

TUGAS 2

ANALISIS MULTIVARIAT DATA HBAT

Mata Kuliah : Analisis Data Multivariat (A)
Dosen Pengampu : Dr. Ahmad Saikhu, S.Si., MT.
NRP : 5025201229
Nama : Surya Abdillah

DESKRIPSI TUGAS

Soal-soal pemodelan data multivariat dari dataset "Multivariate Data Analysis 7e Dataset.xls".

Pemodelan dilakukan dengan menggunakan software R-Studio.

1. Buat analisis Clustering dengan Hierarchical Clustering sampai dengan 3 Klaster untuk data sheet "HBAT_200" dengan pendekatan single linkage, complete linkage, dan average linkage dengan menggunakan variabel X6 - X21!
2. Buat analisis Clustering dengan k-means Clustering sampai dengan 3 Klaster untuk data sheet "HBAT_200" dengan menggunakan variabel X6 - X21!
3. Buat analisis discriminant (Prediksi Grop) dengan menggunakan Fungsi Diskriminan dan Fungsi Klasifikasi untuk data "HBAT" dengan menggunakan variabel prediktor X17-X22 dan variabel respon (label kelas) X5. Lakukan split data-training dan data-testing dengan proporsi 70% dan 30% ! Bandingkan akurasi untuk kedua fungsi !
4. Buat analisis discriminant (Prediksi Grop) dengan menggunakan Fungsi Diskriminan dan Fungsi Klasifikasi untuk data "HBAT" dengan menggunakan variabel prediktor X17-X22 dan variabel respon (label kelas) X1. Lakukan split data-training dan data-testing dengan proporsi 70% dan 30% ! Bandingkan akurasi untuk kedua fungsi !
5. Lakukan analisis Conjoint untuk data sheet HBAT_CONJOINT ! Lakukan analisis terhadap hasil !

Catatan : Dokumen yang diunggah adalah output hasil pemodelan (.pdf) dan code r-studio (.zip atau .rar)

Dalam pengerjaan ini akan digunakan dataset HBAT, yang terdiri dari beberapa bagian seperti dalam tabel berikut:





Nama Sheet	Keterangan
HBAT	dataset HBAT dengan 100 data
HBAT_200	dataset HBAT dengan 200 data
HBAT_MISSING	dataset HBAT dengan terdapat <i>missing value</i>
HBAT_SPLITS	tabel untuk dapat membagi dataset menjadi 60 : 40 (1 menandakn termasuk bagian 40) atau 50 : 50
HBAT_CPLAN	tabel berisi level pada setiap profil (terdapat 25 profil)
HBAT_CONJOINT	urutan nilai dari masing-masing responden terhadap 22 profil (terdapat 86 responden)
HBAT_MDS, HBAT_CORRESP, dan HBAT_CORRESP_INDIV	dataset untuk <i>multidimensional scaling</i>
HBAT_SEM	dataset untuk <i>structural equation analysis</i> (CH 10, 11, dan 12)
HATCO	dataset yang digunakan pada versi terdahulu
SALES	dataset pelatihan <i>sales</i>

Dari keterangan pada tabel di atas, maka, dalam pengerjaan ini akan digunakan beberapa dataset saja, yaitu:

- *clustering*: HBAT_200
- *linear discriminant analysis*: HBAT 200
- *conjoint analysis*: HBAT_CPLAN, HBAT_CONJOINT

IMPORT DATASET

Dataset tersedia dalam bentuk excel, dengan format 1997 - 2003. RStudio menerima format excel .xlsx. Oleh sebab itu, kita perlu merubah format excel menjadi .xlsx, yakni dengan cara *save as* menjadi format .xlsx. Dataset ini merupakan hasil segmentasi market yang telah dilakukan. Skala penilaian 0 - 10 dengan 10 berarti “Excellent” dan 0 berarti “Poor”.

 Multivariate_Data_Analysis_7e_Datasets_EXCEL		04/05/2023 20:24	Microsoft Excel 97-2...	511 KB
 Multivariate_Data_Analysis_7e_Datasets_EXCEL		04/05/2023 20:26	Microsoft Excel Work...	262 KB

PENJELASAN TERKAIT DATA SHEET HBAT_200

24 fitur dengan 200 baris data

Deskripsi Variabel	Tipe Variabel	Pengertian	Value
Data Warehouse Classification Variables Data dasar terkait perusahaan pelanggan dan relasi dengan HBAT			
X ₁ Customer Type	Nonmetric	Lama pelanggan telah membeli produk HBAT	1: $x < 1$ tahun 2: $1 < x < 5$ tahun 3: $x > 5$ tahun
X ₂ Industry Type	Nonmetric	Tipe industri pelanggan	0: industri majalah 1: industri koran
X ₃ Firm Size	Nonmetric	Ukuran karyawan dari perusahaan pelanggan	0: Perusahaan karyawan < 500 1: Perusahaan karyawan > 500
X ₄ Region	Nonmetric	Lokasi pelanggan	0: USA / Amerika Utara 1: di luar Amerika Utara
X ₅ Distribution System	Nonmetric	Cara produk kertas terjual ke pelanggan	0: secara tidak langsung 1: secara langsung
Performance Perceptions Variables Persepsi pelanggan terhadap performa beberapa aspek dari HBAT			
X ₆ Product Quality	Metric	Presepsi atas tingkat kualitas produk kertas HBAT	[0, 10]: poor hingga excellent
X ₇ E-Commerce Activities/Web Site	Metric	Pandangan atas situs web HBAT khususnya keramahan pengguna	[0, 10]: poor hingga excellent
X ₈ Technical Support	Metric	Dukungan teknis ditawarkan untuk memecahkan masalah produk/layanan	[0, 10]: poor hingga excellent
X ₉ Complaint Resolution	Metric	Keluhan diselesaikan secara tepat waktu dan lengkap	[0, 10]: poor hingga excellent
X ₁₀ Advertising	Metric	Persepsi atas iklan HBAT di semua jenis media	[0, 10]: poor hingga excellent

X11 Product Line	Metric	Kedalaman dan keluasan <i>product line</i> untuk kebutuhan pelanggan	[0, 10]: poor hingga excellent
X12 Salesforce Image	Metric	Keseluruhan pandangan pelanggan terhadap karyawan sales	[0, 10]: poor hingga excellent
X13 Competitive Price	Metric	Menawarkan harga kompetitif	[0, 10]: poor hingga excellent
X14 Warranty and Claims	Metric	Mendukung garansi dan klaim produk / layanannya	[0, 10]: poor hingga excellent
X15 New Products	Metric	Mengembangkan dan menjual produk baru	[0, 10]: poor hingga excellent
X16 Ordering and Billing	Metric	Persepsi ordering dan billing dilakukan secara efisien dan benar	[0, 10]: poor hingga excellent
X17 Price Flexibility	Metric	kesediaan perwakilan sales HBAT dalam menegosiasikan harga produk kertas dengan pelanggan	[0, 10]: poor hingga excellent
X18 Delivery Speed	Metric	Jumlah waktu diperlukan untuk mengirimkan produk setelah pesanan dikonfirmasi	[0, 10]: poor hingga excellent
<p style="text-align: center;">Outcome/Relationship Measures Pengukuran spesifik yang menunjukkan relasi pembelian dengan HBAT</p>			
X19 Satisfaction	Metric	Kepuasan pelanggan dengan pembelian yang telah terjadi	Rating dengan skala 10
X20 Likelihood of Recommendation	Metric	Kemungkinan pelanggan merekomendasikan produk HBAT kepada perusahaan lain	Rating dengan skala 10
X21 Likelihood of Future Purchase	Metric	Kemungkinan pelanggan melakukan pembelian ulang kedepannya	Rating dengan skala 10
X22 Current Purchase/Usage Level	Metric	Persentase pemenuhan kebutuhan kertas perusahaan pelanggan yang berasal dari HBAT	Ukuran 100 persen
X23 Consider Strategic Alliance/Partnership in Future	Nonmetric	Pelanggan memerpetimbangkan perusahaannya akan terlibat dalam aliansi/kemitraan strategis dengan HBAT	0: tidak mempertimbangkan 1: ya, mempertimbangkan aliansi strategis atau mitra

NOMOR 1

SOAL

Buat analisis Clustering dengan Hierarchical Clustering sampai dengan 3 Klaster untuk data sheet "HBAT_200" dengan pendekatan single linkage, complete linkage, dan average linkage dengan menggunakan variabel X_6 - X_{21} !

DIKETAHUI

Analisis : hierarchical clustering
Jumlah cluster : 3
Metode Clustering : Single, Complete, dan Average Linkage
Fitur : tabel "HBAT_200", kolom X_6 - X_{21}

JAWAB

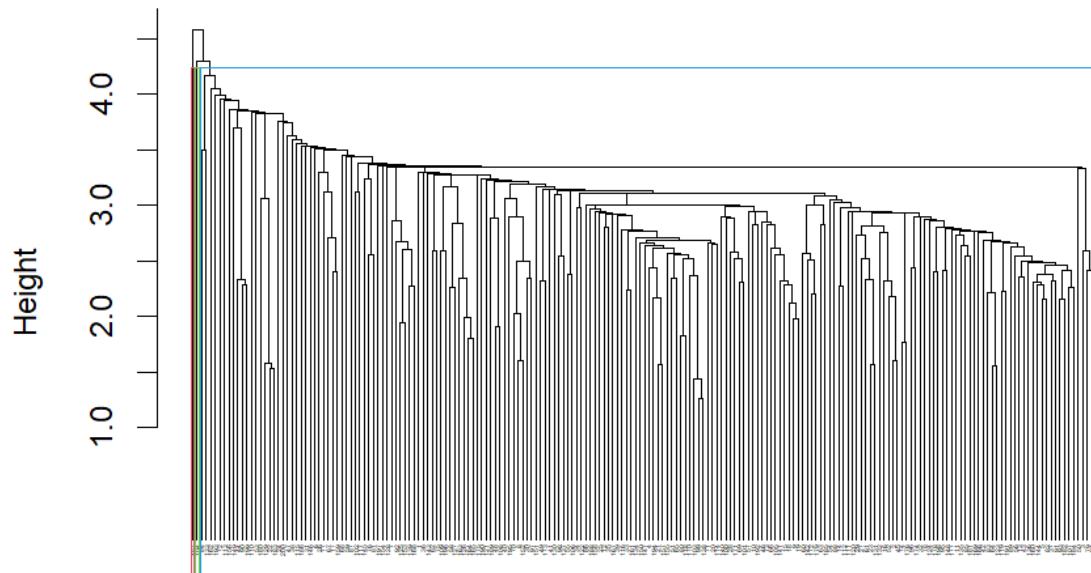
Dalam melakukan hierarchical clustering, perlu dilakukan beberapa langkah, yaitu:

1. Membuat data frame berisi fitur X_6 - X_{21}
2. Membuat distance matrix dengan fungsi `dist()`. Secara default perhitungan jarak yang digunakan adalah euclidean distance
3. Membuat hierarchical clustering dengan metode single, complete, dan average linkage. Pembuatan model menggunakan fungsi `hclust()`, dengan merubah nilai parameter "method" menjadi "single", "complete", atau "average"
4. Memotong histogram sesuai jumlah cluster yang diinginkan, yakni 3. pemotongan dapat dilakukan menggunakan fungsi `cutree()` dengan parameter $k = 3$. Dari hasil pemotongan tersebut didapatkan persebaran data pada masing-masing cluster, sebagai berikut:

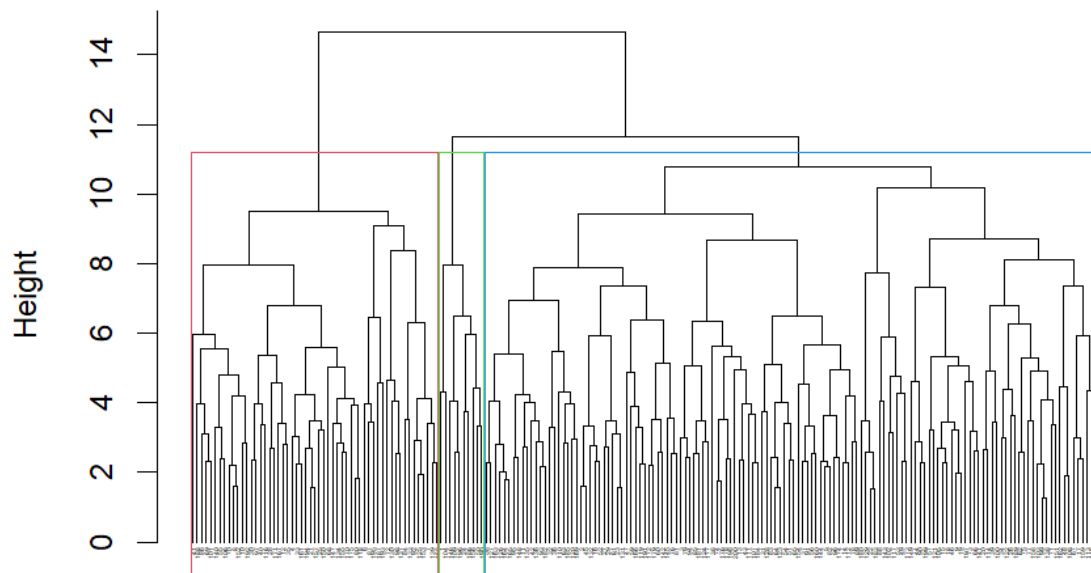
Metode	Cluster		
	1	2	3
Single	198	1	1
Complete	10	135	55
Average	3	189	8

5. Membuat dendrogram hasil clustering.

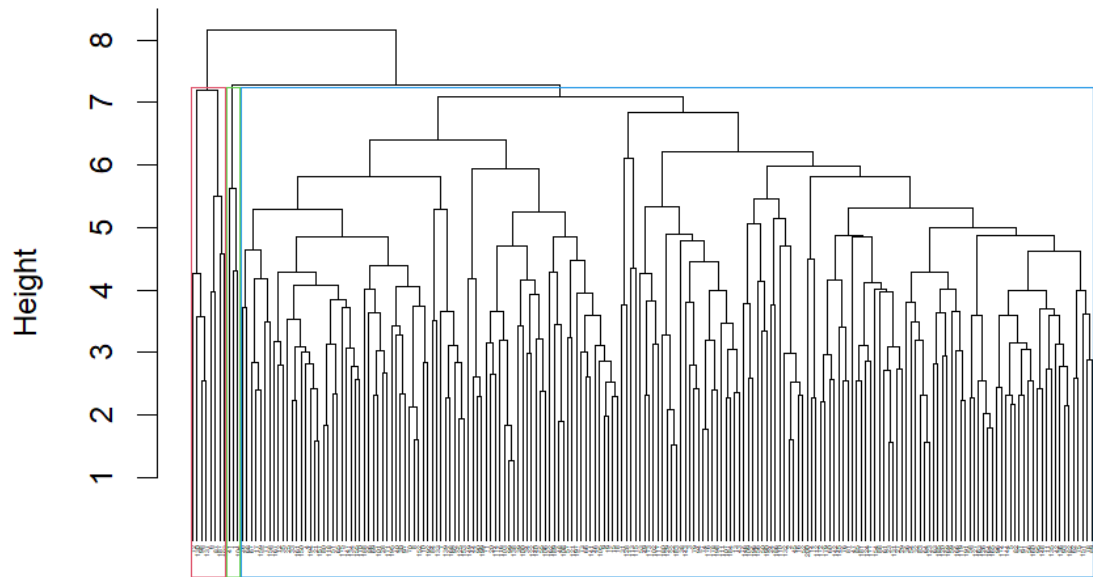
Dendrogram Agglomeratif Single



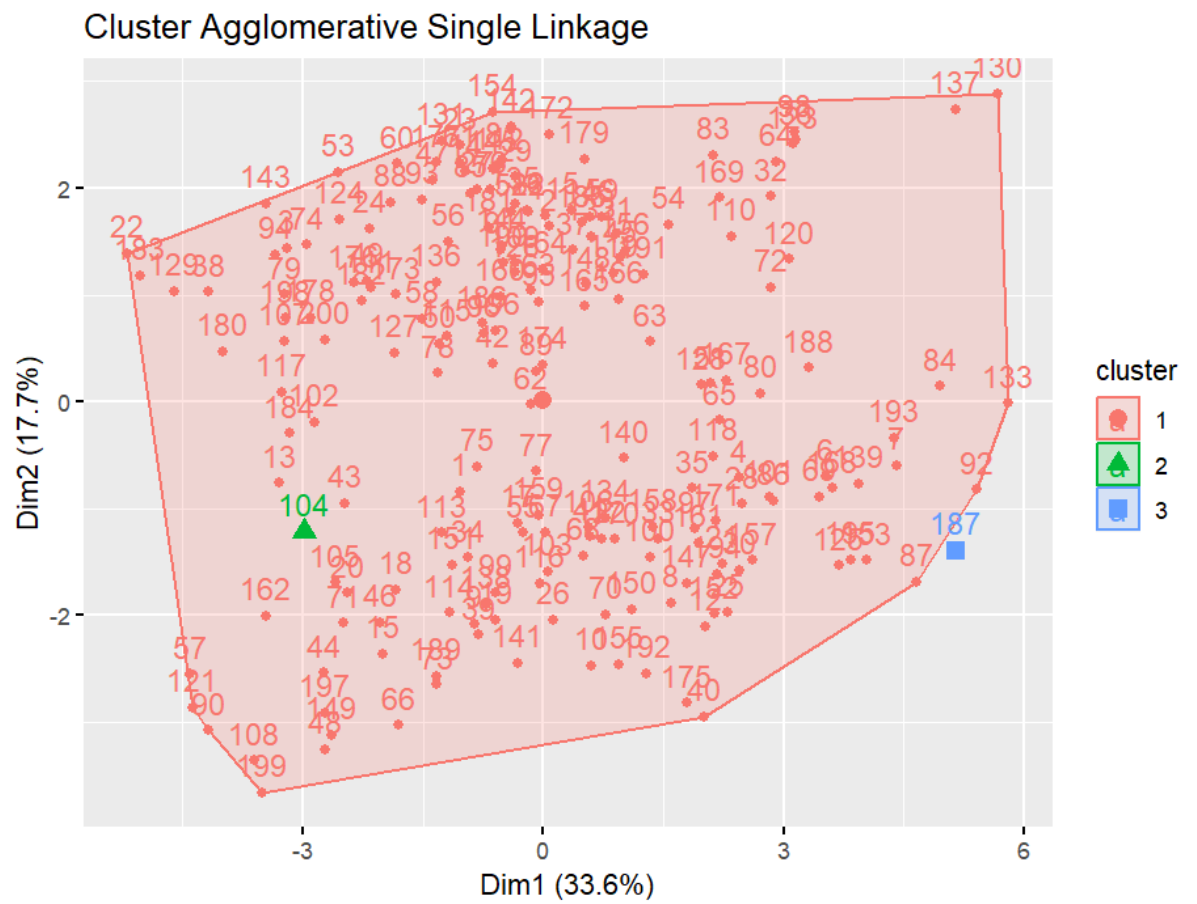
Dendrogram Agglomeratif Complete



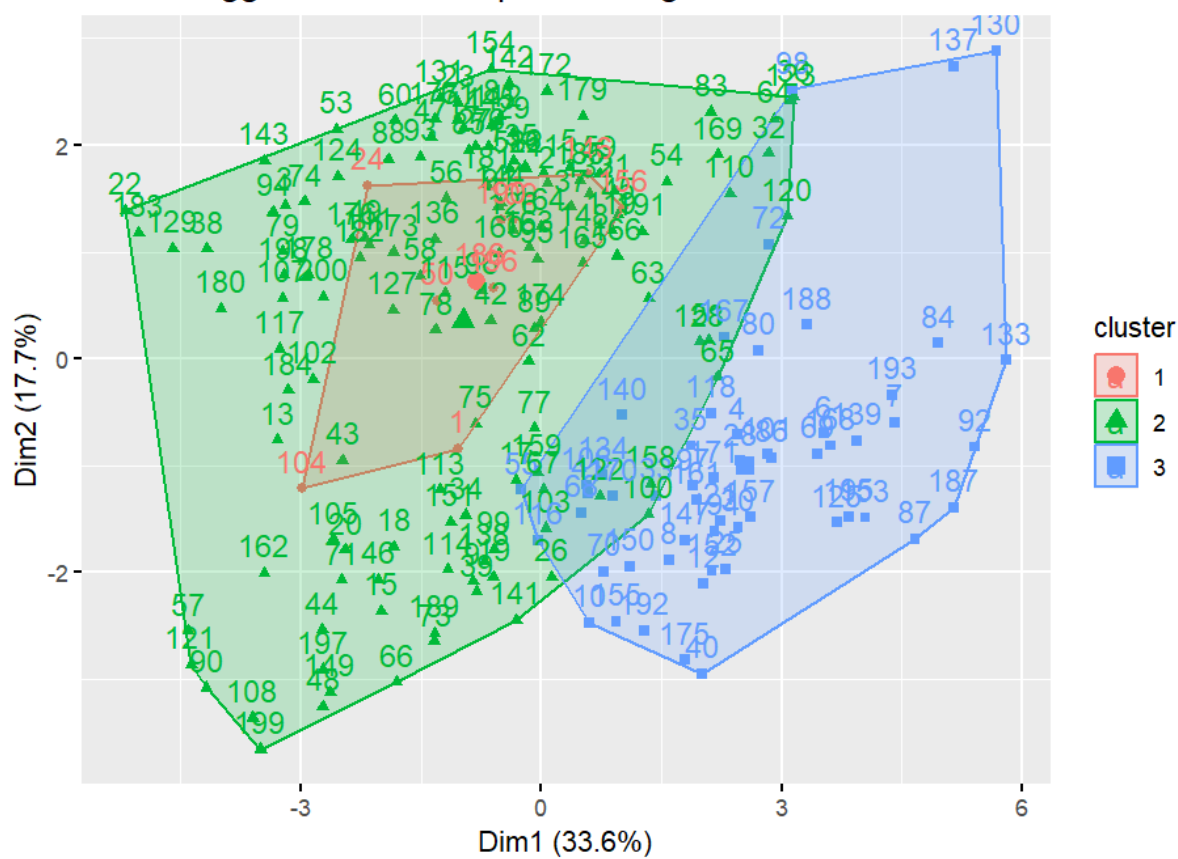
Dendrogram Agglomeratif Average

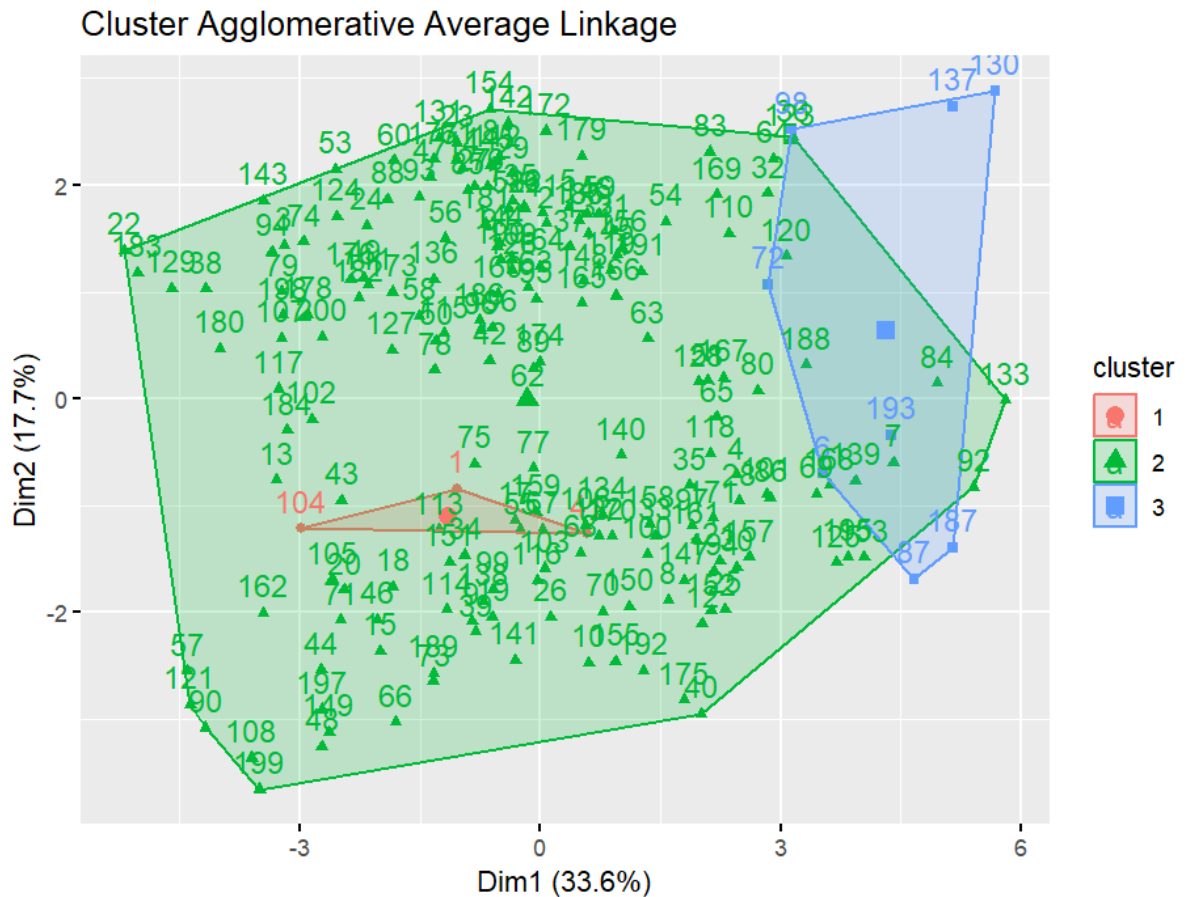


6. Visualisasi cluster dengan fviz_cluster



Cluster Agglomerative Complete Linkage





7. Melakukan evaluasi hasil clustering menggunakan silhouette score. Hal ini dapat dilakukan menggunakan fungsi `silhouette()`, nilai silhouette dari ketiga metode, sebagai berikut:

- Single: 0.07968901
- Complete: -0.02672805
- Average: 0.05301837

Dari nilai silhouette di atas, dapat disimpulkan bahwa metode hierarchichal clustering yang terbaik adalah Single Linkage, lalu dilanjutkan dengan Average Linkage, lalu Complete Linkage. Meski begitu, skor silhouette yang didapatkan masih sangat jauh dari nilai terbaik, yakni 1.

KESIMPULAN

Dari percobaan yang telah dilakukan, dapat ditarik beberapa kesimpulan, antara lain:

- Penggunaan metode hierarchichal clustering yang berbeda, akan menghasilkan clustering dan dendogram yang berbeda.
- Ketiga metode masih menghasilkan clustering dengan terjadi overlay antar cluster.

- Dalam kasus dataset HBAT_200 dengan fitur X6 - X21 didapati bahwa Single Linkage merupakan model clustering yang terbaik karena memiliki silhouette score terbaik, yakni 0.07968901. Metode kedua terbaik adalah Average dengan silhouette score -0.02672805. Sedangkan, metode terakhir adalah Complete Linkage dengan silhouette score -0.02672805.

NOMOR 2

SOAL

Buat analisis Clustering dengan k-means Clustering sampai dengan 3 Klaster untuk data sheet "HBAT_200" dengan menggunakan variabel X6 - X21!

DIKETAHUI

data : HBAT_200, fitur X6 - X21

metode : K-Means

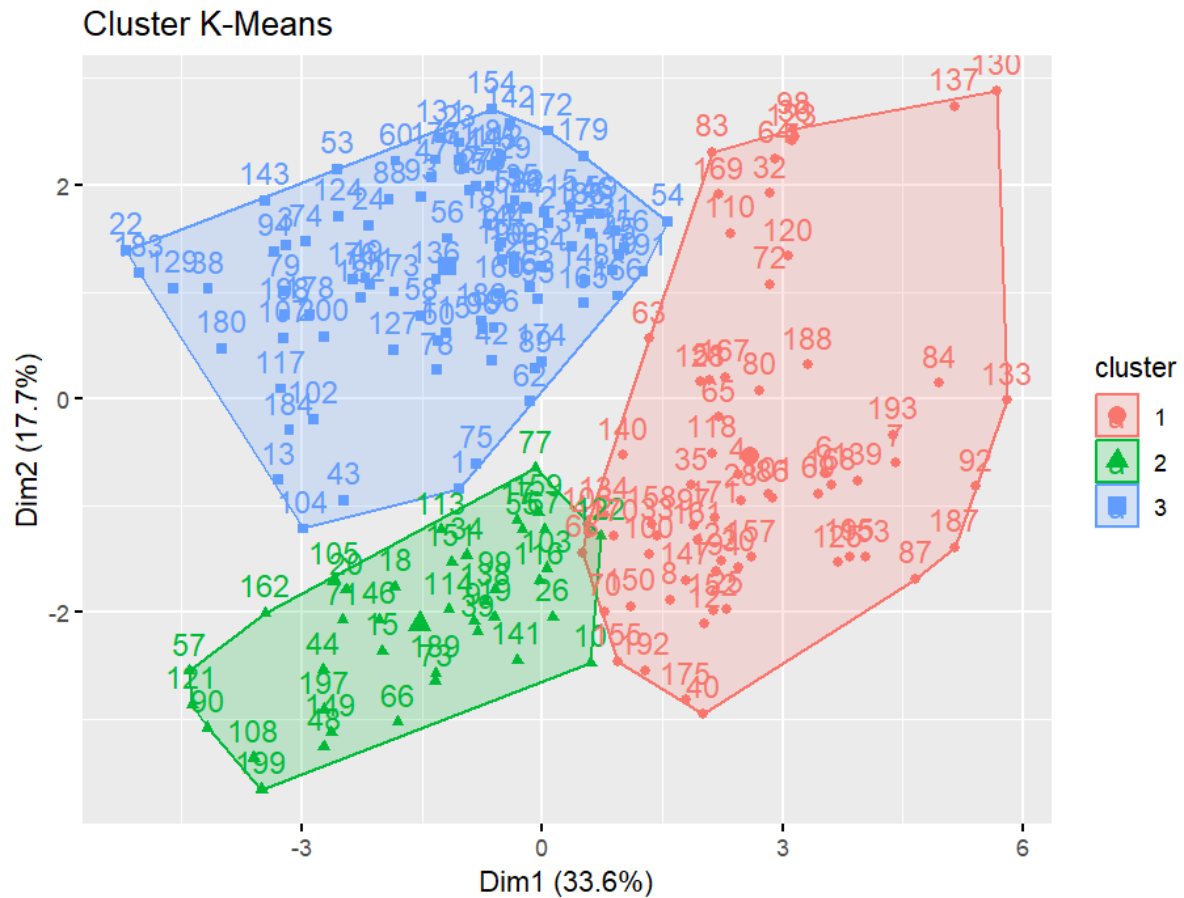
jumlah cluster : 3

JAWAB

Dalam melakukan clustering dengan K-Means, perlu dilakukan beberapa langkah, yaitu:

1. Membuat data frame berisi fitur X6 - X21
2. Membuat model K-Means menggunakan fungsi `kmeans()` dengan parameter `centers = 3` (banyak cluster yang ingin dibentuk) dan `nstart = 1` (random sets awal masing-masing cluster)

3. Membuat visualisasi penyebaran data pada cluster



4. Melakukan evaluasi hasil clustering menggunakan silhouette score. Dari percobaan di atas, didapatkan skor silhouette adalah 0.2665846

KESIMPULAN

- Metode K-Means berhasil membuat model cluster dimana tidak terjadi banyak overlay pada cluster-cluster yang terbentuk
- Pembuatan model clustering dengan K-Means menghasilkan skor silhouette yang terhitung baik, yakni 0.2665846. Nilai ini juga melebihi skor silhouette pada hierarchal clustering, sehingga dapat dikatakan bahwa K-Means menghasilkan model cluster yang lebih baik dibanding hierarchal clustering.

NOMOR 3

SOAL

Buat analisis discriminant (Prediksi Grup) dengan menggunakan Fungsi Diskriminan dan Fungsi Klasifikasi untuk data "HBAT" dengan menggunakan variabel prediktor X17-X22 dan variabel respon (label kelas) X5. Lakukan split data-training dan data-testing dengan proporsi 70% dan 30% ! Bandingkan akurasi untuk kedua fungsi !

DIKETAHUI

Metode : analisis diskriminan dengan fungsi klasifikasi dan diskriminan
Prediktor : $X_{17} - X_{22}$
Target : X_5 (Distribution System)
Split : 70% training, 30% testing

JAWAB

Dalam melakukan analisis diskriminan perlu dilakukan beberapa langkah berikut:

1. Membuat data frame berisi variabel prediktor dan target
2. Membuat correlation matrix dan melakukan feature selection (bila diperlukan).

Adapun, nilai korelasi pada percobaan ini adalah:

	X17	X18	X19	X20	X21	X22	X5
X17	1.00	0.51	0.03	0.05	0.00	0.08	-0.22
X18	0.51	1.00	0.63	0.47	0.43	0.70	0.25
X19	0.03	0.63	1.00	0.76	0.73	0.71	0.55
X20	0.05	0.47	0.76	1.00	0.66	0.55	0.47
X21	0.00	0.43	0.73	0.66	1.00	0.51	0.40
X22	0.08	0.70	0.71	0.55	0.51	1.00	0.27
X5	-0.22	0.25	0.55	0.47	0.40	0.27	1.00

Dalam percobaan ini sendiri tidak ada fitur yang di-drop karena nilai korelasi variabel prediktor dengan target, yakni X_5 sudah cukup baik dan tidak terdapat banyak fitur yang dapat dimanfaatkan

3. Melakukan analisis MANOVA sebagai pemenuhan asumsi pada LDA. Adapun, dalam MANOVA dilakukan beberapa hal, yaitu:
 - a. merubah kolom X_5 sebagai factor
 - b. memenuhi asumsi MANOVA, yakni:

- asumsi normalitas multivariat: pada 5 dari 6 variabel prediktor, didapatkan kesimpulan bahwa H_0 ditolak, sehingga data bukan berdistribusi normal multivariat
- asumsi kesamaan matriks kovarian: tidak dapat ditentukan dikarenakan program r yang berjalan lama dan tidak mengeluarkan hasil

c. Membuat model MANOVA dengan fungsi manova()

- analisis secara simultan:

Test	Nilai	F Hitung	F Tabel	Kesimpulan
Pillai	0.40401	21.805	3.691	tolak H_0
Roy	0.67788	21.805	3.691	tolak H_0
Wilks	0.59599	21.805	3.691	tolak H_0
Hotelling-Lawley	0.67788	21.805	3.691	tolak H_0

Kesimpulan: tolak H_0 , yakni X5 (Distribution System) memiliki pengaruh signifikan terhadap X17 - X22

- analisis secara parsial:


```

> summary.aov(manova_res)
Response 1 :
              Df Sum Sq Mean Sq F value    Pr(>F)
df_classification$X5  1  13.282  13.2825   9.7485 0.002063 **
Residuals           198 269.778   1.3625
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Response 2 :
              Df Sum Sq Mean Sq F value    Pr(>F)
df_classification$X5  1   7.212   7.2115  13.659 0.0002836 ***
Residuals           198 104.537   0.5280
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Response 3 :
              Df Sum Sq Mean Sq F value    Pr(>F)
df_classification$X5  1  92.30  92.300  85.303 < 2.2e-16 ***
Residuals           198 214.24   1.082
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Response 4 :
              Df Sum Sq Mean Sq F value    Pr(>F)
df_classification$X5  1  50.665  50.665  54.91 3.577e-12 ***
Residuals           198 182.694   0.923
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Response 5 :
              Df Sum Sq Mean Sq F value    Pr(>F)
df_classification$X5  1  25.396  25.3959   37.7 4.432e-09 ***
Residuals           198 133.379   0.6736
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Response 6 :
              Df Sum Sq Mean Sq F value    Pr(>F)
df_classification$X5  1 1206.3 1206.26  16.147 8.324e-05 ***
Residuals           198 14791.7   74.71
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Kesimpulan:

- X17 (Price Flexibility)
X5 (Distribution System) berpengaruh signifikan terhadap X17
- X18 (Delivery Speed)
X5 berpengaruh signifikan terhadap X18
- X19 (Satisfaction)
X5 tidak berpengaruh signifikan terhadap X19
- X20 (Likelihood of Recommendation)
X5 berpengaruh signifikan terhadap X20
- X21 (Likelihood of Future Purchase)

X5 berpengaruh signifikan terhadap X21

- X22 (Current Purchase/Usage Level)

X5 berpengaruh signifikan terhadap X22

- d. Kesimpulan: telah memenuhi asumsi dan dapat dilakukan analisis diskriminan
4. Splitting data menjadi training (70%) dan testing (30%) dengan menggunakan fungsi `initial_split()` dengan parameter `prop = 0.7`
 5. Membuat model lda (diskriminan function) dan qda (classification function) dengan dataset training

LDA

```
Call:
lda(X5 ~ ., data = train)

Prior probabilities of groups:
      0      1
0.5395683 0.4604317

Group means:
      X17      X18      X19      X20      X21      X22
0 4.745333 3.676000 6.404000 6.534667 7.366667 56.35333
1 4.298438 3.992188 7.653125 7.553125 8.085938 60.80313

Coefficients of linear discriminants:
      LD1
X17 -0.555030708
X18  0.569262303
X19  0.832386228
X20  0.349126348
X21  0.002846859
X22 -0.071132391
```

QDA

```
> qda_model = qda(X5~., data=train)
> qda_model
Call:
qda(X5 ~ ., data = train)

Prior probabilities of groups:
      0      1
0.5395683 0.4604317

Group means:
      X17      X18      X19      X20      X21      X22
0 4.745333 3.676000 6.404000 6.534667 7.366667 56.35333
1 4.298438 3.992188 7.653125 7.553125 8.085938 60.80313
```

6. Melakukan prediksi pada dataset testing
7. Melakukan evaluasi dengan confusion matrix dan menghitung akurasi

LDA

akurasi: 0.8688525

```
> conf_matLDA
      Truth
Prediction 0  1
      0 28  3
      1  5 25
```

QDA

akurasi: 0.8360656

```
> conf_matQDA
      Truth
Prediction 0  1
      0 26  3
      1  7 25
```

KESIMPULAN

- Pada dataset tersebut dapat dilakukan analisis diskriminan dikarenakan telah memenuhi uji MANOVA
- Kedua metode bisa mendapatkan model dengan nilai akurasi yang cukup tinggi. LDA memiliki nilai akurasi yang lebih tinggi, yakni 0.8688525. Sedangkan, QDA memiliki akurasi 0.8360656.
- Kedua metode mampu menangani klasifikasi pada kasus binary-class dengan baik.

NOMOR 4

SOAL

Buat analisis discriminant (Prediksi Grop) dengan menggunakan Fungsi Diskriminan dan Fungsi Klasifikasi untuk data "HBAT" dengan menggunakan variabel prediktor X17-X22 dan variabel respon (label kelas) X1. Lakukan split data-training dan data-testing dengan proporsi 70% dan 30% ! Bandingkan akurasi untuk kedua fungsi !

DIKETAHUI

Metode : analisis diskriminan dengan fungsi klasifikasi dan diskriminan
Prediktor : $X_{17} - X_{22}$
Target : X_1 (Customer Type)
Split : 70% training, 30% testing

JAWAB

Dalam melakukan analisis diskriminan perlu dilakukan beberapa langkah berikut:

1. Membuat data frame berisi variabel prediktor dan target
2. Membuat correlation matrix dan melakukan feature selection (bila diperlukan).

Adapun, nilai korelasi pada percobaan ini adalah:

```
> cor_matrix
```

	X17	X18	X19	X20	X21	X22	X1
X17	1.00	0.51	0.03	0.05	0.00	0.08	-0.08
X18	0.51	1.00	0.63	0.47	0.43	0.70	0.57
X19	0.03	0.63	1.00	0.76	0.73	0.71	0.71
X20	0.05	0.47	0.76	1.00	0.66	0.55	0.53
X21	0.00	0.43	0.73	0.66	1.00	0.51	0.56
X22	0.08	0.70	0.71	0.55	0.51	1.00	0.83
X1	-0.08	0.57	0.71	0.53	0.56	0.83	1.00

Dalam percobaan ini sendiri tidak ada fitur yang di-drop karena nilai korelasi variabel prediktor dengan target, yakni X_1 sudah cukup baik kecuali pada X17.

8. Melakukan analisis MANOVA sebagai pemenuhan asumsi pada LDA. Adapun, dalam MANOVA dilakukan beberapa hal, yaitu:
 - e. merubah kolom X1 sebagai factor
 - f. memenuhi asumsi MANOVA, yakni:

- asumsi normalitas multivariat: pada 5 dari 6 variabel prediktor, didapatkan kesimpulan bahwa H_0 ditolak, sehingga data bukan berdistribusi normal multivariat
- asumsi kesamaan matriks kovarian: tidak dapat ditentukan dikarenakan program r yang berjalan lama dan tidak mengeluarkan hasil

g. Membuat model MANOVA dengan fungsi manova()

- analisis secara simultan:

Test	Nilai	F Hitung	F Tabel	Kesimpulan
Pillai	1.0408	34.906	2.310	tolak H_0
Roy	2.8657	92.181	3,691	tolak H_0
Wilks	0.1812	43.174	2,310	tolak H_0
Hotelling-La wley	3.2933	52.418	2,310	tolak H_0

Kesimpulan: tolak H_0 , yakni X_1 (Customer Type) memiliki pengaruh signifikan terhadap $X_{17} - X_{22}$

- analisis secara parsial:

```

> summary.aov(manova_res)
Response 1 :
              Df Sum Sq Mean Sq F value    Pr(>F)
df_classification$X1  2  64.65  32.325  29.156 8.096e-12 ***
Residuals           197 218.41   1.109
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Response 2 :
              Df Sum Sq Mean Sq F value    Pr(>F)
df_classification$X1  2 48.596 24.2978  75.794 < 2.2e-16 ***
Residuals           197 63.153   0.3206
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Response 3 :
              Df Sum Sq Mean Sq F value    Pr(>F)
df_classification$X1  2 164.31  82.156 113.79 < 2.2e-16 ***
Residuals           197 142.23   0.722
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Response 4 :
              Df Sum Sq Mean Sq F value    Pr(>F)
df_classification$X1  2  71.043  35.521  43.112 2.953e-16 ***
Residuals           197 162.316   0.824
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Response 5 :
              Df Sum Sq Mean Sq F value    Pr(>F)
df_classification$X1  2  53.545 26.7726  50.121 < 2.2e-16 ***
Residuals           197 105.230   0.5342
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Response 6 :
              Df Sum Sq Mean Sq F value    Pr(>F)
df_classification$X1  2 10997.5  5498.8  216.63 < 2.2e-16 ***
Residuals           197  5000.5   25.4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Kesimpulan:

- X17 (Price Flexibility)
- X1 (Customer Type) berpengaruh signifikan terhadap X17
- X18 (Delivery Speed)

X1 tidak berpengaruh signifikan terhadap X18

- X19 (Satisfaction)

X1 tidak berpengaruh signifikan terhadap X19

- X20 (Likelihood of Recommendation)

X1 berpengaruh signifikan terhadap X20

- X21 (Likelihood of Future Purchase)

X1 tidak berpengaruh signifikan terhadap X21

- X22 (Current Purchase/Usage Level)

X1 tidak berpengaruh signifikan terhadap X22

h. Kesimpulan: telah memenuhi asumsi dan dapat dilakukan analisis diskriminan

9. Splitting data menjadi training (70%) dan testing (30%) dengan menggunakan fungsi `initial_split()` dengan parameter `prop = 0.7`

10. Membuat model `lda` (diskriminan function) dan `qda` (classification function) dengan dataset training

LDA

```

> lda_model
Call:
lda(X1 ~ ., data = train)

Prior probabilities of groups:
      1      2      3
0.3405797 0.3188406 0.3405797

Group means:
      X17      X18      X19      X20      X21      X22
1 4.276596 3.106383 5.640426 6.155319 6.991489 49.33404
2 5.465909 4.284091 7.259091 7.193182 7.868182 58.50909
3 3.980851 4.131915 7.729787 7.474468 8.097872 66.50426

Coefficients of linear discriminants:
      LD1      LD2
X17 -0.42766497 -0.73786029
X18  0.35040832 -0.51766913
X19  0.33373287 -0.41647666
X20 -0.02291219  0.05802764
X21  0.12037397 -0.21705813
X22  0.14981751  0.07754995

Proportion of trace:
      LD1      LD2
0.8252 0.1748

```

QDA

```

> qda_model
Call:
qda(X1 ~ ., data = train)

Prior probabilities of groups:
      1      2      3
0.3405797 0.3188406 0.3405797

Group means:
      X17      X18      X19      X20      X21      X22
1 4.276596 3.106383 5.640426 6.155319 6.991489 49.33404
2 5.465909 4.284091 7.259091 7.193182 7.868182 58.50909
3 3.980851 4.131915 7.729787 7.474468 8.097872 66.50426

```

11. Melakukan prediksi pada dataset testing
12. Melakukan evaluasi dengan confusion matrix dan menghitung akurasi

LDA

akurasi: 0.7741935

```
> conf_matLDA
      Truth
Prediction 1  2  3
      1 19  2  0
      2  2 13  5
      3  0  5 16
```

QDA

akurasi: 0.8387097

```
> conf_matQDA
      Truth
Prediction 1  2  3
      1 19  0  0
      2  2 16  4
      3  0  4 17
```

KESIMPULAN

- Pada dataset tersebut dapat dilakukan analisis diskriminan dikarenakan telah memenuhi uji MANOVA
- Kedua model memiliki perbedaan nilai akurasi yang cukup besar, yakni 0.06451613 dimana metode QDA lebih baik, yakni akurasi 0.8387097. Sedangkan, LDA hanya memiliki akurasi 0.7741935
- Pada kasus data multi-class, yakni variabel target memiliki lebih dari 2 nilai, metode QDA terbukti lebih baik dibandingkan dengan LDA

NOMOR 5

SOAL

Lakukan analisis Conjoint untuk data sheet HBAT_CONJOINT ! Lakukan analisis terhadap hasil !

DIKETAHUI

- metode : conjoint analysis
- dataset : HBAT_CONJOINT (dilakukan drop kolom QN karena identifier responden)
- dataset pendukung : HBAT_CPLAN (level dari masing-masing profile). dilakukan penghapusan baris 23 hingga 25 karena data rating PROD23 hingga PROD25 tidak tersedia pada HBAT_CONJOINT
- df_levels : data frame yang dibuat sendiri dengan isi 1 kolom dan 15 baris, yakni berisi nama level. Tidak ditemukan dokumentasi yang menjelaskan level masing-masing faktor. Sehingga, digunakan penomoran (contoh, facto MIXTURE, terdapat 3 level, sehingga diberi nama level MIXTURE1, MIXTURE2, dan MIXTURE3).

JAWAB dan KESIMPULAN

Diperlukan beberapa langkah dalam analisis conjoint ini, yaitu:

1. Membuat beberapa dataframe, yakni df_CONJOINT (daftar level setiap profile / product), df_RATING (daftar rating responden terhadap profile/product), dan df_levels (berisi nama level)
2. Menghitung preferensi pada seorang responden. Hal ini dapat dilakukan menggunakan fungsi caModel()

```

> caModel(df_RATING[1,], df_profile)

Call:
lm(formula = frm1)

Residuals:
    Min       1Q   Median       3Q      Max
-1.61193 -0.39641 -0.00372  0.39807  1.79210

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    4.66581    0.30074   15.514 2.64e-09 ***
factor(x$MIXTURE)1 -0.05556    0.33650   -0.165 0.871616
factor(x$MIXTURE)2  0.49094    0.30811    1.593 0.137061
factor(x$NUMAPP)1  0.39632    0.31434    1.261 0.231348
factor(x$NUMAPP)2  0.60975    0.30449    2.003 0.068353 .
factor(x$GERMFRE)1 -0.22526    0.23238   -0.969 0.351499
factor(x$BIOPROT)1  0.48763    0.23768    2.052 0.062696 .
factor(x$PRICE)1   1.46701    0.30984    4.735 0.000485 ***
factor(x$PRICE)2   0.87239    0.31691    2.753 0.017511 *
factor(x$STATUS_)1 -0.53104    0.31302   -1.697 0.115552
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.01 on 12 degrees of freedom
Multiple R-squared:  0.8941,    Adjusted R-squared:  0.8146
F-statistic: 11.25 on 9 and 12 DF,  p-value: 0.0001371

```

Hasil analisis:

- Responden 1 memiliki persepsi positif yang kuat terhadap level NUMAPP2, PRICE1, PRICE2
 - Responden 1 memiliki persepsi negatif terhadap level MIXTURE1, GERMFRE1, dan STATUS_1
3. Nilai importance masing-masing faktor dari semua responden bisa didapatkan menggunakan fungsi caImportance()

```

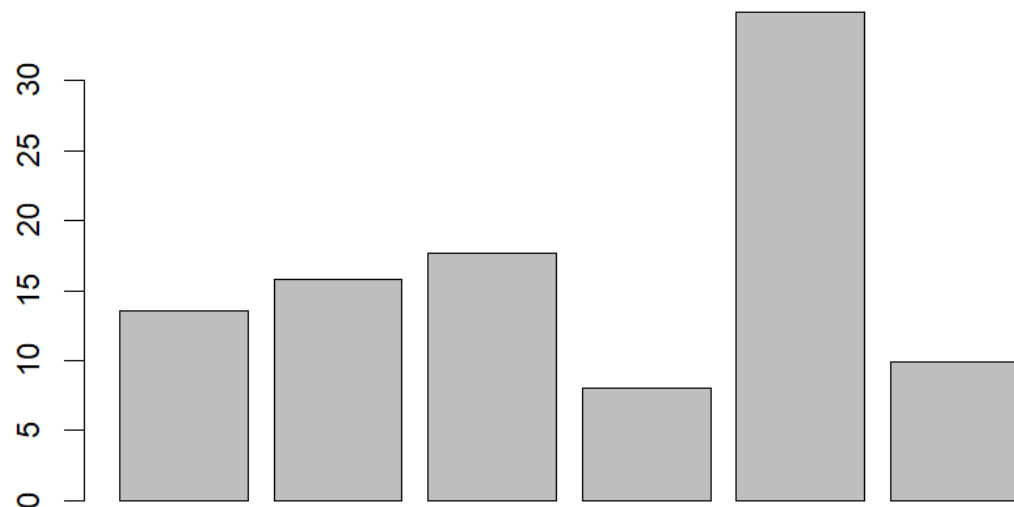
> importance
[1] 13.59 15.83 17.69  8.05 34.95  9.89

```

Hasil analisis:

- Urutan factor yang paling berpengaruh kepada keseluruhan responden adalah:
 1. PRICE: 34.95
 2. GERMFRE: 17.69
 3. NUMAPP: 15.83

4. MIXTURE: 13.59
 5. STATUS: 9.89
 6. BIOPROT: 8.05
- Factor PRICE memiliki nilai tertinggi dan terlampaui jauh dibanding faktor lain. Sehingga, perusahaan HBAT disarankan untuk mengeluarkan produk dengan faktor utama adalah PRICE



4. Part-Worth utilities dengan menimbang semua responden bisa didapatkan menggunakan fungsi `caPartUtilities()`. Dari hasil tersebut didapatkan urutan

```
> list_profile
[1] 5.349 5.949 6.316 7.110 6.680 7.589 6.290 7.471 6.800 6.044 4.703
[12] 6.080 6.114 6.291 6.244 7.357 3.573 5.883 5.225 7.387 6.874 4.631
```

Dari hal ini, bisa didapatkan urutan:

- 1: PROD5
- 2: PROD7
- 3: PROD14
- 4: PROD22
- 5: PROD15
- 6: PROD18

7: PROD12
 8: PROD21
 9: PROD16
 10: PROD8
 11: PROD3
 12: PROD9
 13: PROD10
 14: PROD13
 15: PROD11
 16: PROD19
 17: PROD1
 18: PROD6
 19: PROD4
 20: PROD20
 21: PROD17
 22: PROD2

Maka, perusahaan HBAT disarankan untuk mempertimbangkan memproduksi/meluncurkan produk 5, 7, dan 14

5. urutan level yang paling penting bagi keseluruhan responden

```
> list_level
[1] 4.666 5.539 4.075 6.532 2.980 2.021 2.355 4.465 3.693 3.452 3.812
[12] 4.530 2.924 4.196 4.417
```

MIXTURE1: 6
 MIXTURE2: 7
 MIXTURE3: 13
 NUMAPP1: 5
 NUMAPP2: 10
 NUMAPP3: 9
 GERMFRE1: 11
 GERMFRE2: 3
 BIOPROT1: 14
 BIOPROT2: 15
 PRICE1: 8

PRICE2: 12

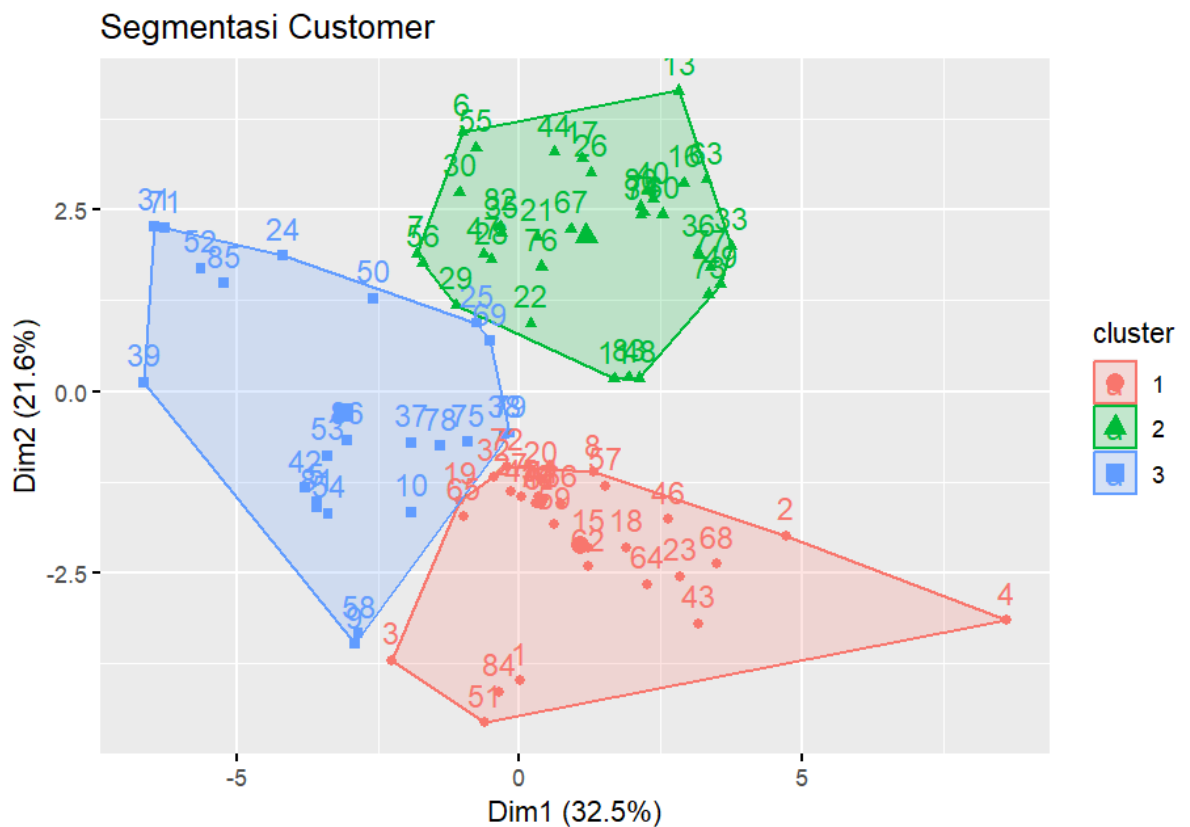
PRICE3: 1

STATUS1: 2

STATUS2: 4

level dengan rating tertinggi adalah level PRICE3, STATUS2, dan GERFRE2. Sehingga, perusahaan HBAT disarankan untuk mengeluarkan produk dengan mempertimbangkan karakteristik faktor tersebut

6. segmentasi customer menjadi 3 cluster menggunakan caSegmentation()



SARAN

- dengan menimbang hasil tersebut diharapkan perusahaan HBAT dapat memiliki gambaran yang cukup terkait pasar yang tersedia, dimana telah didapatkan kesimpulan produk, faktor, dan level yang memiliki persepsi positif di masyarakat.