

**Optimalisasi K-Means Cluster dengan Principal Component Analysis
pada Pengelompokan Kabupaten/Kota di Pulau Kalimantan Berdasarkan
Indikator Tingkat Pengangguran Terbuka**

**Optimization of K-Means Cluster with Principal Component Analysis
on the Grouping of Districts/Cities on the Island of Kalimantan Based on
Unemployment Rate Indicator**

Muhammad Rais¹, Rito Goejantoro, Surya Prangga

Laboratorium Statistika Komputasi Program Studi Statistika, FMIPA, Universitas Mulawarman

E-mail: rais.muhammad.bahri.11011997@gmail.com,

Abstract

Data mining or often also called knowledge discovery in databases is an activity that includes collecting, using historical data to find regularity, patterns, or relationships in large data sets resulting in useful new information. Cluster analysis is an analysis that aims to group data based on its likeness. This research uses the K-Means method combined with PCA. The K-Means method groups data in the form of one or more clusters that share the same characteristics. While the PCA method was used to reduce research variables. This grouping method was applied to the data indicator of the unemployment rate of districts/cities in Kalimantan Island in 2018. The cluster validation used in this study was the Davies-Bouldin Index (DBI). Based on the results of the analysis, it was concluded that the number of principal components formed was as many as 2 principal components. The most optimal grouping of districts/cities in Kalimantan island in 2018 was to use 2 clusters with a DBI value of 0,507. The grouping of districts/cities in Kalimantan Island in 2018 produced 2 clusters, cluster 1 consisting of 51 districts/cities and clusters of 2 consisting of 5 districts/cities. Cluster 1 was a cluster that has the highest percentage of the poor population and the highest labor force participation rate when compared to cluster 2. While cluster 2 was a cluster that has an index value of human development, population, number of the labor force, number of unemployed, population density, and the minimum wage of district/city was high compared to cluster 1.

Keywords: DBI, Unemployment Rate Indicator, K-Means, PCA

Pendahuluan

Masalah pengangguran menjadi salah satu pokok permasalahan dalam pembangunan ekonomi suatu daerah. Tingginya tingkat pengangguran merupakan salah satu cerminan kurang berhasilnya pembangunan dalam suatu wilayah karena terjadi ketidakseimbangan antara jumlah angkatan kerja dengan jumlah lapangan pekerjaan yang tersedia. Tingkat pengangguran terbuka adalah angka yang menunjukkan banyaknya pengangguran terhadap 100 penduduk yang masuk dalam kategori angkatan kerja (Sukirno, 2004). Rata-rata tingkat pengangguran terbuka di Pulau Kalimantan tahun 2018 adalah 4,92%, turun 0,24% dari tahun 2017. Tingkat pengangguran terbuka tertinggi di Pulau Kalimantan pada tahun 2018 adalah Provinsi Kalimantan Timur sebesar 6,60%, hal ini tergolong cukup tinggi jika dilihat dari rata-rata tingkat pengangguran terbuka Indonesia sebesar 5,34% (BPS, 2019).

Analisis multivariat mencakup analisis data penelitian yang menggunakan banyak variabel yang dikenakan pengukuran secara bersamaan.

Analisis Multivariat merupakan teknik analisis yang digunakan untuk menganalisis seperangkat data dengan banyak variabel sebagai objek yang akan diukur (Johnson dan Wichern,

2007). Analisis multivariat dikelompokkan menjadi dua kelompok besar, yaitu metode dependensi dan metode interdependensi (Widarjono, 2015).

Analisis interdependensi (*interdependence methods*) merupakan analisis saling ketergantungan yang sering digunakan untuk memberikan arti kepada suatu set variabel atau mengelompokkan suatu set variabel menjadi kelompok yang lebih sedikit jumlahnya dan masing-masing kelompok membentuk variabel baru yang disebut faktor (Widarjono, 2015).

Salah satu teknik interdependensi adalah *principal component analysis (PCA)*. PCA merupakan suatu teknik mereduksi data multivariat (banyak data) untuk mengubah suatu matriks data awal atau data asli menjadi suatu set kombinasi linier yang lebih sedikit akan tetapi menyerap sebagian besar jumlah varian dari data awal (Supranto, 2004). PCA bertujuan untuk menyederhanakan variabel-variabel yang diamati dengan mereduksi dimensinya tanpa kehilangan banyak informasi dari variabel asalnya.

Data mining merupakan suatu metode pengolahan data untuk menemukan pola atau kecenderungan yang penting yang tersembunyi dari suatu basis data yang tidak diketahui keberadaannya (Santosa, 2007). Analisis *cluster*

merupakan analisis yang bertujuan untuk mengelompokkan data (objek) yang didasarkan hanya pada informasi yang ditemukan pada data yang menggambarkan suatu objek dan hubungan di antaranya. Tujuan analisis *cluster* adalah agar objek-objek yang bergabung dalam sebuah *cluster* merupakan objek-objek yang mirip satu sama lain dan berbeda dengan objek yang berada dalam *cluster* yang lain. Salah satu metode yang biasa digunakan dalam analisis *cluster* adalah metode *K-Means cluster*. Metode *K-Means* digunakan sebagai alternatif untuk data dengan ukuran atau jumlah yang besar karena memiliki kecepatan yang lebih tinggi dibandingkan dengan metode hierarki (Prasetyo, 2012).

Berdasarkan uraian di Atas, *PCA* dapat mereduksi dimensi suatu data tanpa mengurangi karakteristik data tersebut secara signifikan dengan resiko kehilangan informasi yang sangat kecil. Penerapan *PCA* dengan mereduksi dimensi data dapat mengoptimalkan hasil dari analisis *cluster*. Sedangkan *K-Means cluster* dapat mengelompokkan data berdasarkan partisi serta kemiripan objek yang ada di dalam *cluster*. Oleh karena itu, peneliti tertarik melakukan penelitian dengan judul “Optimalisasi *K-Means Cluster* dengan *Principal Component Analysis* pada Pengelompokan Kabupaten/Kota di Pulau Kalimantan Berdasarkan Indikator Tingkat Pengangguran Terbuka”.

Analisis Multivariat

Analisis multivariat sebagai analisis data penelitian yang menggunakan banyak variabel yang dikenakan pengukuran yang secara bersamaan menganalisis beberapa pengukuran terhadap individual atau objek dalam suatu riset data. Analisis Multivariat merupakan teknik analisis yang digunakan untuk menganalisis seperangkat data yang menggunakan banyak variabel sebagai objek yang akan diukur secara simultan (Widarjono, 2015).

Principal Component Analysis

Principal component analysis atau analisis komponen utama adalah teknik yang digunakan untuk menyederhanakan suatu data, dengan cara mentransformasi data secara linier sehingga terbentuk sistem koordinat baru dengan varians maksimum (widarjono, 2015). Tujuan dari *PCA* adalah dapat mereduksi dimensi suatu data menjadi dimensi yang lebih rendah dengan resiko kehilangan informasi yang kecil (Rajashree, 2010). Pereduksian variabel dilakukan dengan melakukan pemilihan *principal component* yang akan menggantikan set variabel asal. *Principal component* yang dipilih adalah *principal component* yang memiliki ragam (variens) yang paling besar, artinya *principal component* tersebut memuat informasi paling besar mewakili variabel

asalnya (Pramana, dkk. 2018).

Data Mining

Data mining adalah suatu metode pengolahan data untuk menemukan pola yang tersembunyi dari suatu basis data. Data mining sering juga disebut *knowledge discovery in database (KDD)* atau *pattern recognition*. *KDD* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar sehingga menghasilkan informasi baru yang berguna. Sedangkan istilah *pattern recognition* atau disebut pengenalan pola mempunyai tujuan pengetahuan yang akan digali dari dalam bongkahan data yang sedang dihadapi (Santosa, 2007).

Analisis Cluster

Analisis *cluster* merupakan analisis yang bertujuan untuk mengelompokkan objek data yang didasarkan hanya pada informasi yang ditemukan pada data yang menggambarkan objek tersebut dan hubungan di antaranya. Tujuannya adalah agar objek-objek yang bergabung dalam sebuah *cluster* merupakan objek-objek yang mirip satu sama lain dan berbeda dengan objek yang berada dalam *cluster* yang lain. Semakin besar kemiripannya (homogenitas) dalam kelompok maka objek tersebut akan semakin dekat jaraknya dan semakin besar perbedaannya di antara kelompok lainnya (Prasetyo, 2012).

K-Means Cluster

K-Means cluster merupakan salah satu metode data non hierarki *clustering* atau biasa disebut dengan metode partisi yang mengelompokkan data dalam bentuk satu atau lebih *cluster*. Data-data yang memiliki karakteristik yang sama akan dikelompokkan dalam satu *cluster* dan data yang memiliki karakteristik yang berbeda akan dikelompokkan dengan *cluster* yang lain sehingga data yang berada dalam satu *cluster* memiliki tingkat variasi yang kecil (Agusta, 2007). Jika jarak dua objek data atau titik data cukup dekat, maka dua objek itu mirip. Semakin dekat berarti semakin tinggi kemiripannya. Sebaliknya semakin tinggi nilai jarak, maka akan semakin tinggi pula ketidakmiripannya (Santosa, 2007).

Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah:

- X_1 : Indeks Pembangunan Manusia (%)
- X_2 : Jumlah Penduduk (Orang)
- X_3 : Jumlah Angkatan Kerja (Orang)
- X_4 : Persentase Penduduk Miskin (%)
- X_5 : Tingkat Partisipasi Angkatan Kerja (%)
- X_6 : Jumlah Pengangguran (Orang)

X_7 : Kepadatan Penduduk (Jiwa/ Km^2)
 X_8 : Upah Minimum Kabupaten/Kota (Rupiah)

Metode Penelitian

Adapun langkah-langkah yang dilakukan dalam analisis data adalah sebagai berikut:

1. Melakukan analisis statistika deskriptif untuk melihat gambaran umum data penelitian.
2. Melakukan reduksi data penelitian menggunakan PCA dengan tahapan sebagai berikut:

Adapun tahapan-tahapan dalam PCA adalah sebagai berikut:

- a. Penormalisasian data dengan persamaan:

$$Z_{ap} = \frac{x_{ap} - \bar{x}_p}{S_p} \quad (1)$$

dengan:

Z_{ap} : hasil normalisasi data

x_{ap} : data ke- a pada variabel ke- p

\bar{x}_p : rata-rata data variabel ke- p

S_p : deviasi standar variabel ke- p

- b. Menghitung koefisien korelasi dengan persamaan:

$$r_{z_p, z_q} = \frac{n(\sum_{a=1}^n z_{ap} z_{aq}) - (\sum_{a=1}^n z_{ap}) \cdot (\sum_{a=1}^n z_{aq})}{\sqrt{n(\sum_{a=1}^n z_{ap}^2) - (\sum_{a=1}^n z_{ap})^2} \cdot \sqrt{n(\sum_{a=1}^n z_{aq}^2) - (\sum_{a=1}^n z_{aq})^2}} \quad (2)$$

- c. Menghitung matriks kovarian berdasarkan matriks korelasi
- d. Menghitung nilai eigen menggunakan persamaan:

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

dan vektor eigen pada persamaan:

$$(R\vec{v} = \lambda\vec{v}) \quad (4)$$

- e. Menentukan jumlah komponen utama yang mungkin terbentuk dengan melihat kriteria nilai eigen yang lebih besar atau sama dengan 1.
- f. Membentuk komponen matriks korelasi yang menunjukkan besarnya korelasi variabel terhadap skor komponen yang terbentuk dengan persamaan:

$$r_{x_p, PC_t} = \vec{v}_{at} \sqrt{\lambda_t} \quad (5)$$

- g. Menghitung transformasi data set baru hasil reduksi dengan PCA dengan persamaan:

$$PC_{at} = \vec{v}_{1a} Z_1 + \vec{v}_{2a} Z_2 + \dots + \vec{v}_{ap} Z_p \quad (6)$$

3. Melakukan pengelompokan data dengan metode *K-Means* dengan tahapan sebagai berikut:

- a. Menentukan banyaknya jumlah *cluster* (k) yang akan digunakan.
- b. Memilih pusat *cluster* awal secara *trial* dan *error* dari data penelitian.
- c. Menghitung jarak *Euclid* data dengan pusat *cluster* dengan persamaan:

$$d(x_{a, c_i}) = \sqrt{\sum_{p=1}^m (x_{ap} - c_{ip})^2} \quad (7)$$

$a = 1, 2, 3, \dots, n$ dan $i = 1, 2, 3, \dots, k$

- d. Mengalokasikan data penelitian ke pusat *cluster* terdekat berdasarkan jarak *Euclid* yang terkecil.
- e. Menghitung pusat *cluster* baru berdasarkan keanggotaan dengan persamaan:

$$\bar{c}_{ip} = \frac{1}{n_i} \sum_{a=1}^{n_i} x_{ap} \quad (8)$$

$i = 1, 2, 3, \dots, k$ dan $p = 1, 2, 3, \dots, n$
dengan:

C_i^a : *centroid* pada *cluster* ke- i iterasi ke- a

\bar{c}_{ip} : rata-rata *cluster* ke- i variabel ke- p

n_i : jumlah objek anggota *cluster* ke- i

x_{ap} : objek pengamatan ke- a variabel ke- p

- f. Kembali ke langkah c,d dan e apabila masih ada data yang berpindah *cluster*.
- g. Mengulangi langkah c, d, e dan f untuk banyak *cluster* (k) yang berbeda.
4. Melakukan uji validitas *cluster* menggunakan nilai *Davies-Bouldin Index* dengan persamaan:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (9)$$

dengan:

DBI : nilai *Davies Bouldin Index*

$R_{i,j}$: rasio seberapa baik nilai perbandingan antara *cluster* ke- i dan *cluster* ke- j

k : jumlah *cluster* yang digunakan

5. Menginterpretasikan hasil *cluster* terbaik berdasarkan nilai *DBI* terkecil.

Hasil dan Pembahasan

1. Statistika Deskriptif

Berdasarkan Tabel 1 dapat dilihat bahwa Indeks pembangunan manusia memiliki nilai minimum 61,82 %, nilai maksimum 80,34 % dan rata-rata 69,85 %. Jumlah penduduk memiliki nilai minimum 28.833 orang, nilai maksimum

778.883 orang dan rata-rata 252.754 orang. Jumlah angkatan kerja memiliki nilai minimum 12.273 orang, nilai maksimum 393.174 orang dan rata-rata 139.612 orang. Persentase penduduk miskin memiliki nilai minimum 2,64 %, nilai maksimum 12,83% dan rata-rata 6,28 %. Tingkat partisipasi angkatan kerja memiliki nilai minimum 61,07 %, nilai maksimum 79,34 % dan rata-rata 69,75 %. Jumlah pengangguran memiliki nilai minimum 616 orang, nilai maksimum 30.272 orang dan rata-rata 6.898 orang. Kepadatan penduduk memiliki nilai minimum 1,35 jiwa/km², nilai maksimum 9.734,29 jiwa/km² dan rata-rata 420,93 jiwa/km². Upah minimum kabupaten/kota memiliki nilai minimum Rp. 2.052.624, nilai maksimum Rp. 3.583.880 dan rata-rata Rp. 2.583.150.

Tabel 1. Statistika deskriptif data indikator tingkat pengangguran terbuka di Kabupaten/Kota di Pulau Kalimantan tahun 2018

Variabel	Minimum	Maksimum	Rata-rata
X ₁	61,82	80,34	69,85
X ₂	28.833	778.883	252.754
X ₃	12.273	393.174	139.612
X ₄	2,64	12,83	6,28
X ₅	61,07	79,34	69,75
X ₆	616	30.272	6.898
X ₇	1,35	9.734,29	420,93
X ₈	2.052.624	3.583.880	2.583.15

2. Principal Component Analysis

Adapun tahapan-tahapan PCA adalah:

a. Normalisasi Data

Normalisasi dilakukan agar data memiliki rentang nilai yang tidak jauh berbeda, sehingga data yang lebih besar tidak akan mendominasi data yang lebih kecil, sehingga agar hasil perhitungan yang diperoleh semakin akurat.

Tabel 2. Normalisasi data indikator tingkat pengangguran terbuka di Kabupaten/Kota di Pulau Kalimantan tahun 2018

Data	X ₁	X ₂	...	X ₈
1	0,398	0,038	...	-0,009
2	0,190	-0,513	...	0,760
3	0,747	2,408	...	0,470
...
56	1,321	0,053	...	2,259

b. Koefisien Korelasi

Berdasarkan Tabel 3, dapat dilihat bahwa terdapat hubungan atau korelasi yang kuat antara variabel X₁ dengan X₆ karena nilai

korelasinya di atas 0,5. Begitu juga X₂ dengan X₃ dan X₆ dan seterusnya.

Tabel 3. Koefisien korelasi data indikator tingkat pengangguran terbuka di Kabupaten/Kota di Pulau Kalimantan tahun 2018

	X ₁	X ₂	...	X ₇	X ₈
X ₁	1,000	0,323	...	0,441	0,423
X ₂	0,323	1,000	...	0,268	-0,189
...
X ₇	0,441	0,268	...	1,000	0,049
X ₈	0,423	-0,189	...	0,049	1,000

c. Matriks varians kovarians

Matriks varians kovarians dibentuk berdasarkan nilai koefisien korelasi.

$$R = \begin{pmatrix} 1 & 0,323 & \dots & 0,441 & 0,423 \\ 0,323 & 1 & \dots & 0,269 & -0,189 \\ \vdots & \vdots & \ddots & \dots & \vdots \\ 0,441 & 0,268 & \dots & 1 & 0,049 \\ 0,423 & -0,189 & \dots & 0,049 & 1 \end{pmatrix}$$

d. Nilai Eigen dan Vektor Eigen

Nilai eigen (λ)

$$\lambda = \begin{pmatrix} 3,562 \\ 1,683 \\ 0,832 \\ 0,756 \\ 0,641 \\ 0,342 \\ 0,117 \\ 0,067 \end{pmatrix}$$

Berdasarkan nilai eigen di atas, dapat dilihat ada 8 nilai eigen. Dalam penentuan banyaknya *principal component* yang akan dipilih menggunakan kriteria nilai eigen ≥ 1 , maka diperoleh nilai eigen yang memenuhi kriteria tersebut yaitu 3,562 dan 6,831.

Vektor eigen (\vec{v})

$$\vec{v} = \begin{pmatrix} -0,377 & 0,398 & \dots & 0,287 & 0,097 \\ -0,414 & -0,393 & \dots & -0,519 & 0,549 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -0,345 & 0,088 & \dots & 0,066 & 0,335 \\ -0,053 & 0,572 & \dots & -0,183 & 0,014 \end{pmatrix}$$

e. Komponen korelasi

Komponen korelasi menunjukkan besarnya korelasi variabel terhadap skor komponen yang terbentuk. Berdasarkan Tabel 4, dapat disimpulkan bahwa:

Korelasi variabel X₁ terhadap PC1 sebesar -0,712 dan terhadap PC2 sebesar 0,516. Karena korelasi variabel X₁ terhadap PC1 lebih besar maka variabel X₁ bergabung dengan PC1. Korelasi variabel X₂ terhadap PC1 sebesar -0,781 dan terhadap PC2 sebesar

-0,510. Karena korelasi variabel X2 terhadap PC1 lebih besar maka variabel X2 bergabung dengan PC1 dan seterusnya sampai dengan variabel X8.

Tabel 4. Komponen matriks korelasi

	PC_1	PC_2
X_1	-0,712	0,516
X_2	-0,781	-0,510
X_3	-0,745	-0,471
X_4	0,343	-0,531
X_5	0,653	-0,269
X_6	-0,953	-0,132
X_7	-0,651	0,114
X_8	-0,100	0,742

f. Persamaan *Principal Component*

Persamaan *principal component* terbentuk berdasarkan vektor eigen, maka diperoleh persamaan sebagai berikut:

$$PC_1 = -0,377Z_1 - 0,414Z_2 - 0,395Z_3 + 0,182Z_4 + 0,346Z_5 - 0,505Z_6 - 0,345Z_7 - 0,053Z_8$$

$$PC_2 = 0,398Z_1 - 0,393Z_2 - 0,363Z_3 + 0,409Z_4 - 0,207Z_5 - 0,102Z_6 + 0,088Z_7 + 0,572Z_8$$

g. Transformasi Data Set

Transformasi data menggunakan persamaan *principal component* dengan memasukkan nilai data yang telah dinormalisasi sehingga diperoleh data baru sebagai berikut:

Tabel 5. Transformasi data set

No	PC_1	PC_2
1	-0,013	-0,233
2	1,067	0,435
3	-3,561	-1,510
4	-0,818	-0,671
5	-0,276	1,478
...
56	-1,198	2,252

3. *K-Means Cluster*

Tabel 6. Nilai *DBI* untuk keseluruhan *cluster*

Banyak <i>Cluster</i> (k)	Nilai <i>DBI</i>
2	0,507
3	0,895
4	0,817
5	0,756
6	0,792
7	0,745
8	0,791

Berdasarkan nilai *DBI* dari banyak *cluster* 2 sampai dengan 8 *cluster* ($k=2,3,4,\dots,8$), nilai *DBI* terkecil adalah 0,507 pada banyak *cluster* 2 ($k=2$), sehingga *cluster* yang optimal yang digunakan untuk pengelompokan adalah sebanyak 2 *cluster*.

Adapun tahapan-tahapan dalam *K-Means Cluster* adalah:

- Menentukan banyaknya *cluster* berdasarkan nilai *DBI*, yaitu 2 *cluster*.
- Menentukan pusat *cluster* awal

Pusat *cluster* awal ditentukan secara acak dari data penelitian. Pusat *cluster* awal ditentukan berdasarkan *trial and error*. Sehingga dipilih data ke-5 dan data ke-36 sebagai berikut:

Tabel 7. Pusat *cluster* awal

<i>Cluster</i> (c_{ij})	Variabel	
	PC_1	PC_2
1	-0,276	1,478
2	-5,883	0,751

- Menghitung jarak Euclid data pengamatan terhadap pusat *cluster* awal

Tabel 8. Jarak Euclid terhadap pusat *cluster* awal

Data	Jarak Euclid	
	<i>Cluster</i> 1	<i>Cluster</i> 2
1	1,731	5,952
2	1,701	6,957
3	4,441	3,242
...
56	1,204	4,919

- Mengalokasikan data penelitian ke pusat *cluster* terdekat iterasi ke-1

Mengalokasikan data pengamatan ke pusat *cluster* terdekat iterasi ke-1 menggunakan persamaan 8. Maka diperoleh bahwa *cluster* 1 beranggotakan 51 kabupaten/kota. Sedangkan *cluster* 2 beranggotakan 5 kabupaten/kota.

- Menghitung pusat *cluster* baru berdasarkan keanggotaan *cluster*.

Tabel 9. Pusat *cluster* baru iterasi ke-1

<i>Cluster</i> (c_{ij})	Variabel	
	PC_1	PC_2
1	0,492	0,031
2	-5,020	-0,321

- Menghitung jarak Euclid data pengamatan terhadap pusat *cluster* baru

Tabel 10. Jarak Euclid terhadap pusat *cluster* baru

Data	Jarak Euclid	
	Cluster 1	Cluster 2
1	0,570	5,008
2	0,702	6,133
3	4,336	1,883
...
56	2,791	4,607

- g. Mengalokasikan data penelitian ke pusat *cluster* terdekat iterasi ke-2
Mengalokasikan data pengamatan ke pusat *cluster* terdekat iterasi ke-2 menggunakan persamaan 8. Maka diperoleh bahwa *cluster* 1 beranggotakan 51 kabupaten/kota, sedangkan *cluster* 2 beranggotakan 5 kabupaten/kota.
- h. Menghitung pusat *cluster* baru
Pusat *cluster* baru ditentukan berdasarkan keanggotaan *cluster*. Berdasarkan hasil alokasi data pada iterasi ke-2 bernilai sama seperti iterasi pertama, di mana tidak terdapat perubahan keanggotaan *cluster* atau tidak ada lagi data pengamatan yang berpindah *cluster*.
Hasil pengelompokan kabupaten/kota menggunakan metode *PCA K-Means* berhenti pada iterasi ke-2, di mana tidak ada lagi data pengamatan yang berpindah *cluster*.

Tabel 11. Hasil pengelompokan *PCA K-Means* untuk $k=2$

Clus ter	Anggota Cluster		
	Kode	Kabupaten/Kota	
1	1, 2, 4, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55 dan 56	Paser, Kutai Barat, Kutai Timur, Berau, Penajam Paser Utara, Mahakan Ulu, Bontang, Sambas, Bengkayang, Landak, Mempawah, Sanggau, Ketapang, Sintang, Kapuas Hulu, Sekadau, Melawi, Kayong Utara, Kubu Raya, Singkawang, Tanah Laut, Kotabaru, Banjar, Barito Kuala, Tapin, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Tabalong, Tanah Bumbu, Balangan, Banjar Baru, Kotawaringin Barat, Kotawaringin Timur, Kapuas, Barito Selatan, Barito Utara, Sukamaru, Lamandau, Seruyan, Katingan, Pulang Pisau, Gunung Mas, Berito Timur, Murung Raya, Palangka Raya, Malinau, Bulungan, Tanah Tidung, Nunukan dan Tarakan	
		2	Kutai Kartanegara, Balikpapan, Samarinda, Pontianak dan Banjarmasin

4. Interpretasi Hasil Pengelompokan *PCA K-Means*

Tabel 12. Rata-rata nilai variabel terhadap *cluster* yang diikutinya

Variabel	Cluster ke-i	
	1	2
X ₁	69,085	77,656
X ₂	217.122,726	616.192
X ₃	109.157,158	279.633
X ₄	6,423	4,764
X ₅	70,378	63,296
X ₆	4.964,059	26.627
X ₇	108,495	3.607,800
X ₈	2.580.309	2.612.131

Berdasarkan Tabel 12 dapat diketahui bahwa: *Cluster* 1 merupakan *cluster* di mana variabel persentase penduduk miskin dan variabel tingkat partisipasi angkatan kerja memiliki angka yang lebih tinggi jika dibandingkan dengan *cluster* 2. Sedangkan *cluster* 2 merupakan *cluster* di mana variabel indeks pembangunan manusia, jumlah penduduk, jumlah angkatan kerja, jumlah pengangguran, kepadatan penduduk dan upah minimum kabupate/kota yang lebih tinggi dibandingkan dengan *cluster* 1.

Kesimpulan

- Jumlah komponen utama (*principal component*) yang terbentuk dari hasil reduksi variabel menggunakan metode *PCA* dengan kriteria nilai eigen yang lebih besar atau sama dengan satu adalah sebanyak dua komponen.
- Pengelompokan kabupaten/kota di pulau Kalimantan tahun 2018 yang paling optimal adalah pengelompokan yang menggunakan 2 *cluster* ($k=2$) dengan nilai *DBI* sebesar 0,507.
- Hasil pengelompokan kabupaten/kota di Pulau Kalimantan tahun 2018 berdasarkan indikator tingkat pengangguran terbuka dihasilkan sebanyak 2 kelompok (*cluster*). *Cluster* 1 beranggotakan 51 kabupaten/kota, sedangkan *cluster* 2 beranggotakan 5 kabupaten/kota.

Daftar Pustaka

Agusta, Y. 2007. K-Means - Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem dan Informatika*, 47-60.

Badan Pusat Statistik Kalimantan Timur. 2019. *Kalimantan Timur Dalam Angka 2019*. Samarinda: Badan Pusat Statistika Provinsi Kalimantan Timur.

Johnson, R.A. and Wichern, D.W. 2007. *Applied Multivariate Statistical Analysis*. New Jersey: Pearson Prentice Hall.

Pramana, S., Yuniarto, B., Mariyah, S., Santoso, I., Nooraeni, R. 2018. *Data Mining dengan*

- R: *Konsep serta Implementasi*. Bogor: IN MEDIA.
- Prasetyo, E. 2012. *Data Mining: Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi.
- Rajashree, D. M. 2010. Hybridized K-Means Clustering Approach For High Dimensional Data. *International Journal of Engineering, Science and Technology*, 59-66.
- Santosa, B. 2007. *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Sukirno, S. 2004. *Pengantar Teori Makro Ekonomi*. Jakarta: Raja Grafindo Persada.
- Supranto, J. 2004. *Analisis Multivariat: Arti dan Interpretasi*. Jakarta: Rineka Cipta.
- Widarjono, A. 2015. *Analisis Multivariat Terapan*. Yogyakarta: UPP STIM YKPN.

