

**Pengelompokan Kabupaten/Kota Di Pulau Kalimantan Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2020 Menggunakan Optimasi K-Means Cluster Dengan Principle Component Analysis (PCA)**

**Grouping Of Regencies/Cities in Kalimantan Island Based On Human Development Index Indicators In 2020 Using K-Means Cluster Optimization Using Principle Component Analysis (PCA)**

**Khoiril Anwar<sup>1</sup>, Rito Goejantoro<sup>2</sup>, Surya Prangga<sup>3</sup>**

<sup>1</sup>Laboratorium Statistika Komputasi Program Studi Statistka, FMIPA, Universitas Mulawarman

<sup>2</sup>Program Studi Statistika, Jurusan Matematika, FMIPA, Universitas Mulawarman

E-mail: [lanwarkhoiril627@gmail.com](mailto:lanwarkhoiril627@gmail.com), [rigoejantoro@fmipa.unmul.ac.id](mailto:rigoejantoro@fmipa.unmul.ac.id), [suryapranngge@gmail.com](mailto:suryapranngge@gmail.com)

**ABSTRACT**

*Data mining is a technique or process to obtain useful information from a large database. Based on its functionality, one of the tasks of data mining is to group data. Cluster analysis is an analysis that aims to group objects based on the information found in the data. One of the cluster analysis methods is the K-Means cluster method, which is a non-hierarchical grouping method by dividing the data set into a number of groups that do not overlap between one group and another. This study aims to classify districts/cities on the island of Kalimantan based on indicators of the human development index and obtain the silhouette coefficient value from the optimal cluster analysis using the K-Means algorithm on principle component analysis. The data used is the 2020 human development index data in districts / cities on the island of Kalimantan and used 8 variables from the human development index indicator. The results of the optimal cluster formed in the grouping of regencies/cities on the island of Kalimantan using the K-Means cluster method on the principle component analysis are 4 clusters. Cluster 1 has 20 regencies/cities, cluster 2 has 3 regencies/cities, cluster 3 has 26 regencies/cities and cluster 4 has 7 regencies/cities. The silhouette coefficient value for data validation from district/city clustering on the island of Kalimantan using the K-Means cluster method on principle component analysis produces 4 clusters of 0.540 which states that the cluster structure formed in this grouping is a medium structure.*

**Keywords:** Human Development Index Indicator, K-Means Cluster, Principle Component Analysis, Silhouette Coefficient,

**Pendahuluan**

Analisis Multivariat berasal dari kata *multy* dan *variate* yang berarti alat analisis lebih dari dua variabel. Analisis multivariat sering digunakan untuk memecahkan permasalahan dalam penelitian yang sifatnya sangat kompleks.

Secara umum analisis multivariat dibagi menjadi dua metode yaitu metode dependensi dan metode interdependensi. Metode dependensi merupakan analisis ketergantungan yang digunakan untuk menjelaskan suatu nilai variabel terikat berdasarkan lebih dari satu variabel bebas yang mempengaruhinya. Sedangkan untuk metode interdependensi merupakan analisis yang digunakan untuk menjelaskan seperangkat variabel dalam pengelompokan berdasarkan variabel-variabel tertentu.

Analisis *cluster* merupakan analisis yang digunakan dalam pengelompokan beberapa objek menjadi beberapa kelompok di mana setiap kelompok yang terbentuk terdiri dari objek yang memiliki kemiripan di dalam kelompok (Supranto, 2010). Menurut Hair dkk (2014), metode dalam analisis *cluster* dibagi menjadi dua yaitu, metode hierarki dan non-hierarki.

*K-Means clustering* adalah salah satu metode *cluster analysis* yang berusaha untuk mempartisi objek yang ada ke dalam satu atau lebih *cluster* atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama dikelompokkan dalam satu *cluster* yang sama begitupun sebaliknya dengan objek yang mempunyai karakteristik yang berbeda akan dikelompokkan ke dalam *cluster* yang berbeda (Ediyanto, 2013). Selain memiliki kelebihan *K-Means* juga memiliki kelemahan yang diakibatkan oleh penentuan pusat awal *cluster*. Hasil *cluster* yang terbentuk dari metode *K-Means* tergantung pada inisiasi nilai pusat awal *cluster* yang diberikan. Sehingga hal ini menyebabkan hasil *cluster* berupa solusi di mana memiliki sifat *local optimal*. Selain itu algoritma *K-Means* dihadapkan pada persoalan *curse of dimensionality* saat dihadapkan pada persoalan data dengan fitur besar. Algoritma klasifikasi dan *clustering* menjadi bermasalah pada data dengan dimensi tinggi yaitu dengan menurunnya akurasi klasifikasi dan kualitas *cluster* yang rendah dan juga berpengaruh pada waktu komputasi yang lama. Akan tetapi permasalahan ini dapat diatasi

dengan melakukan reduksi dimensi (Wahono, 2015).

Reduksi dimensi digunakan untuk menghilangkan fitur yang tidak relevan, mengurangi *noise*, dan mengurangi *curse of dimensionality*. Banyak metode yang bisa digunakan dalam reduksi dimensi. Salah satunya yaitu metode *Principal Component Analysis (PCA)*. *PCA* merupakan teknik yang digunakan untuk menyederhanakan suatu data dengan cara mentransformasi linier sehingga terbentuk suatu sistem koordinat baru dengan varians minimum. *PCA* digunakan untuk mengetahui kemiripan atau perbedaan dengan data lain. *PCA* dapat mereduksi dimensi data yang lebih rendah dengan resiko kehilangan informasi yang sangat kecil. (Smith & Lindsay, 2002).

*Silhouette Coefficient* merupakan sebuah metode evaluasi untuk menguji optimal atau ketepatan dari sebuah *cluster* yang telah terbentuk dari proses *clustering* (Furqon & Muflikah, 2016).

Muningsih dkk (2020), melakukan penelitian tentang penerapan metode *PCA* untuk *clustering* data kunjungan wisatawan mancanegara ke Indonesia diketahui pengolahan data yang dilakukan menggunakan metode *K-Means clustering* dan metode reduksi *PCA* mendapatkan hasil model *K-Means* dan *PCA* menghasilkan nilai validasi yang lebih baik dibanding model *K-Means* tanpa *PCA*. Alwi dan Hasrul (2018), melakukan penelitian tentang analisis *cluster* untuk pengelompokan kabupaten/kota di Provinsi Sulawesi Selatan berdasarkan indikator kesejahteraan rakyat mendapatkan hasil analisis *cluster* menggunakan penerapan *PCA* memberikan hasil lebih optimal dibandingkan dengan analisis *cluster* tanpa *PCA*. Talakua dkk (2017), melakukan penelitian tentang analisis *cluster* dengan menggunakan metode *K-Means* untuk pengelompokan kabupaten/kota di Provinsi Maluku berdasarkan indikator indeks pembangunan manusia tahun 2014 mendapatkan hasil bahwa analisis *cluster* pada *K-Means* baik digunakan dalam pengelompokan indikator indeks pembangunan manusia.

IPM merupakan salah satu cara untuk mengukur keberhasilan kinerja suatu Negara atau Wilayah dalam suatu bidang pembangunan manusia. IPM dapat mengukur capaian pembangunan manusia berbasis sejumlah komponen dasar kualitas hidup yaitu umur panjang dan hidup sehat, pengetahuan, dan standar hidup layak. IPM memberikan gambaran mengenai tingkat pencapaian pembangunan manusia sebagai dampak dari kegiatan pembangunan manusia yang dilakukan oleh suatu Negara atau daerah (Badan Pusat Statistika, 2020).

Tujuan dari penelitian ini adalah untuk memperoleh nilai *silhouette coefficient* hasil analisis *cluster* terbaik dan mengetahui hasil pengelompokan optimal pada kabupaten/kota di Pulau Kalimantan pada indikator IPM Tahun 2020 dengan menggunakan algoritma *K-Means Cluster* dengan *Principle Component Analysis*.

### Analisis Multivariat

Analisis multivariat merupakan analisis yang berhubungan dengan semua teknik statistik yang secara simultan menganalisis sejumlah pengukuran objek (Widarjono, 2015).

Menurut Johnson & Wichern (2007), analisis multivariat dikelompokkan menjadi dua kelompok besar yaitu:

1. Metode dependensi
2. Metode interdependensi

### Principal Component Analysis (PCA)

*Principal Component Analysis (PCA)* merupakan teknik analisis yang digunakan untuk menyederhanakan suatu data, dengan cara mentransformasi data secara linier sehingga terbentuk sistem koordinat baru dengan varians maksimum. Tujuan utama *PCA* adalah untuk mengurangi dimensi peubah yang saling berhubungan dan cukup banyak variabel, sehingga lebih mudah untuk menginterpretasikan data-data tersebut. Metode yang digunakan yaitu menentukan komponen utama dengan kombinasi linier (Johnson & Wichern, 2007).

### Data Mining

Data mining merupakan Proses untuk menemukan informasi yang menarik dan tersembunyi dari suatu kumpulan data yang berukuran besar yang tersimpan dalam suatu basis data, data *warehouse* atau tempat penyimpanan data lainnya (Tan dkk, 2006). Menurut Prasetyo (2012), terdapat tiga langkah utama dalam *data mining* yaitu eksplorasi atau pemrosesan awal data, membangun model dan melakukan validasi, dan penerapan.

### Analisis Cluster

Teknik *data mining* yang digunakan untuk mengelompokkan himpunan objek (*dataset*) ke dalam beberapa grup/*cluster* di mana hanya berdasarkan pada kemiripan karakteristik dari atribut yang dimiliki sehingga data objek yang berada di dalam *cluster* yang sama memiliki kemiripan satu sama lain tetapi mereka tidak mirip dengan data objek yang berada dalam *cluster* yang berbeda disebut dengan analisis *cluster* (Han dkk, 2012).

Menurut Xu dan Donald (2009), algoritma *clustering* bekerja dengan mengelompokkan objek-objek data (pola, entitas, kejadian, unit dan

hasil observasi) ke dalam sejumlah *cluster* tertentu. Algoritma *cluster* melakukan pemisahan, pemecahan atau segmentasi ke dalam sejumlah kelompok (*cluster*) menurut karakteristik tertentu.

**K-Means Clustering**

*K-Means clustering* merupakan salah satu metode terbaik dan populer dalam algoritma *clustering* di mana *K-Means* mencari partisi yang optimal dari suatu data dengan cara meminimalkan kriteria jumlah kesalahan kuadrat dengan prosedur iterasi yang optimal (Muningsih & Kiswati, 2015).

**Silhouette Coefficient (SC)**

*Silhouette Coefficient* merupakan sebuah metode evaluasi untuk menguji optimal atau ketepatan dari sebuah *cluster* yang telah terbentuk dari proses *clustering* (Furqon & Muflikhah, 2016). *Silhouette Coefficient* memberikan hasil kualitas visual dalam setiap *cluster*, memberikan informasi sesuai dengan jumlah *cluster* pada *dataset*. *SC* adalah salah satu metode yang dapat digunakan untuk melihat kualitas dan kekuatan *cluster*, di mana metode ini merupakan metode validasi *cluster* yang menggabungkan metode *cohesion* dan *separation* (Kaufman & Rousseeuw, 1990).

**Metodologi Penelitian**

**Sumber Data dan Variabel Penelitian**

Penelitian ini menggunakan data sekunder yang diperoleh dari Badan Pusat Statistik. Variabel yang digunakan dalam penelitian ini adalah sebagai berikut:

- X<sub>1</sub> : Usia Harapan Hidup (Tahun)
- X<sub>2</sub> : Harapan Lama Sekolah (Tahun)
- X<sub>3</sub> : Rata-rata Lama Sekolah (Tahun)
- X<sub>4</sub> : Pengeluaran Perkapita (Ribu Rupiah/Kapita)
- X<sub>5</sub> : Laju Pertumbuhan Ekonomi (%)
- X<sub>6</sub> : Persentase Penduduk Miskin (%)
- X<sub>7</sub> : Tingkat Pengangguran Terbuka (%)
- X<sub>8</sub> : Jumlah Tenaga Kesehatan (Orang)

**Metode Penelitian**

1. Melakukan analisis statistika deskriptif digunakan untuk melihat gambaran umum data penelitian.
2. Melakukan reduksi data penelitian menggunakan metode *principal component analysis* dengan tahapan adalah sebagai berikut:

- a. Melakukan penormalisasian data menggunakan persamaan:

$$Z_{ap} = \frac{x_{ap} - \bar{x}_p}{s_p} \tag{1}$$

di mana:

Z<sub>ap</sub> : hasil normalisasi data

x<sub>ap</sub> : data ke-a pada variabel ke-p

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

s<sub>p</sub> : deviasi standar variabel ke-p

- b. Menghitung koefisien korelasi untuk melihat keeratan hubungan antar variabel menggunakan persamaan:

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

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

$$|\lambda I - R| = \vec{0} \tag{3}$$

dan vektor eigen pada persamaan:

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

- 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 menggunakan persamaan:

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

- g. Menghitung transformasi *dataset* baru hasil reduksi dengan *PCA* menggunakan persamaan:

$$PC_{at} = \vec{v}_{at1}Z_1 + \vec{v}_{at2}Z_2 + \dots + \vec{v}_{ap}Z_{ap} \tag{6}$$

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

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

a = 1,2,3, ..., n dan i = 1,2,3, ..., k

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

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

i = 1,2,3, ..., k dan p = 1,2,3, ..., n

dengan:

C<sub>i</sub><sup>a</sup> : *centroid* pada *cluster* ke-i iterasi ke-a

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

n<sub>i</sub> : jumlah objek anggota *cluster* ke-i

x<sub>ap</sub> : 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. Menghitung nilai *Silhouette Coefficient* untuk mengetahui kualitas dari hasil suatu pengelompokan dengan tahapan adalah sebagai berikut:
  - a. Menghitung rata-rata jarak data ke-*i* dengan semua data pada *cluster* yang sama
 
$$a_i = \frac{1}{N_p - 1} \sum_{r=1}^{N_p - 1} d_{i,r}, r \neq i \quad (9)$$
  - b. Menghitung rata-rata jarak data ke-*i* dengan semua data pada *cluster* yang berbeda
 
$$b_i = \min\{d_i(p)\}, r \neq i \quad (10)$$
  - c. Menghitung nilai  $SC_1(i)$ 

$$SC_1(i) = \frac{b_i - a_i}{\max\{b_i - a_i\}}, i = 1, 2, 3, \dots, \quad (11)$$
  - d. Menghitung rata-rata nilai  $SC_2(p)$ 

$$SC_2(p) = \frac{1}{N_p} \sum_{x_i \in C_p} SC_1(i) \quad (12)$$
  - e. Menghitung nilai  $SC$  global
 
$$SC = \frac{\sum_{p=1}^K (N_p \times SC_2(p))}{\sum_{p=1}^K N_p} \quad (13)$$
  - f. Menentukan nilai *k* optimal berdasarkan nilai *silhouette coefficient* terbesar
5. Menginterpretasikan hasil *cluster* berdasarkan nilai indeks validitas *Silhouette Coefficient*.

**Hasil dan Pembahasan**

**1. Statistika Deskriptif**

**Tabel 1.** Statistika Deskriptif Data Indikator Indeks Pembangunan Manusia di Kabupaten/Kota di Pulau Kalimantan Tahun 2020

Variabel	Minimum	Maksimum	Rata-rata
X <sub>1</sub>	63,83	74,49	70,63
X <sub>2</sub>	11,16	15,00	12,72
X <sub>3</sub>	6,01	11,52	8,32
X <sub>4</sub>	6.804	16.278	10.661
X <sub>5</sub>	-1,21	0,65	-0,004
X <sub>6</sub>	2,55	12,04	5,97
X <sub>7</sub>	2,24	12,36	4,94
X <sub>8</sub>	167	4.587	1.186

Berdasarkan Tabel 1 dapat diketahui bahwa Usia harapan hidup (X<sub>1</sub>) memiliki nilai minimum 63,83 tahun, nilai maksimum 74,49 tahun dan rata-rata 70,63 tahun. Harapan lama sekolah (X<sub>2</sub>) memiliki nilai minimum 11,16 tahun, nilai maksimum 15,00 tahun dan rata-rata 12,75 tahun. Rata-rata lama sekolah (X<sub>3</sub>) memiliki nilai minimum 6,01 tahun, nilai maksimum 11,52 tahun dan rata-rata 8,32 tahun. Pengeluaran perkapita (X<sub>4</sub>) memiliki nilai minimum Rp. 6.804, nilai maksimum Rp. 16.278 dan rata-rata Rp. 10.611. Laju pertumbuhan ekonomi (X<sub>5</sub>) memiliki nilai minimum 1,21, nilai maksimum

0,65%. Dan rata-rata -0,004%. Persentase penduduk miskin (X<sub>6</sub>) memiliki nilai minimum 2,55%, nilai maksimum 12,04% dan rata-rata 5,97%. Tingkat pengangguran (X<sub>7</sub>) memiliki nilai minimum 2,24%, nilai maksimum 12,36% dan rata-rata 5,97%. Jumlah tenaga kesehatan (X<sub>8</sub>) terbuka memiliki nilai minimum 167 orang, nilai maksimum 4.587 orang dan rata-rata 1.186 orang.

**2. Principal Component Analysis (PCA)**

Adapun tahapan-tahapan pada PCA adalah sebagai berikut:

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

**Tabel 2.** Normalisasi Data Indikator Indeks Pembangunan Manusia di Kabupaten/Kota di Pulau Kalimantan Tahun 2020

Data	X <sub>1</sub>	X <sub>2</sub>	...	X <sub>8</sub>
1	-0,703	-0,129	...	0,417
2	1,325	-0,735	...	0,138
3	0,915	-0,374	...	-0,150
⋮	⋮	⋮	⋮	⋮
56	1,383	-1,513	...	0,164

- b. Koefisien Korelasi

**Tabel 3.** Koefisien Korelasi Indikator Indeks Pembangunan Manusia di Kabupaten/Kota di Pulau Kalimantan Tahun 2020

	X <sub>1</sub>	X <sub>2</sub>	...	X <sub>7</sub>	X <sub>8</sub>
X <sub>1</sub>	1,000	0,353	...	0,450	0,419
X <sub>2</sub>	0,353	1,000	...	0,631	0,539
⋮	⋮	⋮	⋮	⋮	⋮
X <sub>7</sub>	0,450	0,631	...	1,000	0,613
X <sub>8</sub>	0,419	0,539	...	0,613	1,000

Berdasarkan Tabel 3, dapat dilihat bahwa terdapat hubungan atau korelasi yang sedang antara variabel X<sub>1</sub> dengan X<sub>7</sub> yaitu sebesar 0,450, hubungan atau korelasi yang kuat antara variabel X<sub>2</sub> dengan X<sub>3</sub> yaitu sebesar 0,755 begitupun seterusnya.

- c. Matriks Varians Kovarians  
Matriks varians kovarians dibentuk berdasarkan nilai koefisien korelasi.

$$R = \begin{pmatrix} 1 & 0,353 & \dots & 0,450 & 0,419 \\ 0,353 & 1 & \dots & 0,631 & 0,539 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0,450 & 0,631 & \dots & 1 & 0,613 \\ 0,419 & 0,539 & \dots & 0,613 & 1 \end{pmatrix}$$

- d. Nilai Eigen dan Vektor Eigen  
Nilai eigen ( $\lambda$ )

$$\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \lambda_7 \\ \lambda_8 \end{pmatrix} = \begin{pmatrix} 3,594 \\ 1,805 \\ 0,939 \\ 0,519 \\ 0,421 \\ 0,333 \\ 0,287 \\ 0,099 \end{pmatrix}$$

Berdasarkan nilai eigen, terdapat delapan nilai eigen yang diperoleh. Dalam menentukan banyaknya *PCA* yang akan dipilih menggunakan kriteria nilai eigen  $\geq 1$ . Sehingga dari 8 nilai eigen diperoleh dua nilai eigen yang memenuhi kriteria tersebut, yaitu  $\lambda_1 = 3,594$  dan  $\lambda_2 = 1,805$ .

Vektor eigen ( $\vec{v}$ )

$$\vec{v} = \begin{pmatrix} -0,266 & 0,375 & \dots & 0,059 & -0,239 \\ -0,464 & 0,005 & \dots & 0,537 & -0,394 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -0,420 & 0,068 & \dots & -0,445 & 0,152 \\ -0,366 & 0,153 & \dots & 0,269 & 0,182 \end{pmatrix}$$

- e. Komponen Korelasi  
Komponen korelasi menunjukkan besarnya korelasi variabel terhadap skor komponen yang terbentuk.

**Tabel 4.** Komponen Matriks Korelasi

	$PC_1$	$PC_2$
$X_1$	-0,504	0,638
$X_2$	-0,880	0,007
$X_3$	-0,851	-0,026
$X_4$	-0,751	-0,496
$X_5$	0,191	-0,688
$X_6$	0,356	0,791
$X_7$	-0,694	0,091
$X_8$	-0,693	0,206

Berdasarkan Tabel 4, dapat disimpulkan bahwa korelasi variabel  $X_1$  terhadap  $PC_1$  sebesar -0,504 dan terhadap  $PC_2$  sebesar 0,638. Karena korelasi variabel  $X_1$  terhadap  $PC_2$  lebih besar maka variabel  $X_1$  tergabung dengan  $PC_2$ . Korelasi variabel  $X_2$  terhadap  $PC_1$  sebesar -0,880 dan terhadap  $PC_2$  sebesar 0,007. Karena korelasi variabel  $X_2$  terhadap  $PC_1$  lebih besar maka variabel  $X_2$  tergabung dengan  $PC_1$  dan seterusnya sampai dengan variabel  $X_8$ .

- f. Persamaan *Principal Component*  
Persamaan *principal component* terbentuk berdasarkan vektor eigen, maka diperoleh persamaan sebagai berikut:

$$PC_1 = -0,266Z_1 - 0,464Z_2 - 0,499Z_3 - 0,393Z_4 + 0,101Z_5 + 0,188Z_6 - 0,420Z_7 - 0,366Z_8$$

$$PC_2 = 0,475Z_1 + 0,005Z_2 - 0,019Z_3 - 0,369Z_4 - 0,512Z_5 + 0,589Z_6 - 0,068Z_7 - 0,153Z_8$$

- 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	1,235	0,262
2	1,192	0,425
3	1,792	2,197
4	0,953	0,322
5	1,892	0,634
...	...	...
56	-0,656	1,180

**3. K-Means Cluster**

Langkah-langkah pengelompokan *K-Means cluster* adalah sebagai berikut:

- Menentukan banyaknya *cluster* yang akan digunakan dengan menggunakan data set baru hasil perhitungan *PCA*, dengan *cluster* yang digunakan adalah  $k=2,3,4$ , dan 5.
- Menentukan pusat *cluster* awal  
Pada penelitian ini diambil contoh dengan menggunakan untuk *cluster*  $k = 2$ . Pusat *cluster* awal dipilih secara acak dari data yang telah ditransformasi menggunakan *PCA*. Penentuan pusat *cluster* awal dilakukan secara *trial* dan *error*. Sehingga dipilih data ke-5 dan data ke-36 sebagai berikut:

**Tabel 6.** Pusat *Cluster* ( $c^{(0)}$ ) Untuk  $k = 2$

<i>Cluster</i> ( $c_{i,j}$ )	Variabel	
	$PC_1$	$PC_2$
1	1,892	0,634
2	1,381	-1,366

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

**Tabel 7.** Pusat *Cluster* ( $c^{(0)}$ ) Untuk  $k = 2$

Data	Jarak Euclid	
	<i>Cluster</i> 1	<i>Cluster</i> 2
1	0,755	1,635
2	0,731	1,801
3	1,566	3,587
...	...	...
56	2,606	3,261

- 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 23 kabupaten/kota. Sedangkan *cluster* 2 beranggotakan 33 kabupaten/kota.

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

**Tabel 8.** Pusat *Cluster* Baru ( $c^{(1)}$ ) Untuk  $k = 2$

Cluster ( $c_{i,j}$ )	Variabel	
	PC <sub>1</sub>	PC <sub>2</sub>
1	0,761	1,358
2	-0,485	-0,900

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

**Tabel 9.** Jarak Euclid Terhadap Pusat *Cluster* Baru ( $c^{(1)}$ ) Untuk  $k = 2$

Data	Jarak Euclid	
	Cluster 1	Cluster 2
1	1,194	2,076
2	1,028	2,137
3	1,329	3,844
...	...	...
56	1,428	2,087

- g. Mengalokasikan data penelitian ke pusat *cluster* terdekat iterasi ke-1  
Mengalokasikan data pengamatan ke pusat *cluster* terdekat iterasi ke-2 menggunakan persamaan 8. Maka diperoleh bahwa *cluster* 1 beranggotakan 24 kabupaten/kota, sedangkan *cluster* 32 beranggotakan 5 kabupaten/kota.
- h. Menghitung pusat *cluster* baru  
Pusat *cluster* baru ditentukan berdasarkan keanggotaan *cluster* hingga tidak ada lagi *cluster* yang berpindah tempat. Diketahui bahwa hasil alokasi data iterasi ke-6 sama seperti alokasi iterasi ke-5, di mana tidak terdapat perubahan pada keanggotaan *cluster* atau tidak ada lagi data pengamatan yang berpindah *cluster*. Sehingga pusat *cluster* baru ke-5 akan sama dengan pusat *cluster* ke-6.

**Tabel 10.** Hasil Pengelompokan *K-Means Cluster* terhadap *PCA*

Clus Ter	Anggota Cluster	
	Kode	Kabupaten/Kota
1	1, 2, 3, 4,	Sambas, Bengkayang,
	5, 6, 7, 8,	Landak, Mempawah,
	9, 10, 11,	Sanggau, Ketapang,
	22, 27,	Sintang, Kapuas Hulu,
	42, 43,	Sekadau, Melawi,
	45, 47	Kayong Utara, Seruyun,

**Tabel 10.** Hasil Pengelompokan *K-Means Cluster* terhadap *PCA* (lanjutan)

Clus Ter	Anggota Cluster	
	Kode	Kabupaten/Kota
1	48, 53,	Murung Raya, Paser, Kutai Barat, Kutai Timur, Penajam Paser Utara, Mahakam Ulu, Bulungan, Tana Tidung, Nunukan, dan Kota Tarakan
	54, 55,	Kubu Raya, Pontianak, Kota Singkawang, Kotawaringin Barat, Kotawaringin Timur,
	dan 56	12, 13, Kapuas, Barito Selatan, Barito Utara, Sukamara, Lamadau, Katingan, Pulang Pisau, Gunung Mas, Barito Timur, Kota Palangka Raya, Tanah Laut, Kota Baru,
		14, 15, 16, 17, 18, 19, 20, 21, 23, 24, 25, 26, 28, 29,
		30, 31, Banjar, Barito Kuala, Tapin, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Tabalong, Tanah Bumbu, Balangan, Kota Banjarmasin, Kota Banjar Baru, Kutai Kartanegara, Berau, Kota Balikpapan, Kota Samarinda, Kota Bontang, dan Malinau
2	12, 13,	Kotawaringin Barat, Kotawaringin Timur,
	14, 15,	Kapuas, Barito Selatan, Barito Utara, Sukamara, Lamadau, Katingan, Pulang Pisau, Gunung Mas, Barito Timur, Kota Palangka Raya, Tanah Laut, Kota Baru,
	16, 17,	Barito Utara, Sukamara, Lamadau, Katingan, Pulang Pisau, Gunung Mas, Barito Timur, Kota Palangka Raya, Tanah Laut, Kota Baru,
	18, 19,	Lamadau, Katingan, Pulang Pisau, Gunung Mas, Barito Timur, Kota Palangka Raya, Tanah Laut, Kota Baru,
	20, 21,	Pulang Pisau, Gunung Mas, Barito Timur, Kota Palangka Raya, Tanah Laut, Kota Baru,
	22, 23,	Mas, Barito Timur, Kota Palangka Raya, Tanah Laut, Kota Baru,
	24, 25,	Mas, Barito Timur, Kota Palangka Raya, Tanah Laut, Kota Baru,
	26, 27,	Kota Palangka Raya, Tanah Laut, Kota Baru,
	28, 29,	Tanah Laut, Kota Baru,
	30, 31,	Banjar, Barito Kuala, Tapin, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Tabalong, Tanah Bumbu, Balangan, Kota Banjarmasin, Kota Banjar Baru, Kutai Kartanegara, Berau, Kota Balikpapan, Kota Samarinda, Kota Bontang, dan Malinau
32, 33,	Tapin, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Tabalong, Tanah Bumbu, Balangan, Kota Banjarmasin, Kota Banjar Baru, Kutai Kartanegara, Berau, Kota Balikpapan, Kota Samarinda, Kota Bontang, dan Malinau	
34, 35,	Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Tabalong, Tanah Bumbu, Balangan, Kota Banjarmasin, Kota Banjar Baru, Kutai Kartanegara, Berau, Kota Balikpapan, Kota Samarinda, Kota Bontang, dan Malinau	
36, 37,	Tengah, Hulu Sungai Utara, Tabalong, Tanah Bumbu, Balangan, Kota Banjarmasin, Kota Banjar Baru, Kutai Kartanegara, Berau, Kota Balikpapan, Kota Samarinda, Kota Bontang, dan Malinau	
38, 39,	Utara, Tabalong, Tanah Bumbu, Balangan, Kota Banjarmasin, Kota Banjar Baru, Kutai Kartanegara, Berau, Kota Balikpapan, Kota Samarinda, Kota Bontang, dan Malinau	
40, 41,	Bumbu, Balangan, Kota Banjarmasin, Kota Banjar Baru, Kutai Kartanegara, Berau, Kota Balikpapan, Kota Samarinda, Kota Bontang, dan Malinau	
42, 43,	Banjarmasin, Kota Banjarmasin, Kota Banjar Baru, Kutai Kartanegara, Berau, Kota Balikpapan, Kota Samarinda, Kota Bontang, dan Malinau	
44, 45,	Banjarmasin, Kota Banjarmasin, Kota Banjar Baru, Kutai Kartanegara, Berau, Kota Balikpapan, Kota Samarinda, Kota Bontang, dan Malinau	
46, 47,	Banjarmasin, Kota Banjarmasin, Kota Banjar Baru, Kutai Kartanegara, Berau, Kota Balikpapan, Kota Samarinda, Kota Bontang, dan Malinau	
48, 49,	Banjarmasin, Kota Banjarmasin, Kota Banjar Baru, Kutai Kartanegara, Berau, Kota Balikpapan, Kota Samarinda, Kota Bontang, dan Malinau	
50, 51,	Banjarmasin, Kota Banjarmasin, Kota Banjar Baru, Kutai Kartanegara, Berau, Kota Balikpapan, Kota Samarinda, Kota Bontang, dan Malinau	
52, 53,	Banjarmasin, Kota Banjarmasin, Kota Banjar Baru, Kutai Kartanegara, Berau, Kota Balikpapan, Kota Samarinda, Kota Bontang, dan Malinau	
54, 55,	Banjarmasin, Kota Banjarmasin, Kota Banjar Baru, Kutai Kartanegara, Berau, Kota Balikpapan, Kota Samarinda, Kota Bontang, dan Malinau	
56, 57,	Banjarmasin, Kota Banjarmasin, Kota Banjar Baru, Kutai Kartanegara, Berau, Kota Balikpapan, Kota Samarinda, Kota Bontang, dan Malinau	

- i. Mengulang kembali dengan perhitungan yang sama untuk mendapatkan nilai  $k = 3, 4$  dan  $5$ . Anggota dari *cluster* yang terbentuk adalah sebagai berikut:

**Tabel 11.** Anggota *Cluster* Yang Terbentuk

cluster	Anggota Cluster
2	{(1,2,...,56),(12,13,...,52)}
3	{(9,12,...,39),(13,28,...,51),(1,2,...,56)}
4	{(9,12,...,39),(49,50),(1,2,...,56),(14,28,...,51)}
5	{(1,2,...,27),(14,42,...,56),(12,15,...,39),(2,6,...,55),(13,28,...,51)}

**4. Silhouette Coefficient (SC)**

Adapun tahapan-tahapan pada *PCA* adalah sebagai berikut:

- a. Menghitung rata-rata jarak data ke- $i$  dengan semua data yang berada pada *cluster* yang sama dengan menggunakan persamaan:

$$a_1 = \frac{1}{N_1 - 1} \sum_{r=1}^{N_1-1} d_{1,r}$$

$$\begin{aligned}
 &= \frac{1}{22-1} \sum_{r=1}^{21} d_{1,r} \\
 &= \frac{1}{22} [d(1,2) + d(1,3) + \dots + d(1,56)] \\
 &= 1,479
 \end{aligned}$$

Dengan cara yang sama dihitung semua nilai  $a_i$  untuk anggota dalam cluster 1 sampai dengan terakhir yaitu  $a_{56}$ . Begitupun pada cluster 2.

- b. Menghitung rata-rata jarak data ke- $i$  dengan semua data yang berada pada cluster yang berbeda dengan menggunakan persamaan:

$$\begin{aligned}
 d_1(2) &= \frac{1}{N_2} \sum_{r=1}^{N_2} d_{1,r} \\
 &= \frac{1}{34} [d(1,12) + d(1,13) + \dots + d(1,52)] \\
 &= \frac{1}{34} [1,661 + 6,384 + \dots + 1,769] \\
 &= \frac{1}{34} [86,387] \\
 &= 2,541
 \end{aligned}$$

Selanjutnya dengan cara yang sama dihitung semua nilai  $d_i$  untuk anggota dalam cluster 1 sampai dengan terakhir yaitu  $d_{56}$ . Begitupun pada cluster 2.

- c. Menghitung nilai  $SC_1$  data ke- $i$  dengan menggunakan persamaan:

$$\begin{aligned}
 SC_1(1) &= \frac{b_1 - a_1}{\max\{b_1, a_1\}} \\
 &= \frac{2,541 - 1,479}{\max\{(2,451), (1,479)\}} \\
 &= 0,418
 \end{aligned}$$

Demikian seterusnya dengan cara yang sama dihitung  $SC_1(i)$  untuk  $i = 1, 2, \dots, 56$ .

- d. Menghitung rata-rata nilai  $SC_2$  pada cluster ke- $p$ . Perhitungan nilai  $SC_1$  untuk cluster 1 adalah sebagai berikut:

$$\begin{aligned}
 SC_2(1) &= \frac{1}{N_1} \sum_{x_i \in C_1} SC_1(i) \\
 &= \frac{1}{22} (SC_1(1) + SC_1(2) + \dots + SC_1(56)) \\
 &= \frac{1}{22} (0,418 + 0,446 + \dots + 0,270) \\
 &= 0,435
 \end{aligned}$$

Perhitungan nilai  $SC_2$  untuk cluster 2 adalah sebagai berikut:

$$\begin{aligned}
 SC_2(1) &= \frac{1}{N_2} \sum_{x_i \in C_1} SC_1(i) \\
 &= \frac{1}{34} (SC_1(12) + SC_1(13) + \dots + SC_1(52)) \\
 &= \frac{1}{34} (0,253 + 0,318 + \dots + (-0,208)) \\
 &= 0,310
 \end{aligned}$$

- e. Menghitung nilai  $SC$  Global dengan menggunakan persamaan:

$$\begin{aligned}
 SC &= \frac{\{(n_1 \times SC_2(1)) + (n_2 \times SC_2(2))\}}{(n_1 + n_2)} \\
 &= \frac{\{(22 \times 0,435) + (34 \times 0,310)\}}{(22 + 34)} \\
 &= 0,359
 \end{aligned}$$

Hasil validasi untuk 2 cluster berdasarkan nilai  $SC$  global adalah sebesar 0,359 artinya pengelompokan secara keseluruhan dapat dikatakan *weak structure* (struktur cluster yang lemah/buruk), artinya sedikit sekali kumpulan objek  $i$  yang berada dalam cluster yang tepat dan sisanya bergabung pada cluster yang tidak semestinya.

Dilakukan perhitungan  $SC$  seperti langkah 1 sampai dengan langkah 5 untuk clustering dengan metode *K-Means* pada pembentukan 3 cluster, 4 cluster dan 5 cluster.

- f. Menentukan nilai  $k$  optimal berdasarkan nilai *silhouette coefficient* terbesar. Peneliti membandingkan nilai  $SC$  Global dari hasil pembentukan 2, 3, 4 dan 5 cluster yang dapat dilihat pada Tabel berikut ini:

Jumlah Cluster	$SC$
2	0,359
3	0,467
4	0,540
5	0,445

Berdasarkan Tabel 12, dapat diketahui bahwa nilai  $SC$  untuk evaluasi hasil pengelompokan data indeks pembangunan manusia berdasarkan variabel yang diduga sebagai penentu rendahnya indeks pembangunan manusia di kabupaten/kota di Pulau Kalimantan dengan metode *K-Means Cluster* dengan *Principle Component Analysis* memiliki nilai  $SC$  yang berbeda-beda. Nilai  $SC$  yang terbesar adalah pengelompokan pada 4 cluster yaitu 0,540. Oleh karena itu, dapat diputuskan bahwa pengelompokan yang paling optimal pada pengelompokan kabupaten/kota di Pulau Kalimantan dengan metode *K-Means Cluster* dengan *Principle Component Analysis* adalah 4 cluster. Oleh karena itu, dapat diputuskan bahwa pengelompokan yang paling optimal pada pengelompokan kabupaten/kota di Pulau Kalimantan dengan metode *K-Means Cluster* dengan *Principle Component Analysis* adalah 4 cluster. Cluster 1 beranggotakan 26 kabupaten/kota yang terdiri dari kabupaten/kota Sekadau, Kubu Raya, Kotawaringin Barat, Kotawaringin Utara, Kapuas, Barito Selatan, Barito Utara, Sukamara, Lamandau, Seruyun, Katingan, Pulang Pisau, Gunung Mas, Barito Timur, Murung Raya, Tanah Laut, Kota Baru, Banjar, Barito Kuala, Tapin, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara,

Tabalong, Tanah Bumbu dan Balangan. *Cluster 2* beranggotakan 3 kabupaten/kota yang terdiri dari kabupaten/kota Pontianak, Kota Balikpapan dan Kota Samarinda. *Cluster 3* beranggotakan 20 kabupaten/kota yang terdiri dari kabupaten/kota Sambas, Bengkayang, Landak, Mempawah, Sanggau, Ketapang, Sintang, Kapuas Hulu, Melawi, Kayong Utara, Paser, Kutai Barat, Kutai Timur, Penajam Paser Utara, Mahakam Ulu, Malinau, Bulungan, Tana Tidung, Nunukan dan Kota Tarakan. Sedangkan *Cluster 4* beranggotakan 7 kabupaten/kota yang terdiri dari kabupaten/kota Singkawang, Kota Palangka Raya, Kota Banjarmasin, Kota Banjar Baru, Kutai Kartanegara, Berau dan Kota Bontang.

**Karakteristik Hasil Analisis Cluster**

Berdasarkan hasil pengelompokan 4 *cluster*, dilakukan perhitungan nilai rata-rata variabel untuk melihat karakteristik dari masing-masing *cluster*.

**Tabel 13.** Nilai rata-rata pada setiap *cluster*

Variabel	Cluster ke- <i>i</i>			
	1	2	3	4
X <sub>1</sub>	68,80	73,91	71,88	72,44
X <sub>2</sub>	12,42	14,68	12,45	13,79
X <sub>3</sub>	8,00	10,44	7,84	9,98
X <sub>4</sub>	10.834	13.384	8.380	13.243
X <sub>5</sub>	0,19	-0,04	-0,23	-0,07
X <sub>6</sub>	4,77	4,01	8,25	4,75
X <sub>7</sub>	4,05	6,98	4,67	9,87
X <sub>8</sub>	794	3.762	1.160	1.620

Berdasarkan Tabel 13, dapat diketahui bahwa *Cluster 1* merupakan *cluster* di mana variabel laju pertumbuhan ekonomi lebih tinggi jika dibandingkan dengan *cluster 2, 3* dan *4*. Hal ini berarti kabupaten/kota pada *cluster 1* cenderung memiliki usia harapan hidup, harapan lama sekolah, rata-rata lama sekolah, pengeluaran perkapita, persentase penduduk miskin, tingkat pengangguran terbuka dan jumlah tenaga kesehatan kabupaten/kota yang rendah jika dibandingkan dengan *cluster 2, 3* dan *4*. *Cluster 2* merupakan *cluster* di mana variabel usia harapan hidup, harapan lama sekolah, rata-rata lama sekolah, pengeluaran perkapita dan jumlah tenaga kesehatan kabupaten/kota yang lebih tinggi dibandingkan dengan *cluster 1, 3* dan *4*. Hal ini berarti kabupaten/kota pada *cluster 2* cenderung memiliki laju pertumbuhan ekonomi, persentase penduduk miskin dan tingkat pengangguran terbuka yang rendah jika dibandingkan dengan *cluster 1, 3* dan *4*. *Cluster 3* merupakan *cluster* di mana variabel persentase penduduk miskin lebih tinggi jika dibandingkan dengan *cluster 1, 2* dan *4*. Hal ini berarti kabupaten/kota pada *cluster 3* cenderung memiliki usia harapan hidup, harapan lama sekolah, rata-rata lama sekolah, pengeluaran perkapita, laju pertumbuhan ekonomi, tingkat

pengangguran terbuka dan jumlah tenaga kesehatan kabupaten/kota yang rendah jika dibandingkan dengan *cluster 1, 2* dan *4*. *Cluster 4* merupakan *cluster* di mana variabel tingkat pengangguran terbuka lebih tinggi jika dibandingkan dengan *cluster 1, 2* dan *3*. Hal ini berarti kabupaten/kota pada *cluster 4* cenderung memiliki usia harapan hidup, harapan lama sekolah, rata-rata lama sekolah, pengeluaran perkapita, laju pertumbuhan ekonomi, persentase penduduk miskin, dan jumlah tenaga kesehatan kabupaten/kota yang rendah jika dibandingkan dengan *cluster 1, 2* dan *3*.

**Kesimpulan**

1. Nilai *Silhouette Coefficient* untuk pengelompokan kabupaten/kota di Pulau Kalimantan berdasarkan indikator indeks pembangunan manusia dengan menggunakan algoritma *K-Means Cluster* dengan *Principle Component Analysis* menghasilkan 4 *cluster* dengan nilai sebesar 0,540 yang menyatakan bahwa *structure cluster* yang terbentuk pada pengelompokan ini adalah *medium structure*.
2. *Cluster* optimal yang terbentuk pada pengelompokan kabupaten/kota di Pulau Kalimantan berdasarkan indikator indeks pembangunan manusia dengan menggunakan algoritma *K-Means Cluster* dengan *Principle Component Analysis* adalah sebanyak 4 *cluster* yaitu *cluster 1, cluster 2, cluster 3* dan *cluster 4*. *Cluster 1* beranggotakan 24 kabupaten/kota, *cluster 2* beranggotakan 3 kabupaten/kota, *cluster 3* beranggotakan 20 kabupaten/kota dan *cluster 4* beranggotakan 7 kabupaten/kota.

**Daftar Pustaka**

Alwi, W. & Hasrul, M. (2018). Analisis *Cluster* untuk Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat. *Jurnal MSA* 35-41.

Badan Pusat Statistik. (2020). *Indeks Pembangunan Manusia 2020*. Indonesia: Badan Pusat Statistika Republik Indonesia.

Ediyanto, M.N. (2013). Pengklasifikasian Karakteristik dengan Metode K-Means Cluster Analysis. *Buletin Ilmiah Mat, Stat, dan terapannya*, 133-136.

Furqon, M.T., & Muflikhah, L. (2016). Clustering the Potential Risk of Tsunami Using Density-Based Spatial Clustering of Application With Noise (DbSCAN). *Jurnal of Enviromental Engineering and Sustainable Technology* 1-8.

- Hair, J. F., Black, W. C., Babin, B. J. & Anderson, R. E. (2014). *Multivariate Data Analysis 7<sup>th</sup> Edition*. New York: Pearson Education, Inc.
- Han, J., Kamber, M. & Pei, J. (2012). *Data Mining Concepts and Techniques Third Edition*. New York: Elsevier.
- Johnson, R. A. and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. New Jersey: Pearson Prentice Hall.
- Kaufman, E., & Rousseeuw, P.J. (1990). *Finding Groups in Data*. New York: John Wiley & Sons.
- Muningsih, E., & Kiswati, S. (2015). Penerapan Metode K-Means untuk Clustering Produk Online Shop dalam Penentuan Stok Barang. *Jurnal Bianglala Informatika* 10-17.
- Muningsih, E., Hasan, N., & Sulisty, G.B. (2020). Penerapan Metode Principal Component Analysis (PCA) untuk Clustering Data Kunjungan Wisatawan Mancanegara ke Indonesia. *Jurnal Bianglala Informatika* 58-62.
- Prasetyo, E. (2012). *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Andi.
- Smith I.L., & Lindsay. (2002). *A Tutorial on Principal Component Analysis*. New York: Cornell University.
- Supranto, J. (2010). *Analisis Multivariate Arti dan Interpretasi, cet. Kedua*. Jakarta: Rineka Cipta.
- Tan, P., Steinbach, M. & Kumar, V. (2006). *Introduction to Data Mining*. New York: Pearson Education, Inc.
- Talakua, M.W., Leleury, Z.A., & Talluta, A.W. (2017). Analisis Cluster dengan Menggunakan Metode K-Means untuk Pengelompokan Kabupaten/Kota di Provinsi Maluku berdasarkan Indikator Indeks Pembangunan Manusia tahun 2014. *Jurnal Ilmu Matematika dan Terapan* 119-127.
- Wahono, R.S. (2015). Algoritma Cluster Dinamik untuk Optimasi Cluster pada Algoritma K-Means dalam Pemetaan Nasabah Potensial. *Journal of Intelligent Systems* 33-36.
- Windarjono, A. (2015). *Analisis Multivariat Terapan dengan Program SPSS, AMOS, dan SMARTPLS Edisi Kedua*. Yogyakarta: UPM STIM YKPN.
- Xu, R. And Donald, C. W. (2009). *Clustering*. Hoboken: Wiley.

