

BAB III

DIVISIVE ANALISIS

Pada bab ini akan dipaparkan bagaimana konsep dari *divisive analysis* serta algoritma dari metode tersebut.

3.1 DEFINISI METODE *DIVISIVE*

Teknik *divisive* klustering termasuk dalam analisis kluster hierarchical. Pada setiap langkahnya, metode *divisive* terjadi penambahan kelompok kedalam dua nilai terkecil, sampai akhirnya semua elemen terkelompokan. Ini berarti bahwa kluster hierarchical dibangun dalam $n-1$ langkah ketika data mengandung n objek.

Teknik *divisive* merupakan proses pengklasteran yang didasarkan pada persamaan nilai rata-rata antar objek. Jika sebuah objek memiliki persamaan nilai rata-rata terbesar maka objek tersebut akan terpisah dan berubah menjadi *splinter group*. Pada teknik *divisive* ini perhitungan juga dilihat dari perbedaan atau selisih antara persamaan nilai rata-rata dengan nilai elemen matrik yang telah menjadi *splinter group*. Jika selisih nilai antara persamaan nilai rata-rata dengan nilai elemen matrik *splinter group* bernilai negatif, maka perhitungan terhenti sehingga harus dibuat matrik baru untuk mendapatkan kluster yang lain. Perhitungan ini terus dilakukan sedemikian sehingga semua objek terpisah.

3.2 ALGORITMA METODE *DIVISIVE*

Misalkan diberikan data X matriks berukuran $n \times p$ (n = jumlah sampel data, p = variabel setiap data). X_{ji} = data sampel ke- j ($j = 1, 2, \dots, n$) dan variabel ke- i ($i = 1, 2, \dots, p$).

1. Bentuk suatu matriks jarak dengan menggunakan jarak Euclid. Rumusnya adalah:

$$= \sqrt{\sum_{k=1}^i (x_{jk} - x_{lk})^2} ; i = 1, 2, \dots, p$$

Asumsikan setiap data dianggap sebagai klaster. Jika diberikan n data dan c klaster maka $n = c$, maka diperoleh matriks jaraknya, yaitu:

$$D(1)_{n \times n} = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix}$$

2. Hitung nilai rata-rata setiap objek dengan objek lainnya.
3. Dari tabel temukan objek yang memiliki nilai rata-rata yang terbesar, objek yang memiliki nilai rata-rata yang terbesar akan terpisah dan berubah menjadi *splinter group*.
4. Hitung selisih nilai antara elemen matrik *splinter group* dengan nilai rata-rata setiap objek yang tersisa.
5. Temukan objek yang memiliki nilai selisih terbesar antara elemen matrik *splinter group* dengan nilai rata-rata. Jika nilai selisih tersebut bernilai

positif, maka objek yang memiliki nilai selisih terbesar bergabung dengan *splinter group*.

- Ulangi langkah satu sampai enam sedemikian sehingga semua nilai selisih antara elemen matrik *splinter group* dengan nilai rata-rata bernilai negatif dan klaster terbagi menjadi dua klaster baru.

CONTOH :

Untuk mengilustrasikan algoritma analisis *divisive* clustering, akan diberikan contoh sebagai berikut dimana data tersebut telah di konversi menjadi sebuah matriks jarak Euclid

$$\begin{array}{c}
 a \\
 b \\
 c \\
 d \\
 e
 \end{array}
 \begin{array}{c}
 a \\
 b \\
 c \\
 d \\
 e
 \end{array}
 \begin{array}{c}
 c \\
 d \\
 e
 \end{array}
 \begin{array}{c}
 d \\
 e
 \end{array}
 \begin{array}{c}
 e
 \end{array}
 \left[\begin{array}{ccccc}
 0.0 & 2.0 & 6.0 & 10.0 & 9.0 \\
 2.0 & 0.0 & 5.0 & 9.0 & 8.0 \\
 6.0 & 5.0 & 0.0 & 4.0 & 5.0 \\
 10.0 & 9.0 & 4.0 & 0.0 & 3.0 \\
 9.0 & 8.0 & 5.0 & 3.0 & 0.0
 \end{array} \right]$$

Data yang terdapat dalam matrik memiliki nilai yang berbeda antara objek a, b, c, d, dan e. dalam contoh ini didapat :

<i>Objects</i>	<i>Average dissimilarity to the other objects</i>
a	$(2.0+6.0+10.0+9.0)/4 = 6.75$
b	$(2.0+5.0+9.0+8.0)/4 = 6.00$
c	$(6.0+5.0+4.0+5.0)/4 = 5.00$
d	$(10.0+9.0+4.0+3.0)/4 = 6.50$
e	$(9.0+8.0+5.0+3.0)/4 = 6.25$

Objek a disebut *splinter group*. Pada langkah ini mempunyai dua grup yaitu grup {a} dan grup {b,c,d,e}, tetapi perhitungan tidak berhenti sampai disini. Untuk setiap objek dari grup yang besar harus dihitung average dissimilarity dengan objek yang tersisa, dan membandingkan itu dengan average dissimilarity dengan objek dari *splinter group* :

<i>Objects</i>	<i>Average dissimilarity to remain objects</i>	<i>Average dissimilarity to objects of splinter group</i>	<i>Difference</i>
b	$(5.0+9.0+8.0)/3 = 7.33$	2.00	5.33
c	$(5.0+4.0+5.0)/3 = 4.67$	6.00	-1.33
d	$(9.0+4.0+3.0)/3 = 5.33$	10.00	-4.67
e	$(8.0+5.0+3.0)/3 = 5.33$	9.00	-3.67

Terlihat bahwa perbedaan terbesar terdapat pada objek b. Oleh sebab itu, objek b berpindah bagian kepada *splinter group*. Jadi, *splinter group* yang satu {a,b} dan grup sisanya {c,d,e}. Ketika dilakukan perhitungan ulang, didapat :

<i>Objects</i>	<i>Average dissimilarity to remain objects</i>	<i>Average dissimilarity to objects of splinter group</i>	<i>Difference</i>
c	$(5.0+4.0)/2 = 4.50$	$(6.0+5.0)/2 = 5.50$	-1.00
d	$(4.0+3.0)/2 = 3.50$	$(10.0+9.0)/2 = 9.50$	-6.00
e	$(5.0+3.0)/2 = 4.00$	$(9.0+8.0)/2 = 8.50$	-4.50

Pada bagian ini terlihat, untuk semua difference bernilai negatif. Oleh sebab itu tidak terjadi perpindahan. Proses dihentikan dan langkah divisive pertama selesai. Dimana data dibagi dalam dua kluster, yaitu {a,b}, dan {c,d,e}.

Dalam langkah berikutnya akan dilakukan pembagian kluster. Pembagian kluster dilakukan pada kluster yang memiliki Rata-rata terbesar. Rata-rata dari {a,b} adalah 2, dan untuk {c,d,e} adalah 5. Oleh sebab itu, akan terjadi pembagian kluster {c,d,e}, dengan matrik :

$$\begin{matrix} & c & d & e \\ c & \begin{bmatrix} 0.0 & 4.0 & 5.0 \end{bmatrix} \\ d & \begin{bmatrix} 4.0 & 0.0 & 3.0 \end{bmatrix} \\ e & \begin{bmatrix} 5.0 & 3.0 & 0.0 \end{bmatrix} \end{matrix}$$

Untuk menentukan *splinter* group, maka dihitung :

<i>Objects</i>	<i>Average dissimilarity to the other objects</i>
c	$(4.0+5.0)/2 = 4.50$
d	$(4.0+3.0)/2 = 3.50$
e	$(5.0+3.0)/2 = 4.00$

Ternyata objek c yang memiliki nilai positif terbesar, maka objek c masuk dalam *splinter* group. Jadi, terdapat dua grup yaitu {c} dan {d,e}, selanjutnya akan ditentukan :

<i>Objects</i>	<i>Average dissimilarity to remain objects</i>	<i>Average dissimilarity to objects of splinter group</i>	<i>Difference</i>
d	3.0	4.00	-1.00
e	3.0	5.00	-2.00

Proses ini dihentikan karena semua difference bernilai negatif. Karena itu, pembagian pada langkah dua {c,d,e} adalah {c} dan {d,e}. jadi, kluster yang

didapat $\{a,b\}$, $\{c\}$, dan $\{d,e\}$. kluster $\{c\}$ disebut singleton karena hanya mengandung satu objek.

Hasil dari hierarchi digambarkan sebagai berikut :

