

---

## Pengelompokkan Tingkat Kriminalitas di Indonesia Menggunakan Algoritma *Average Linkage*

Azman<sup>1\*</sup>, Anisa<sup>2</sup>, Sri Astuti Thamrin<sup>3</sup>  
<sup>1,2,3</sup>Departemen Statistika, Fakultas MIPA,  
Universitas Hasanuddin, Makassar, 90245, Indonesia

\* Corresponding author, email address: [azman2028@gmail.com](mailto:azman2028@gmail.com)

### Abstract

Crime needs to be analyzed and grouped so that the act does not cause harm either ecologically or psychologically. The statistical method that can be used to classify crime is the Average Linkage Algorithm. The study aims to group and analyze the characteristics of criminal cases in Indonesia. From the results of the analysis, 3 clusters were formed based on the average of each cluster. Cluster 1 consists of Aceh, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, West Java, Central Java, DI Yogyakarta, East Java, Banten, Bali, West Nusa Tenggara, East Nusa Tenggara, West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, North Sulawesi, Central Sulawesi, South Sulawesi, Southeast Sulawesi, Gorontalo, Maluku, North Maluku and Papua. Cluster 2 consists of North Sumatra while Cluster 3 consists of Metro Jaya. The results of the grouping are the basis of the government, apparatus and the community in implementing the handling of criminal acts that occur in the area of each cluster so that prevention can minimize the losses caused by these crimes

**Keywords:** Average Linkage Algorithm, Classify, Crime.

### Abstrak

Tindak kriminal perlu dianalisis dan dikelompokkan agar perbuatan tersebut tidak menyebabkan kerugian secara ekonomis maupun psikologis. Metode statistika yang dapat digunakan untuk mengelompokkan kriminalitas adalah Algoritma *Average Linkage*. Penelitian bertujuan untuk melakukan pengelompokkan dan menganalisis karakteristik kasus kriminalitas di Indonesia. Dari hasil analisis, terbentuk 3 cluster berdasarkan dari rata rata setiap *cluster*. *Cluster* 1 beranggotakan Aceh, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Maluku, Maluku Utara dan Papua. *Cluster* 2 beranggotakan Sumatera Utara sedangkan *Cluster* 3 beranggotakan Metro Jaya. Hasil pengelompokkan tersebut menjadi landasan pemerintah, aparat dan masyarakat dalam melaksanakan penanggulangan tindak kriminal yang terjadi di wilayah setiap *cluster* sehingga pencegahan yang dilakukan dapat meminimalisir kerugian akibat tindak kriminal tersebut.

**Kata Kunci:** Algoritma *Average Linkage*, Klasifikasi, Kriminalitas.

## 1. Pendahuluan

Kriminalitas merupakan segala macam bentuk tindakan dan perbuatan yang merugikan secara ekonomis dan psikologis yang melanggar hukum yang berlaku dalam

negara Indonesia serta norma-norma sosial dan agama. Dapat diartikan bahwa, tindak kriminalitas adalah segala sesuatu perbuatan yang melanggar hukum dan melanggar norma-norma sosial, sehingga masyarakat menentangnya [1]. Pencegahan tindak kriminal dapat dilakukan dengan mengetahui daerah rawan terjadi tindak kriminal dengan cara mengelompokkan dan menganalisis karakteristik setiap daerah. Untuk itu diperlukan suatu metode statistika dalam mengelompokkan dan menganalisis karakteristik setiap daerah yang rawan kriminal. Metode statistika yang dapat digunakan untuk mengkaji kasus kriminal ini adalah analisis *cluster*.

Analisis Cluster merupakan suatu teknik analisis multivariat yang bertujuan untuk mengelompokkan data observasi ataupun variabel-variabel ke dalam cluster sedemikian rupa sehingga masing-masing cluster bersifat homogen sesuai dengan faktor yang digunakan untuk melakukan pengelompokkan. Karena yang diinginkan adalah untuk mendapatkan cluster yang sehomogen mungkin, maka yang digunakan sebagai dasar untuk mengelompokkan adalah kesamaan skor nilai yang dianalisis. Data mengenai ukuran kesamaan tersebut dapat dianalisis dengan analisis cluster sehingga dapat ditentukan objek yang masuk cluster tertentu [2].

Untuk melakukan pengelompokkan setiap observasi dapat digunakan Algoritma *Average Linkage*. Algoritma *Average Linkage* merupakan proses *clustering* yang didasarkan pada jarak rata-rata antar observasinya (*average distance*). Algoritma ini memiliki kelebihan yaitu dalam menentukan jarak antara dua *cluster* didefinisikan sebagai rata-rata jarak antara semua pasangan observasi dengan menggunakan informasi pada semua pasangan jarak, tidak hanya jarak maksimum atau minimum sehingga cenderung lebih stabil [3].

Dalam penelitian ini akan dilakukan pengelompokkan dan menganalisis karakteristik kasus kriminalitas dengan menggunakan Algoritma *Average Linkage*. Penelitian ini bertujuan untuk pengelompokkan dan menganalisis karakteristik kasus kriminalitas tiap Provinsi di Indonesia.

## **2. Material dan Metode**

### **2.1 Sumber Data**

Penelitian ini menggunakan data sekunder jumlah kasus kriminal di Indonesia yang diperoleh dalam buku *Statistika Kriminal 2019* yang diterbitkan oleh Badan Pusat Statistika Pusat. Data yang diambil yaitu jumlah kasus kriminal setiap provinsi di Indonesia pada tahun 2018.

### **2.2 Algoritma Average Linkage**

Algoritma *average linkage* adalah proses *clustering* yang didasarkan pada jarak rata-rata antar observasinya (*average distance*). Pada dasarnya adalah jarak rata-rata antara observasi pengelompokkan dimulai tengah atau pasangan observasi dengan jarak paling mendekati jarak rata-rata. Jarak antara *cluster* pada metode ini ditentukan dan rata-rata

jarak seluruh observasi suatu kelompok lainnya. Metode ini bertujuan meminimumkan rata-rata jarak semua pasangan pengamatan dan dua *cluster* yang digabungkan serta cenderung membentuk cluster dengan ragam kecil. Pada berbagai keadaan, metode ini dianggap lebih stabil dibandingkan dengan kedua metode lainnya yakni Algoritma *Single Linkage* dan *Complete Linkage* [4].

Untuk lebih memahami cara kerja metode ini perhatikan algoritma berikut ini :

1. Menentukan observasi bersesuaian yang memiliki jarak terdekat dalam matriks jarak  $D = \{d_{ik}\}$
2. Menggabungkan observasi yang bersesuaian tersebut, katakanlah observasi U dan observasi V, yang kemudian didapatkan *Cluster* (UV).
3. Menghitung jarak minimum antara *Cluster* (UV) dengan observasi lain katakanlah W yang belum bergabung, dengan rumus :

$$d_{UV(W)} = \frac{\sum_i \sum_k d_{ik}}{N_{(UV)}N_W} \quad (1)$$

dan menghitung kembali matriks jarak baru dengan cara pada langkah ketiga algoritma *Cluster* hierarki *agglomerative* secara umum, dan beri nama D2.

4. Mengulangi langkah 2 sampai bergabung menjadi satu *Cluster*.

Dimana  $d_{ik}$  adalah jarak antara observasi i dalam *Cluster* (UV) dan observasi k dalam *Cluster* W.  $N_{(UV)}$  dan  $N_W$  adalah jumlah observasi dalam *Cluster* (UV) dan W [5].

### 2.3 Uji Kaiser Meyer Olkin (KMO)

Untuk mengetahui apakah sampel yang diambil benar-benar dapat mewakili populasi yang ada dibutuhkan nilai *Kaiser-Meyer Olkin (KMO)*. Dimana nilai *KMO* kurang dari 0.5 menandakan bahwa sampel yang diambil tidak dapat mewakili populasi yang ada [6].

Rumus KMO adalah sebagai berikut:

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2} \quad (2)$$

### 2.4 Uji Multikolinearitas

Menurut Montgomery, Peck, & Vining (1992) dalam Fitriani (2019), kolinearitas terjadi karena terdapat korelasi yang cukup tinggi di antara variabel independen. Untuk mendeteksi ada atau tidaknya multikolinearitas dapat dilakukan dengan menghitung koefisien korelasi sederhana (*korelasi Pearson*) antar variabel bebas, jika terdapat nilai yang mencapai atau melebihi 0,8 maka terjadi multikolinearitas [5].

### 2.5 Analisis Komponen Utama

Analisis Komponen Utama (AKU) adalah suatu teknik analisis statistik untuk mentransformasi variabel-variabel asli yang masih saling berkorelasi satu dengan yang

lain menjadi satu himpunan variabel baru yang tidak berkorelasi lagi. Secara umum tujuan dari AKU adalah mereduksi dimensi data [7].

Adapun langkah-langkah dalam analisis komponen utama berdasarkan matriks korelasi :

1. Menghitung matriks korelasi  $\mathbf{R}$
2. Menentukan nilai eigen  $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_p$ , dengan persamaan :

$$|\mathbf{R} - \lambda \mathbf{I}| = 0 \quad (3)$$

nilai *eigen* selalu diurutkan dari yang terbesar sampai nilai terkecil. Nilai *eigen* menunjukkan besarnya total varian yang dijelaskan oleh KU yang terbentuk. KU dibentuk berdasarkan persamaan :

$$KU_i = \hat{y}_i = \hat{\mathbf{e}}_i \mathbf{X} = \hat{e}_{i1}x_1 + \hat{e}_{i2}x_2 + \dots + \hat{e}_{ip}x_p \quad (4)$$

$$i = 1, 2, \dots, p$$

nilai *eigen* dapat menjelaskan besarnya kontribusi keragaman masing-masing komponen utama dalam menjelaskan keragaman data asal. Besarnya kontribusi keragaman relatif yang mampu dijelaskan oleh komponen utama ke-*i* adalah sebesar:

$$Total\ Varian = \frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \quad (5)$$

Sedangkan besarnya keragaman kumulatif untuk *q* buah komponen utama dapat dituliskan sebagai berikut :

$$Total\ Kumulatif\ Varian = \frac{\sum_{j=1}^q \lambda_j}{\sum_{i=1}^p \lambda_i} \times 100\% \quad (6)$$

3.  $\hat{\mathbf{e}}_i$  adalah vector eigen yang diperoleh dari setiap nilai eigen  $\lambda_p$  yang memenuhi persamaan:

$$(\mathbf{R} - \lambda \mathbf{I}) \hat{\mathbf{e}}_i = 0 \quad (7)$$

Salah satu tujuan dari Analisis Komponen Utama adalah mereduksi dimensi data asal yang semula terdapat *p* variabel bebas menjadi *k* komponen utama (dimana  $k < p$ ). Kriteria pemilihan *k* :

1. Proporsi kumulatif keragaman data asal yang dijelaskan oleh *k* komponen utama minimal 80%.
2. Komponen utama yang dipilih adalah komponen utama yang mempunyai nilai eigen lebih besar satu atau  $\lambda_i > 1$ .

## 2.6 Ukuran Jarak Euclidean

Tujuan analisis *cluster* adalah mengelompokkan observasi yang mirip kedalam *Cluster* yang sama. Oleh karena itu memerlukan beberapa ukuran untuk mengetahui seberapa mirip atau berbeda observasi-observasi tersebut. Pendekatan yang biasa digunakan adalah mengukur kemiripan yang dinyatakan dalam jarak (*distance*) antara pasangan observasi [7].

Jarak *Euclidean* antara *Cluster* ke-*i* dan ke-*j* dari *p* peubah didefinisikan:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2} \quad (8)$$

### 2.7 Interpretasi Cluster

Proses ini dimulai dengan suatu ukuran yang sering digunakan yaitu *Centroid cluster*. Membuat profil dan interpretasi *cluster* tidak hanya untuk memperoleh suatu gambaran saja, melainkan untuk menyediakan rata-rata untuk menilai korespondensi pada *cluster* yang terbentuk serta profil *cluster* memberikan arahan bagi signifikansi praktis. Selain itu juga dilakukan untuk menjelaskan karakteristik dari setiap kelompok berdasarkan dengan tujuan untuk memberi label pada masing-masing kelompok tersebut.

Adapun cara *menghitung mean (centroid)* yaitu: [7].

$$C = \frac{\sum_{j=1}^n \sum_{k=1}^p x_{jk}}{N} \quad (9)$$

## 3. Hasil dan Diskusi

Analisis data pertama kali dianalisis secara deskriptif. Berikut adalah statistik deskriptif data kriminalitas di Indonesia yang disajikan pada Tabel 1.

Tabel 1. Statistik Deskriptif Data Kriminalitas

Variabel	Mean	Maksimum	Minimum	StDev
$x_1$	50,74	186	3	44,12
$x_2$	41,23	161	3	40,86
$x_3$	458,5	2959	5	627,06
$x_4$	337,5	1406	4	361,56
$x_5$	252	1724	16	331,92
$x_6$	876,9	5969	36	1159,59
$x_7$	1137	7214	2	1673,83

Sumber: Data olah, 2020

Berdasarkan Tabel 1 diperoleh nilai minimum, maksimum, mean dan standar variasi dari tujuh variabel. Nilai tersebut menggambarkan perbedaan karakteristik dari setiap variabel. Selanjutnya ditentukan uji Kaiser Meyer Olkin yang dapat dilihat pada Tabel 2.

Tabel 2. Uji KMO

Uji	Nilai
KMO	0,731

Sumber: Data olah, 2020

Berdasarkan Tabel 2 menunjukkan bahwa nilai KMO sebesar  $0,731 > 0,5$  sehingga dapat disimpulkan sampel digunakan sudah cukup untuk dianalisis lebih lanjut. Selanjutnya dilakukan uji multikolinearitas dengan menentukan nilai korelasi person, dengan menggunakan program R Studio 3.5.0. Hasil yang diperoleh bahwa variabel  $x_2$

dengan  $x_3$  mengalami multikolinearitas karena nilai korelasi lebih dari  $\pm 0,80$ . Oleh sebab itu dilakukan analisis komponen utama untuk mengatasi masalah tersebut.

Selanjutnya, analisis komponen utama pada data menghasilkan proporsi kumulatif dan proporsi variansi kumulatif seperti pada Tabel 3.

Tabel 3. Proporsi variansi dan proporsi variansi kumulatif

KU	Nilai <i>Eigen</i>	Total Varian (%)	Total Kumulatif Varian (%)
<b>KU<sub>1</sub></b>	2004,87	84,02	84,02
<b>KU<sub>2</sub></b>	642,34	8,625	92,642
<b>KU<sub>3</sub></b>	523,56	5,73	98,37
<b>KU<sub>4</sub></b>	242,09	1,225	99,597
<b>KU<sub>5</sub></b>	134,67	0,379	99,976
<b>KU<sub>6</sub></b>	27,02	0,015	99,992
<b>KU<sub>7</sub></b>	20,12	0,008	100

Sumber: Data olah, 2020

Berdasarkan Tabel 3, diperoleh bahwa Komponen Utama 1 dan Komponen Utama 2 secara bersama-sama telah dapat menjelaskan keragaman atau variansi dari ketujuh variabel sebesar 92,642 %. Artinya, cukup penggunaan dua komponen utama dalam proses analisis data. Selanjutnya menentukan *eigen vector* atau koefisien dari komponen utama yang digunakan untuk membentuk persamaan KU sebagai berikut:

$$KU_1 = 0.011x_1 + 0.011x_2 + 0.237x_3 + 0.112x_4 + 0.049x_5 + 0.51x_6 + 0.817x_7$$

$$KU_2 = -0.011x_1 + 0.001x_2 - 0.181x_3 + 0.104x_4 - 0.001x_5 + 0.846x_6 - 0.49x_7$$

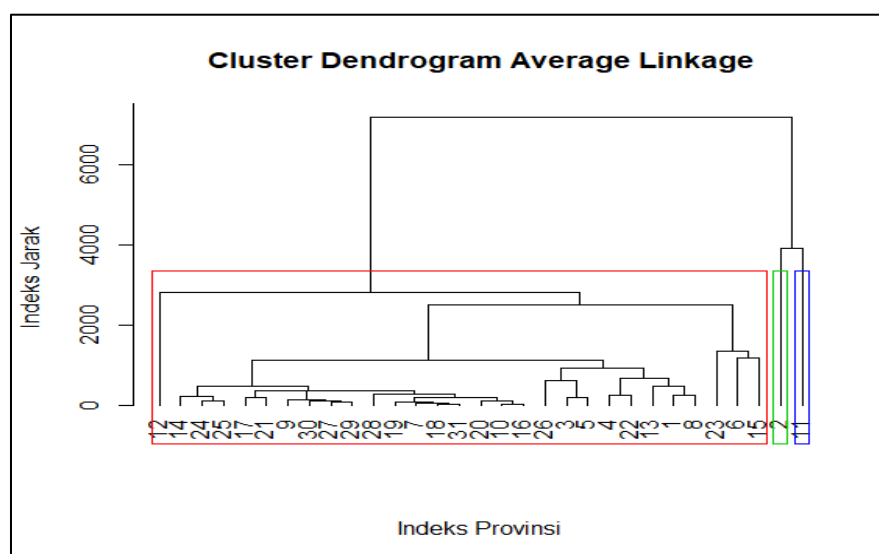
Selanjutnya, menghitung kemiripan antar observasi dengan perhitungan jarak. Perhitungan untuk jarak antara Provinsi Aceh dan Provinsi Sumatera Utara (observasi 1 dan 2 ).

$$d(1,2) = \sqrt{(154,67 - (-345,73))^2 + (5221,4 - (-1699))^2} = 5244,57$$

Hasil perhitungan jarak dari Provinsi Aceh dan Provins Sumatera Utara diperoleh sebesar 5244,57. Cara perhitungan tersebut digunakan untuk menentukan jarak antara pasangan observasi lainnya yang nantinya membentuk matriks jarak antara pasangan observasi.

Hasil perhitungan jarak antara objek melalui algoritma *Average Linkage* menunjukkan bahwa observasi nomor 18 (Provinsi Nusa Tenggara Barat) dan nomor 31 (Provinsi Papua) bergabung menjadi satu *cluster*. Hal ini disebabkan karena kedua provinsi tersebut memiliki jarak paling kecil yaitu 34,51. Hasil tersebut merupakan jarak yang paling dekat dari banyaknya kombinasi jarak 31 Provinsi. Selanjutnya, jarak yang terdekat kedua adalah objek nomor 10 (Provinsi Kepulauan Riau) bergabung dengan nomor 16 (Provinsi Banten) dengan jarak yaitu 51,5. Demikian seterusnya, sehingga

semua observasi bergabung menjadi satu *cluster*. Hasil dari proses tersebut ditampilkan dalam suatu dendrogram seperti yang terlihat pada Gambar 1.



Gambar 1. Dendrogram Analisis *Cluster* Algoritma *Average Linkage*

Pada Gambar 1 menunjukkan bahwa terdapat 3 *cluster* yaitu *cluster* 1 terdiri 29 provinsi, *cluster* 2 terdiri satu provinsi, dan *cluster* 3 terdiri satu provinsi. Selengkapnya ditunjukkan dalam Tabel 4 sebagai berikut:

Tabel 4. Cluster dan Anggotanya

Cluster	Provinsi
Cluster 1	Aceh, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Maluku, Maluku Utara, Papua.
Cluster 2	Sumatera Utara
Cluster 3	Metro Jaya

Sumber: Data olah, 2020

Selanjutnya berdasarkan Tabel 4, dilakukan perhitungan *centroid cluster* dan hasilnya dapat dilihat pada Tabel 5. Berdasarkan nilai rata-rata cluster setiap variable, Cluster 1 memiliki rata-rata terendah dari semua variabel jika dibandingkan dengan 2 cluster lainnya. Adapun Cluster 2 memiliki rata-rata tertinggi untuk variabel kasus pemerkosaan ( $x_1$ ), kasus pembunuhan ( $x_2$ ), kasus penganiayaan ( $x_3$ ), kasus pencurian ( $x_4$ ) dan kasus pengrusakan ( $x_5$ ). Untuk Cluster 3 memiliki rata-rata tertinggi

untuk variabel kasus penipuan ( $x_6$ ) dan kasus narkoba ( $x_7$ ). Hal tersebut berarti bahwa jika dilihat dari nilai rata-rata setiap cluster maka tingkat kriminalitas Cluster 1 berada di tingkat rendah, Cluster 3 berada di tingkat sedang dan Cluster 2 berada di tingkat tinggi.

Tabel 5. *Centroid Cluster*

Variabel	Cluster 1	Cluster 2	Cluster 3
$x_1$	45	186	70
$x_2$	36	161	69
$x_3$	332	2959	1601
$x_4$	295	1153	743
$x_5$	232	804	253
$x_6$	660	2068	5969
$x_7$	763	5907	7214

Sumber: Data olah, 2020

#### 4. Kesimpulan

Berdasarkan hasil analisis cluster yang dilakukan, tingkat kriminalitas setiap provinsi di Indonesia diperoleh 3 cluster berdasarkan nilai rata-rata setiap cluster. Cluster 1 berada pada tingkat rendah untuk tingkat kriminalitas dengan anggota yang paling banyak antara lain Aceh, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, Pulau Jawa, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Pulau Kalimantan dan Sulawesi, Maluku, Maluku Utara dan Papua. Cluster 2 berada pada tingkat tinggi yang beranggotakan Sumatera Utara sedangkan Cluster 3 berada pada tingkat sedang yang beranggotakan Metro Jaya. Hasil pengelompokan tersebut dapat menjadi landasan fokus pemerintah, aparat dan masyarakat dalam melaksanakan penanggulangan dan penertiban tindak kriminal lainnya yang terjadi di wilayah masing-masing setiap cluster sehingga pencegahan yang dilakukan dapat meminimalisir dampak dan kerugian akibat tindak kriminal tersebut.

#### Daftar Pustaka

- [1] K. Kartono. *Patologi Sosial*, Jakarta: Raja grafindo Persada, 1999.
- [2] Gudono. *Analisis Data Multivariat Edisi Pertama*, Yogyakarta: BFE, 2011.
- [3] N. W. A. Aprilia, I. G. A. M. Srinadi dan K. Sari. Pengelompokan Desa/Kelurahan Di Kota Denpasar Menurut Indiktor Pendidikan, *E-Jurnal Matematika*, Vol. 5, No. 2, pp. 38-43, 2016.
- [4] S. Santoso. *Buku Latihan SPSS Statistik Multivariat*, Jakarta: PT Elex Komputindo, 2002.
- [5] R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistical Analysis*. 6th Edition, United States of America: Prentice Hall International.Inc, 2007.



- [6] S. Fitriani. *Pemodelan Modified Jackknife Ridge Regression pada Data yang Mengandung Multikolinearitas*. Skripsi, Makassar: Universitas Hasanuddin, 2019.
- [7] C. E. Mongi. Two Step Clustering Analysis for Combination Data. *JdC*, Vol. 4, No. 1, pp. 9-19, 2015.