7
6	6	32	,004	4	0	11
7	2	3	,004	0	5	14
8	11	25	,005	0	0	18
9	9	26	,007	0	0	12
10	5	7	,009	0	0	12
11	6	10	,012	6	2	14
12	5	9	,014	10	9	15
13	1	16	,015	0	0	17
14	2	6	,024	7	11	22
15	5	12	,033	12	3	22
16	20	27	,044	0	0	20
17	1	8	,052	13	0	25
18	11	31	,055	8	0	27
19	13	17	,073	0	0	21
20	19	20	,093	0	16	23
21	13	28	,094	19	0	28
22	2	5	,100	14	15	27
23	15	19	,112	0	20	26
24	14	24	,128	0	0	28
25	1	33	,229	17	0	29
26	15	30	,244	23	0	30
27	2	11	,248	22	18	29
28	13	14	,284	21	24	30
29	1	2	,380	25	27	31
30	13	15	,603	28	26	31
31	1	13	3,636	29	30	32
32	1	18	12,940	31	0	33
33	1	34	63,260	32	0	0

Gambar 3.2: Tabel *Agglomeration Schedule*.

sebelumnya, seperti tampak dalam *Proximity matrix* dan dapat dihitung sebagai berikut:

- (a) Jarak Kota Sumatera Barat dan Riau = 0,004
- (b) Jarak Kota Sumatera Barat dan Kalimantan Tengah = 0,004
- (c) Jarak rata-rata =  $(0,004 + 0,004) / 2 = 0,004$

3. Stage 2 : terjadi pembentukan *cluster* Kota 10 (Kepulauan Riau) dan Kota 23 (Kalimantan Timur) yang berjarak 0,002, yang kemudian berlanjut ke *Stage 11*.
4. Demikian seterusnya dari *stage 3* sampai ke *stage* terakhir.

Proses aglomerasi ini bersifat kompleks, khususnya perhitungan koefisien yang melibatkan sekian banyak obyek dan terus bertambah. Proses aglomerasi pada akhirnya akan menyatukan semua obyek menjadi satu *cluster*. Hanya saja dalam prosesnya dihasilkan beberapa *cluster* dengan masing-masing anggotanya, tergantung jumlah *cluster* yang dibentuk.

- Hasil *output cluster membership* dapat dilihat pada tabel berikut ini:

Cluster Membership			
Case	4 Clusters	3 Clusters	2 Clusters
1:ACEH	1	1	1
2:SUMATERA UTARA	1	1	1
3:SUMATERA BARAT	1	1	1
4:RIAU	1	1	1
5:JAMBI	1	1	1
6:SUMATERA SELATAN	1	1	1
7:BENGKULU	1	1	1
8:LAMPUNG	1	1	1
9:KEP. BANGKA BELITUNG	1	1	1
10:KEP. RIAU	1	1	1
11:DKI JAKARTA	1	1	1
12:JAWA BARAT	1	1	1
13:JAWA TENGAH	2	1	1
14:DI YOGYAKARTA	2	1	1
15:JAWA TIMUR	2	1	1
16:BANTEN	1	1	1
17:BALI	2	1	1
18:NUSA TENGGARA BARAT	3	2	1
19:NUSA TENGGARA TIMUR	2	1	1
20:KALIMANTAN BARAT	2	1	1
21:KALIMANTAN TENGAH	1	1	1
22:KALIMANTAN SELATAN	1	1	1
23:KALIMANTAN TIMUR	1	1	1
24:KALIMANTAN UTARA	2	1	1
25:SULAWESI UTARA	1	1	1
26:SULAWESI TENGAH	1	1	1
27:SULAWESI SELATAN	2	1	1
28:SULAWESI TENGGARA	2	1	1
29:GORONTALO	1	1	1
30:SULAWESI BARAT	2	1	1
31:MALUKU	1	1	1
32:MALUKU UTARA	1	1	1
33:PAPUA BARAT	1	1	1
34:PAPUA	4	3	2

Gambar 3.3: Tabel *Cluster Membership*.

Perincian jumlah cluster dengan anggota yang terbentuk dijabarkan sebagai berikut:

1. Apabila diinginkan dibentuk 4 cluster, maka:

- (a) Anggota cluster 1 adalah Kota Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Banten, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tenggara, Gorontalo, Maluku, Maluku Utara, dan Papua Barat
  - (b) Anggota cluster 2 adalah Kota Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Utara, Sulawesi Selatan, Sulawesi Tengah, dan Sulawesi Barat.
  - (c) Anggota cluster 3 adalah Kota Nusa Tenggara Barat.
  - (d) Anggota cluster 4 adalah Kota Papua.
2. Apabila ditentukan dibentuk 3 cluster, maka :
- (a) Anggota cluster 1 adalah Kota Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Utara, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tenggara, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Barat, Gorontalo, Maluku, Maluku Utara, dan Papua Barat.
  - (b) Anggota cluster 2 adalah Kota Nusa Tenggara Barat.
  - (c) Anggota cluster 3 adalah Kota Papua.
3. Apabila ditentukan dibentuk 2 cluster, maka :
- (a) Anggota cluster 1 adalah Kota Kota Aceh, Sumatera Utara,



Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Timur, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Utara, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tenggara, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Barat, Gorontalo, Maluku, Maluku Utara, dan Papua Barat .

(b) Anggota cluster 2 adalah Kota Papua.

Dari hasil di atas dapat dilihat bahwa peralihan dari kolom 4 *Clusters* ke kolom 3 *Clusters*, yang terjadi adalah penggabungan variabel-variabel yang sudah ada, bukan mengacak variabel dari awal. Pada kolom 3 *Clusters*, terlihat anggota pada *cluster* 1 adalah nama-nama kota yang sebelumnya menjadi anggota *cluster* 1 dan *cluster* 2 dari kolom 4 *Clusters*. Demikian pula dari anggota *cluster* 1 dari kolom 2 *Clusters* adalah penggabungan dari anggota *cluster* 1 dan *cluster* 2 dari kolom 3 *Clusters*. Dari proses tersebut terlihat bahwa kota Nusa Tenggara Barat dan Papua memang jauh berbeda (tidak similar) dengan kota lainnya, dan kota Papua adalah yang paling berbeda dengan kota-kota lainnya.

- Terdapat satu lagi hasil *output SPSS* yaitu dendogram. Dendogram berguna untuk menunjukkan anggota *cluster* yang ada jika akan ditentukan berapa *cluster* yang seharusnya dibentuk. Sebagai contoh yang terlihat dalam dendogram, apabila akan dibentuk 2 *cluster*, maka *cluster* 1 beranggotakan Kota DKI Jakarta sampai dengan Kota Jawa Tengah (sesuai urutan dalam dendogram); dan *cluster* 2 beranggotakan sisanya. Demikian seterusnya dapat dengan mudah dilihat anggota tiap cluster

sesuai jumlah *cluster* yang diinginkan. Berikut adalah grafik dendrogram:

***** H I E R A R C H I C A L C L U S T E R A N A L Y S I S *****						
Dendrogram using Average Linkage (Between Groups)						
Rescaled Distance Cluster Combine						
C A S E	0	5	10	15	20	25
Label	Num	+-----+-----+-----+-----+				
RIAU	4	--				
KALIMANTAN TENGAH	21	--				
SUMATERA BARAT	3	--				
SUMATERA UTARA	2	--				
KEP. RIAU	10	--				
KALIMANTAN TIMUR	23	--				
SUMATERA SELATAN	6	--				
GORONTALO	29	--				
MALUKU UTARA	32	--				
JAWA BARAT	12	--				
KALIMANTAN SELATAN	22	--				
KEP. BANGKA BELITUNG	9	--				
SULAWESI TENGAH	26	--				
JAMBI	5	--				
BENGKULU	7	--				
DKI JAKARTA	11	--+				
SULAWESI UTARA	25	--				
MALUKU	31	--				
ACEH	1	--				
BANTEN	16	--				
LAMPUNG	8	--				
PAPUA BARAT	33	-- +-----+				
KALIMANTAN BARAT	20	--				
SULAWESI SELATAN	27	--				
NUSA TENGGARA TIMUR	19	--				
JAWA TIMUR	15	--				
SULAWESI BARAT	30	--	+-----+			
JAWA TENGAH	13	--+				
BALI	17	--				
SULAWESI TENGGARA	28	--				
DI YOGYAKARTA	14	--				
KALIMANTAN UTARA	24	--				
NUSA TENGGARA BARAT	18	-----+				
PAPUA	34	-----+				

Gambar 3.4: Dendrogram Analisis Klaster Hierarki