

PENGUNAAN ANALISIS DISKRIMINAN PADA KANDUNGAN MINYAK MENTAH PADA BATU PASIR

THE USE OF DISCRIMINANT ANALYSIS FOR SAND STONE SUBPOPULATION CLASSIFICATION BASED ON CRUDE OIL CONTENT

Teguh Yuniarko

Program Studi Matematika, Fakultas Sains dan Teknologi Informasi,
Institut Sains dan Teknologi Nasional,
Jl. Moh. Kahfi II, Bhumi Srengseng Indah, Jagakarsa, Jakarta Selatan 12640
Telp. (021) 7874647, Fax. (021) 7866955
teguhyuniarko@istn.ac.id

ABSTRACT

To determine the independent discriminant analysis of crude oil, two populations of sand stone was utilized. The rock contains five variables, namely crude oil *Vanadium*, *Iron*, *Beryllium*, *Hydrocarbon*, and *Aromatic Hydrocarbon Saturate*. Data was analyzed from *Hydrocarbon*, *Iron*, and *Vanadium*. These three variables are used to inform the discriminant function and classification process. The classification result obtained using the SPSS program, showed that the discriminant function is significant and with a validated test accuracy of 93.9 %.

Keywords: *Discriminant Analysis, Classification, Multivariate, Discriminant Function*

ABSTRAK

Penerapan analisis diskriminan kandungan minyak mentah pada dua populasi batu pasir adalah menentukan variabel independen yang secara signifikan dapat membedakan antara batu yang mengandung lima variabel minyak mentah yaitu Vanadium, Iron, Beryllium, Saturade Hydrocarbon dan Aromatic Hydrocarbon. Hasil analisis data dari kelima variabel independen tersebut, variabel yang dapat membedakan antara kedua populasi batu pasir adalah variabel Saturade Hydrocarbon, Iron dan Vanadium. Oleh karena itu, ketiga variabel tersebut yang digunakan dalam membentuk fungsi diskriminan dan proses klasifikasi. Hasil klasifikasi menggunakan program SPSS menunjukkan bahwa fungsi diskriminan tersebut signifikan dan uji validasi fungsi diskriminan diperoleh ketepatan klasifikasi sebesar 93.9%.

1. PENDAHULUAN

Perkembangan pengetahuan meng-alami fenomena kompleksitas yang mem-butuhkan penyelidikan lebih mendalam untuk mengklasifikasikan penelitian ke banyak variabel yang berbeda. Klasifikasi biasanya membutuhkan struktur masalah lebih dari diskriminasi. Ketika melakukan klasifikasi, sangat mungkin akan terjadi kesalahan pengklasifikasian objek/observasi.

Prosedur pengklasifikasian yang baik harus memperkecil peluang terjadinya misklasifikasi. Dengan kata lain, pro-babilitas misklasifikasi harus kecil. Oleh karena itu maka perlu dilakukan evaluasi terhadap prosedur klasifikasi. Seorang bernama Ronal A.Fisher mengembangkan sebuah metode fungsi diskriminan. Fisher menyatakan bahwa apabila dua atau lebih populasi telah diukur dalam beberapa karakter X_1, X_2, \dots, X_p , maka dapat dibangun fungsi linear tertentu dari pengukuran itu dimana fungsi itu merupakan fungsi pembeda (pemisah) terbaik bagi populasi -populasi yang dipelajari. Fungsi linear yang dibangun itu disebut sebagai fungsi diskriminan.

Analisis diskriminan adalah teknik statistika yang digunakan untuk mengelaskan individu atau

obyek ke dalam suatu kelompok atau kelas berdasarkan sekumpulan peubah bebas .(*Dillon dan Goldstein, 1984*).

Analisis diskriminan digunakan untuk membuat satu model prediksi keanggotaan kelompok yang didasarkan pada karakteristik-karakteristik yang diobservasi untuk masing-masing kasus. Prosedur ini akan menghasilkan fungsi diskriminan yang didasarkan pada kombinasi-linier yang berasal dari variabel-variabel bebas yang dapat menghasilkan perbedaan paling baik antara kelompok-kelompok yang dianalisis.

Penerapan analisis diskriminan pada bidang kesehatan diantaranya untuk memisahkan populasi pasien

depresi.¹ aplikasi pada bidang lainnya yaitu untuk menganalisis faktor yang mempengaruhi loyalitas pelanggan kartu seluler CDMA dan GSM. [Ar.Rasyid, Gunadarma], analisis Data Survei Kepuasan Pelanggan PDAM Kota Bukittinggi Sumatera Barat Tahun 2005.[Mellyna, 2005], penggunaan fungsi diskriminan untuk mengklasifikasikan minyak mentah ke dalam familinya.²[Brian Hitchon,1984], dan penggunaan fungsi diskriminan untuk pendistribusian minyak mentah.³

Adapun dalam tulisan ini dibahas mengenai penggunaan analisis diskriminan pada dua kelompok populasi batu pasir (Kelompok Sub Mulinia dan Kelompok upper (Mulinia, second subscales, first subscales)) berdasarkan kandungan minyak mentah yang terdapat pada batu pasir.

2 ANALISIS

ANALISIS KOMPONEN UTAMA

Analisis Komponen Utama dian-taranya terdiri dari proses Kesamaan Kovariansi untuk dua populasi dengan matriks kovarians yang berbentuk :

$$S_{1(p \times p)} = \frac{1}{n_1-1} \sum_{j=1}^{n_1} (X_{1j} - \bar{X}_1)(X_{1j} - \bar{X}_1)'$$

$$S_{2(p \times p)} = \frac{1}{n_2-1} \sum_{j=1}^{n_2} (X_{2j} - \bar{X}_2)(X_{2j} - \bar{X}_2)'$$

Karena matriks kovarians Σ populasi 1 dengan populasi 2 dianggap sama ($\Sigma_1 = \Sigma_2$), maka matriks kovarians gabungan kedua populasinya menjadi :

$$S_G = \frac{(n_1-1)S_1 + (n_2-1)S_2}{(n_1+n_2-2)}$$

Sebelum kita membangun fungsi diskriminan, maka kita perlu menguji perbedaan vektor rata-rata apakah vektor rata-rata itu berbeda. Oleh karena itu, fungsi diskriminan pada dasarnya dibangun untuk menerangkan perbedaan di antara populasi.

Pengujian vektor rata-rata dilakukan untuk menguji perbedaan sifat-sifat diantara kedua populasi yang kita uji. Jika hasil hipotesisnya menolak H_0 , maka nilai vektor rata-rata dari kedua populasi tersebut berbeda, dengan demikian kita dapat membangun fungsi diskriminan untuk mengkaji perbedaan sifat-sifat yang ada di antara kedua populasi yang dipelajari. Jika sebaliknya maka fungsi diskriminan tidak layak dibangun karena tidak ada perbedaan sifat-sifat diantara kedua populasi yang kita uji.

Hipotesis :

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Statistik pengujianya :

T^2 -Hotelling

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{X}_1 - \bar{X}_2)' S_G^{-1} (\bar{X}_1 - \bar{X}_2)$$

Kaidah Penentuan Keputusannya, yaitu :

$$T^2 > \frac{(n_1+n_2-2)p}{n_1+n_2-p-1} F_{\alpha; v_1, v_2} \text{ maka tolak } H_0 \text{ atau } T^2 \leq$$

$$\frac{(n_1+n_2-2)p}{n_1+n_2-p-1} F_{\alpha; v_1, v_2} \text{ maka terima } H_0.$$

Keterangan :

$$F_{hitung} = \frac{n_1+n_2-p-1}{(n_1+n_2-2)p} T^2$$

$$v_1 = p$$

$$v_2 = n_1 + n_2 - p - 1$$

$$n_1 = \text{banyak sampel populasi 1;}$$

$$n_2 = \text{banyak sampel populasi 2}$$

$$p = \text{sifat dari populasi yang diuji.}$$

ANALISIS DISKRIMINAN

Asumsi-asumsi yang harus terpenuhi :

1. Sejumlah p variabel independen harus berdistribusi normal.
2. Matriks varian-kovarians variabel independen berukuran p x p pada kedua kelompok harus sama.

Fungsi diskriminan linear Fisher yang dapat dirumuskan sebagai berikut :

$$Y = (\bar{X}_1 - \bar{X}_2)' S_G^{-1} X$$

Keterangan :

X = vektor variabel acak yang diidentifikasi dalam model fungsi diskriminan

\bar{X}_1 = vektor nilai rata-rata variabel acak dari kelompok pertama.

\bar{X}_2 = vektor nilai rata-rata variabel acak dari kelompok kedua.

S_G^{-1} = invers matriks gabungan

S_G^{-1} = inverse of combined matrix

3. METODE DAN PEMBAHASAN

Metode yang digunakan dalam peneli-tian ini adalah studi pustaka.

Batu pasir merupakan bahan material yang mengandung berbagai jenis minyak mentah. Data pada tabel 1 merupakan data sekunder sampel minyak mentah dua kelompok populasi batu pasir dari tiga populasi yang diperoleh dari sumber literatur. [Johnson and Wichern, 1998]

Brian Hitchon (2), R. H. Filby (3)

AAPG Bulletin
Volume 68 (1984)

³ Distribution of Transition Metals in North Alaskan Oils, Joseph A. Curiale. Unocal Research, *Metal Complexes in Fossil Fuels*. 1987. American Chemical Society

¹ The Use of Discriminant Analysis to Separate a Study Population by Treatment Subgroups in a Clinical Trial with a New Pentapeptide Antidepressant*. John P. Feighner, MD Lev Sverdlov, PhD Innapharma, Inc. 1 Maynard Drive, Suite 205 Park Ridge, NJ 07656* This study was sponsored by Innapharma, Inc.

² Use of Trace Elements for Classification of Crude Oils into Families--Example from Alberta, Canada

Popuation group	Variable					
	X1	X2	X3	X4	X5	
1	5	47	0.07	7.06	6.1	
	3.4	32	0.2	5.82	4.69	
	1.2	12	0	5.54	3.15	
	8.4	17	0.07	6.31	4.55	
	4.2	36	0.5	9.25	4.95	
	4.2	35	0.5	5.69	2.22	
	3.9	41	0.1	5.63	2.94	
	3.9	36	0.07	6.19	2.27	
	7.3	32	0.3	8.02	12.92	
	4.4	46	0.07	7.54	5.76	
	3	30	0	5.12	10.77	
	2	6.3	13	0.5	4.24	8.27
		1.7	5.6	1	5.69	4.64
7.3		24	0	4.34	2.99	
7.8		18	0.5	3.92	6.09	
7.8		25	0.7	5.39	6.2	
7.8		26	1	5.02	2.5	
9.5		17	0.05	3.52	5.71	
7.7		14	0.3	4.65	8.63	
11		20	0.5	4.27	8.4	
8		14	0.3	4.32	7.87	
8.4		18	0.2	4.38	7.98	
10		18	0.1	3.06	7.67	
7.3		15	0.05	3.76	6.84	
9.5		22	0.3	3.98	5.02	
8.4		15	0.2	5.02	10.12	
8.4		17	0.2	4.42	8.25	
9.5		25	0.5	4.44	5.95	
7.2		22	1	4.7	3.49	
4		12	0.5	5.71	6.32	
6.7		52	0.5	4.8	3.2	
9		27	0.3	3.69	3.3	
7.8		29	1.5	6.72	5.75	
4.5		41	0.5	3.33	2.27	
6.2		34	0.7	7.56	6.93	
5.6		20	0.5	5.07	6.7	
9		17	0.2	4.39	8.33	
8.4		20	0.1	3.74	3.77	
9.5		19	0.5	3.72	7.37	
9		20	0.5	5.97	11.17	
6.2		16	0.05	4.23	4.18	
7.3	20	0.5	4.39	3.5		
3.6	15	0.7	7	4.82		
6.2	34	0.07	4.84	2.37		
7.3	22	0	4.13	2.7		
4.1	29	0.7	5.78	7.76		
5.4	29	0.2	4.64	2.65		
5	34	0.7	4.21	6.5		
6.2	27	0.3	3.97	2.97		

Sebelum membangun fungsi diskriminan, perlu dilakukan pengujian perbedaan vektor rata-rata kedua kelompok dengan menggunakan T^2 -Hotelling yang

menyatakan bahwa kedua vektor nilai rata-rata yang diuji berbeda nyata pada taraf $\alpha = 0.05$.

Karena kedua vektor rata-rata yang diuji berbeda secara nyata, maka dapat dilakukan uji lanjutan dengan fungsi diskriminan. Lalu dilakukan uji apakah benar kelima variabel dalam fungsi diskriminan semuanya cukup berarti menerangkan perbedaan kandungan minyak mentah pada kedua populasi, ataukah ada variabel tertentu yang tidak memiliki kontribusi yang berarti sehingga dapat dikeluarkan dari model. Akan dilihat pada uji selang kepercayaan disimpulkan nilai rata-rata variabel X_1, X_2 dan X_4 **berbeda secara nyata** serta X_3 dan X_5 **tidak berbeda secara nyata** diantara dua kelompok populasi yang ada pada selang kepercayaan simultan 95%.

Uji korelasi antara setiap variabel dalam model dan fungsi diskriminan itu sendiri dengan formula :

$$R_{Y,X_i} = \frac{d_i}{\sqrt{s_{ii}D^2}}$$

Keterangan :

R_{Y,X_i} = korelasi anatar fungsi diskriminan dan varabel X_i dalam model.

d_i = selisih nilai rata-rata variabel X_i diantara kedua kelompok.

s_{ii} = ragam (varians) dari variabel X_i diperoleh dari matriks S_G .

$$D^2 = (\bar{X}_1 - \bar{X}_2)' S_G^{-1} (\bar{X}_1 - \bar{X}_2)$$

Untuk kasus di atas, nilai $D^2 = 11.845$

Uji korelasi antara variabel, terdapat 3 variabel dengan nilai korelasi terendah, yaitu X_5, X_3 dan X_1 . Dan dikarenakan pada selang kepercayaan simultan 95% variabel X_5 dan X_3 tidak berbeda secara nyata, kemudian variabel X_5 dan X_3 dikeluarkan dari fungsi Diskriminan. Sehingga populasi tersebut sekarang hanya memiliki 3 variabel. Dari tiga variabel X_1, X_2 dan X_4 , diperoleh fungsi diskriminan baru, kemudian dilakukan uji selang kepercayaan dan uji korelasi kembali pada fungsi diskriminan.

Tabel 2 menunjukkan angka akhir dari nilai Wilks' Lambda. Angka chi-square sebesar 47,329 dengan tingkat signifikansi yang tinggi menunjukkan perbedaan yang jelas antara dua kelompok.

while the correlation test returns to the discriminant function.

DATA

Data 3.1 terdiri dari sampel kedua minyak mentah dari 2 grup populasi (%)

Penjelasan

1 = sub Mulinia = 11

2 = upper (Mulinia, second subscales, first subscales) = 38

X1 = Vanadium

X2 = Iron

- X3 = Beryllium
- X4 = Saturate Hydrocarbon
- X5 = Aromatic Hydrocarbon

ANALISIS DATA

Table 1 Rata-rat variabel tiap Group

Variabl e	Group 1 n ₁ = 11	Group 2 n ₂ = 38
X1	4,45	7,33
X2	33,09	22,25
X3	0,17	0,43
X4	5,65	4,66
X5	5,48	5,77

$$S_G = \begin{bmatrix} 3,95 & -0,95 & -0,14 & -0,46 & 1,68 \\ -0,9 & 85,05 & 0,43 & 1,35 & -6,92 \\ -0,14 & 0,43 & 0,09 & 0,18 & -0,01 \\ -0,46 & 1,35 & 0,18 & 1,11 & 0,56 \\ 1,68 & -6,92 & -0,01 & 0,5 & 6,94 \end{bmatrix}$$

Sebelum membangun sebuah fungsi diskriminan, perlu melakukan uji beda vektor rata-rata dari 2 grup populasi dengan stistik Hotelling T .

$$(\bar{X}_1 - \bar{X}_2) = \begin{bmatrix} 4,45 & - & 7,33 \\ 33,09 & - & 22,25 \\ 0,17 & - & 0,43 \\ 6,56 & - & 4,66 \\ 5,48 & - & 5,77 \end{bmatrix} = \begin{bmatrix} -2,88 \\ 10,84 \\ -0,26 \\ 1,9 \\ -0,29 \end{bmatrix}$$

$$S_G^{-1} = \begin{bmatrix} 0,318 & -0,007903 & 0,208 & -0,097 & 0,017 \\ -0,007903 & 0,014 & -0,02 & -0,024 & 0,017 \\ 0,208 & -0,027 & 17,04 & -2,729 & 0,19 \\ 0,157 & -0,024 & -2,729 & 1,534 & 0,19 \\ -0,097 & 0,017 & 0,167 & -0,19 & 0,201 \end{bmatrix}$$

$$(\bar{X}_1 - \bar{X}_2)^t S_G^{-1} (\bar{X}_1 - \bar{X}_2) = 11,845$$

diperoleh

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{X}_1 - \bar{X}_2)^t S_G^{-1} (\bar{X}_1 - \bar{X}_2) = \frac{(11)(38)}{11 + 38_2} (11,845) = 101,045$$

$$\frac{(n_1+n_2-2)p}{n_1+n_2-p-1} F_{0,005;(5)(38+11-5-1)} =$$

$$\frac{(11+38-2)p}{11+38-5-1} (2,435) = 13,3075$$

Karena $T^2 = 101,045 > 13,3075$, H_0 ditolak, jadi kedua vektor rata-rata berbeda secara signifikan pada $\alpha = 0.05$. Pengujian dilanjutkan dengan sebuah fungsi diskrimin

$$Y = a^t X = (\bar{X}_1 - \bar{X}_2)^t S_G^{-1} X =$$

$$\begin{bmatrix} -0,73 \\ 0,126 \\ -10,555 \\ 2,964 \\ 0,00643 \end{bmatrix} X$$

Diperoleh fungsi :

$$\hat{Y} = -073X_1 + 0,126X_2 - 10,555X_3 + 2,964X_4 + 0,00643X_5$$

INTERVAL KEPERCAYAAN

Selanjutnya, 5 variabel didalam fungsi diskriminan diuji untuk menentukan mana saja yang berbeda secara signifikan didalam minyak mentah dan bila ada variabel yang perlu dikeluarkan dari model matematika. Pengujian ini adalah menentukan interval kepercayaan dengan menggunakan formula sebagai berikut,

$$(\bar{X}_1 - \bar{X}_2) \pm c \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) S_{ii,G}} \text{ Where}$$

$$c = \sqrt{\frac{(n_1+n_2-2)p}{n_1+n_2-p-1}} F_{p;n_1+n_2-p-1}$$

Penjelasan:

1. Taksiran 95 % interval kepercayaan simultan untuk variabel X_1 :

$$(\bar{X}_{11} - \bar{X}_{21}) = (4,45 - 7,33) = -2,88$$

$$c =$$

$$\sqrt{\frac{(11+38-2)5}{11+38-5-1}} F_{5;43} = 3,647$$

$$\text{And } \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) S_{ii,G}} =$$

$$\sqrt{\left(\frac{1}{11} + \frac{1}{38}\right) 3,95} = 0,680$$

Terdapat interval kepercayaan $(-2,88 \pm (3.647)(0,680))$ atau $[-5,359, -0,400]$. Karena interval kepercayaan untuk X_1 tidak mengandung 0, maka kita simpulkan bahwa nilai rata-rata dari variabel X_1 berbeda secara signifikan antara dua grup populasi yang ada. Taksiran 95 % interval kepercayaan simultan untuk variabel X_2 adalah

$$(\bar{X}_{12} - \bar{X}_{22}) = (33,09 - 22,25) = 10,84$$

$$c = \sqrt{\frac{(11+38-2)5}{11+38-5-1}} F_{5;43} = 3,647$$

$$\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) S_{ii,G}} = \sqrt{\left(\frac{1}{11} + \frac{1}{38}\right) 0,014} = 0,040$$

Terdapat interval $[10,693, 10,987]$.

Karena interval kepercayaan untu X_2 tidak mengandung 0, maka kita simpulkan rata-rata nilai variabel X_2 dua grup populaasi yang ada. Taksiran 95 % interval kepercayaan simultan untuk X_3 adalah :

$$(\bar{X}_{13} - \bar{X}_{23}) = (0,17 - 0,43) = -0,26$$

$$c = \sqrt{\frac{(11+38-2)5}{11+38-5-1}} F_{5;43} = 3,647$$

$$\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) S_{ii,G}} = \sqrt{\left(\frac{1}{11} + \frac{1}{38}\right) 17,04} = 1,413$$

Terdapat sebuah interval kepercayaan [-5,414, 4,894]. Karena interval kepercayaan untuk X_3 tidak mengandung 0, maka disimpulkan nilai rata-rata X_3 iberbeda secara signifikan antara dua grup populasi.

2. Taksiran 95 % interval kepercayaan simultan

$$(\bar{X}_{14} - \bar{X}_{24}) = (6,56 - 4,66) = 1,9$$

$$c = \sqrt{\left(\frac{(11+38-2)5}{11+38-5-1}\right) F_{5,43}} = 3,647$$

$$\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) S_{ii,G}} = \sqrt{\left(\frac{1}{11} + \frac{1}{38}\right) 1,534} = 0,424$$

Terdapat sebuah interval kepercayaan [0,354, 3,446]. Karena interval kepercayaan untuk X_4 tidak mengandung sebuah nilai 0, maka disimpulkan nilai rata-rata X_4 berbeda secara signifikan antara dua populasi yang ada.

3. Taksiran 95 % interval kepercayaan simultan untuk variabel X_5 adalah,

$$((\bar{X}_{15} - \bar{X}_{25})) = (5,48 - 5,77) = -0,29$$

$$c = \sqrt{\left(\frac{(11+38-2)5}{11+38-5-1}\right) F_{5,43}} = 3,647$$

$$\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) S_{ii,G}} = \sqrt{\left(\frac{1}{11} + \frac{1}{38}\right) 0,201} = 0,151$$

Terdapat interval kepercayaan [-0,843, 0,263].

Karena interval kepercayaan untuk X_5 tidak mengandung sebuah nilai 0, maka disimpulkan bahwa nilai rata-rata variabel X_5 berbeda secara signifikan antara dua grup populasi yang ada.

KORELASI ANTAR VARIABEL AWAL

Formula berikut digunakan untuk menentukan korelasi dari tiap-tiap variabel didalam fungsi diskriminan.

$$R_{Y,X_i} = \frac{d_i}{\sqrt{S_{ii}D^2}}$$

Dimana

R_{Y,X_i} = Korelasi antar fungsi diskriminan dan X_i variable didalam model.

d_i = perbedaan didalam rata-rata variabel X_i antar dua grup. S_{ii} = variansi dari variabel X_i yang diperoleh dari matriks S_G

$$D^2 = (\bar{X}_1 - \bar{X}_2)^t S_G^{-1} (\bar{X}_1 - \bar{X}_2) = 11,845$$

Perhitungan antar variabel X_i adalah sebagai berikut:

$$\begin{aligned} R_{Y,X_1} &= -0,421 ; R_{Y,X_2} = 0,342 ; R_{Y,X_3} \\ &= -0,252 ; R_{Y,X_4} \\ &= 0,524 ; R_{Y,X_5} = -0,032 \end{aligned}$$

Dari korelasi antar variabel, terdapat 3 variabel dengan nilai korelasi terendah yaitu, X_5 , X_3 , dan X_2 . Bagaimanapun didalam 95 % taksiran interval kepercayaan X_5 , and X_3 tidak berbeda secara signifikan, maka keduanya dihilangkan dari fungsi diskriminan. Sehingga populasi memiliki 3 karakteristik matriks baru sebagai berikut,

$$S_G = \begin{bmatrix} 3,95 & -0,95 & -0,46 \\ -0,95 & 85,05 & 1,35 \\ -0,46 & 1,35 & 1,11 \end{bmatrix} \text{ dan } S_G^{-1} = \begin{bmatrix} 0,266 & 0,001246 & 0,109 \\ 0,001246 & 0,012 & -0,014 \\ 0,109 & -0,014 & 0,963 \end{bmatrix}$$

Fungsi diskriminan menjadi $Y = \alpha^t X = (\bar{X}_1 - \bar{X}_2)^t S_G^{-1} X = \begin{bmatrix} -0,546 \\ 0,1 \\ 1,364 \end{bmatrix} X$

Dan formulanya dapat ditulis sebagai berikut.: $\hat{Y} = -0,546X_1 + 0,1X_2 + 1,364X_4$

TAKSIRAN INTERVAL KEPERCAYAAN PADA VARIABEL BARU

Dengan menggunakan metode yan sama, interval kepercayaan untuk X_1 , X_2 , and X_4 masing-masing adalah:

1. Untuk X_1 adalah [-4,899, -0,861], ini artinya bahwa rata-rata nilai X_1 berbeda secara signifikan dalam dua grup populasi. Untuk X_2 adalah [10,721, 10,958], ini artinya bahwa nilai rata-rata X_2 berbeda secara signifikan dalam dua grup populasi
2. Untuk X_4 adalah [0,0641, 3,15], ini berarti nilai rata-rata bahwa X_4 berbeda secara nyata dalam dua grup populasi.

KORELASI ANTAR VARIABEL BARU

Menggunakan rumus yang sama seperti dalam 5 variabel, korelasi antar variabel X_i dan fungsi diskriminan diperoleh sebagai berikut

$$\begin{aligned} D^2 &= 5,246 ; R_{Y,X_1} = -0,632 ; R_{Y,X_2} \\ &= 0,513 ; R_{Y,X_4} = 0,787 \end{aligned}$$

Untuk tujuan pengklasifikasian, nilai m adalah = $\frac{1}{2}D^2 = \frac{1}{2}(5,246) = 2,263$

Sedemikian sehingga objek dengan kriteria berikut $X_1 = 5$; $X_2 = 47$; $X_4 = 706$, akan menentukan grup populasi 1 atau 2 melalui rumus berikut,

$\hat{Y} = -0,546(5) + 0,1(47) + 1,364(7,06) = 11,599$. Karena $Y_0 - m = 11,599 - 2,263$ lebih besar dari 0, jadi objek yang diklasifikasi berada didalam grup populasi 1.

OUTPUT PROGRAM SPSS

Group		Predicted group membership		Total
		1	2	
Original	count 1	10	1	11
	2	2	36	38
	% 1	90.9	9.1	100.0
	2	5.3	94.7	100.0
cross-validated ^a	count 1	10	1	11
	2	2	36	38
	% 1	90.9	9.1	100.0
	2	5.3	94.7	100.0

Classification Results^{b,c}

- a. Cross validation was carried out in the cases analyzed alone and in accordance to the function derived.
- b. Initial cross population group is 93.9% and it suitable to be classified.
- c. Corrected cross population groups 93.9% are valid and suitable to be classified.

4. SIMPULAN

Fungsi diskriminan dapat membantu kita dalam pengujian masalah dimana jika terdapat dua atau lebih populasi yang telah diukur dalam beberapa karakter X_1, X_2, \dots, X_p , maka dapat diuji apakah antara variabel tersebut cukup berarti kehadirannya atau tidak dalam fungsi diskriminan atau dapat dikatakan juga diantara 2 populasi tersebut dimana letak perbedaan sifatnya. Perhitungannya dimulai dengan menghitung pengujian vektor rata-rata dua populasi, jika hasil H_0 diterima maka kita tidak dapat menguji perbedaan sifat antara populasi karena sifat populasinya sama, jika hasil H_0 ditolak maka kita dapat melanjutkan proses analisis untuk mendapatkan fungsi diskriminannya hingga menentukan penggolongan objek.

Ada perbedaan yang signifikan antara populasi kelompok 1 (Sub-Mulinia) dengan populasi kelompok 2 (upper : Mulinia, Subskala pertama, Subskala kedua). Hal ini dibuktikan pada analisis nilai Wilks' Lambda.

Variabel yang membuat perbedaan antara kedua populasi adalah kandungan minyak mentah jenis Saturade Hydrocarbon(x_4), Iron (x_2) dan Vanadium (x_1). Hal ini dibuktikan dengan tingginya nilai korelasi ketiga variabel.

Model atau fungsi Diskriminan untuk kasus tersebut adalah :

$$\hat{Y} = -0.546X_1 + 0.1X_2 + 1.364X_4$$

Model (fungsi) Diskriminan di atas mempunyai ketepatan mengklasifikasikan kasus sebesar 93,9% (Tabel 3). Model di atas bisa digunakan untuk mengklasifikasi sebuah kasus pada tipe kandungan minyak mentah tertentu.

DAFTAR PUSTAKA

A.Johnson and Wichern. (1998), *Applied Multivariate Statistic Analysis Fourth Edition*, Prentice-Hall International, Inc.

R. Y. Liu ; M.Parelius and K Singh (1999), *Multivariate Analysis by data depth*, The Annals of Statistics, Vol 27, pp 783-858

C.Croux ; P.Filzmoser and K.Joossens (2008), *Classification Efficiency for Robust Linear Discriminant Analysis* ; Statistica Sinica, Vol 18 pp 581-599

Friday Zinzendoff Okwonu & Abdul Rahman Othman (2012) ; *A Model classification Technique for Linear Discriminant Analysis for Two Groups* ; UCSI International Journal of Computer Science issues ; Vol 9, issue 3, No 2