



## PEMETAAN KABUPATEN/KOTA DI PROVINSI PAPUA BERDASARKAN INDIKATOR DAERAH TERTINGGAL DENGAN METODE *ENSEMBLE CLUSTERING*

Evi Fauziyari\*, Dhoriva Urwatul Wustqa

Program Studi Statistika, Universitas Negeri Yogyakarta

\*e-mail: [evi8825fmipa.2018@student.uny.ac.id](mailto:evi8825fmipa.2018@student.uny.ac.id); [dhoriva\\_uw@uny.ac.id](mailto:dhoriva_uw@uny.ac.id)

**Abstrak** Provinsi Papua menjadi provinsi yang memiliki daerah tertinggal terbanyak di Indonesia. Untuk pemerataan pembangunan, perlu diketahui kabupaten/kota di Provinsi Papua yang termasuk dalam daerah sangat tertinggal dengan memetakan daerah tertinggal berdasarkan indikator daerah tertinggal menggunakan metode *ensemble clustering*. Terdapat 29 kabupaten dan 13 variabel yang digunakan dengan 10 variabel data bertipe numerik dan 3 variabel bertipe kategorik. Pembentukan *cluster* diawali dengan menentukan banyak *cluster*. Kemudian dilakukan pembentukan *cluster* dengan metode *K-means* untuk variabel numerik dan *K-modes* untuk variabel kategorik. Validasi dilakukan untuk mendapatkan *cluster* terbaik menggunakan *Davies Index Bouldin* dan *Silhouette*. Hasil dari kedua metode yang berupa data kategorik dianalisis dengan metode *K-modes*. Hasil penelitian menghasilkan 5 *cluster* terbaik dan dipetakan menjadi 5 tingkat daerah tertinggal. Setiap tingkat memiliki banyak kabupaten dan karakteristik yang berbeda.

**Kata Kunci:** *Daerah Tertinggal, Davies Index Bouldin, Ensemble Clustering, Silhouette.*

**Abstrac** Papua Province is the province with the most underdeveloped districts in Indonesia. For equitable development, it is necessary to know the regencies/ districts in Papua Province which are included in very underdeveloped districts by mapping underdeveloped districts based on indicators of underdeveloped districts using the ensemble clustering method. There are 29 districts and 13 variables used with 10 data variables of numeric type and 3 variables of categorical type. Cluster formation begins with determining the number of clusters. Then, cluster formation was performed using the *K-means* method for numerical variables and *K-modes* for categorical variables. Validation is done to get the best cluster using *Davies Index Bouldin* and *Silhouette*. The results of the two methods in the form of categorical data were analyzed by the *K-modes* method. The results of the study produced the best 5 clusters and were mapped into 5 levels of underdeveloped districts. Each level has many districts and different characteristics.

**Keywords:** *Davies Index Bouldin, Ensemble Clustering, Silhouette, Underdeveloped District*

## PENDAHULUAN

Indonesia merupakan salah satu negara berkembang yang terus berupaya untuk meningkatkan kesejahteraan masyarakat dengan kebijakan-kebijakan yang telah ditetapkan. Salah satu upaya yang dilakukan adalah dengan kebijakan tentang pembangunan nasional. Di era globalisasi ini, Indonesia semakin gencar untuk melakukan pembangunan di berbagai daerah. Namun, dengan maraknya kegiatan pembangunan belum tentu dapat mensejahterakan masyarakat. Terdapat kecenderungan bahwa daerah yang telah berkembang akan menyerap lebih banyak investasi dan intervensi pembangunan (Budianta, 2010). Artinya, terdapat perbedaan atau ketidakseimbangan yang terjadi antar daerah. Perbedaan tersebut dapat memunculkan adanya daerah yang lebih tertinggal dibandingkan dengan daerah lainnya. Berdasarkan Peraturan Pemerintah No.78 Tahun 2014, daerah tertinggal merupakan daerah yang masyarakat dan wilayahnya kurang berkembang dibandingkan daerah lain dalam skala nasional. Dalam Peraturan Presiden No 63 Tahun 2020, suatu daerah ditetapkan sebagai berdasarkan kriteria yaitu perekonomian masyarakat, sumber daya manusia, sarana dan prasarana, kemampuan keuangan daerah, aksesibilitas, dan karakteristik daerah. Dalam peraturan tersebut juga ditetapkan 62 daerah sebagai daerah tertinggal yang didominasi oleh daerah dari wilayah Indonesia bagian timur.

Dari 11 provinsi yang termasuk didalam status daerah tertinggal, Provinsi Papua memiliki wilayah kabupaten paling banyak yang masuk dalam status daerah tertinggal. Terdapat 22 kabupaten dari 29 provinsi Papua yang termasuk dalam daerah tertinggal. Banyaknya daerah tertinggal di Provinsi Papua dapat dilihat dengan rata-rata Indeks Desa Membangun (IDM) yang bernilai paling rendah dari provinsi yang lainnya, yakni 0,4563 (Dihni & Mutia, 2021). Mantan Menteri Desa, Pembangunan Daerah Tertinggal dan Transmigrasi Eko Putro Sandjojo pernah mengatakan bahwa masalah akses di Papua masih menjadi kendala dalam meningkatkan status desa. Wilayah Papua minim infrastruktur utama seperti jalan dan jembatan. Penyebab ketertinggalan suatu daerah kabupaten di wilayah Papua menurut perspektif pemerintah pusat adalah aksesibilitas, ekonomi dan Sumber Daya Manusia (Badan Perencanaan Pembangunan Daerah Provinsi Papua, 2020). Perlu kebijakan yang tepat untuk menangani masalah desa tertinggal sehingga dapat memberikan kemajuan pembangunan untuk Indonesia, khususnya Provinsi Papua. Untuk dapat menetapkan kebijakan yang tepat, diperlukan pemetaan terhadap kabupaten/kota di Provinsi Papua berdasarkan indikator daerah tertinggal.

Pemetaan tersebut dapat dilakukan dengan menggunakan teknik *clustering*. Klusterisasi (*clustering*) adalah proses pengelompokan himpunan data ke dalam beberapa grup atau *cluster* sedemikian hingga objek-objek dalam suatu *cluster* memiliki kemiripan yang tinggi, namun sangat berbeda (ketidakmiripan yang tinggi) dengan objek-objek di *cluster-cluster* lainnya. Metode yang digunakan untuk klusterisasi secara umum dibagi menjadi 2 yaitu klusterisasi berbasis hierarki dan klusterisasi berbasis partisi. Pengelompokan berbasis hierarki bekerja dengan mengelompokkan objek data ke dalam pohon dari *cluster*. Sedangkan pengelompokan berbasis partisi bekerja dengan mengatur objek ke dalam  $k$  partisi ( $k \leq n$ ), di mana setiap partisi mewakili sebuah *cluster* (Han et al., 2012).

Proses *clustering* dianggap hanya berfokus pada data dengan 1 (satu) tipe saja. Padahal, bisa saja dalam suatu pengelompokan terdapat dua tipe data, numerik dan kategorik (Hunt & Jorgensen, 2011). Data yang memiliki dua tipe data dapat disebut data campuran. Pada pengelompokan objek, data campuran tidak bisa hanya menggunakan satu metode pengelompokan saja, data campuran harus dibedakan dalam analisisnya. Metode *K-means* merupakan salah satu algoritma pengelompokan yang cocok digunakan data bertipe numerik.

Penelitian dengan menggunakan metode *K-means* pernah dilakukan oleh Hasugian (2020) untuk mengelompokkan desa yang ada di Padang Lawas untuk menggali potensi atau pengetahuan dari desa yang ada. Sedangkan untuk data bertipe kategorik, dapat digunakan metode *K-modes* yang merupakan pengembangan dari metode *K-means* dengan menggantikan rata-rata *cluster* dengan modus dan menggunakan metode berbasis frekuensi untuk memperbarui modus dalam proses *clustering* (Huang, 1998). Metode *ensemble clustering* dapat digunakan untuk menggabungkan kedua hasil pengelompokkan dengan metode yang berbeda. Dalam penelitian yang dilakukan Strehl & Ghosh (2002), mereka dapat menunjukkan bagaimana *ensemble clustering* dapat meningkatkan kualitas suatu kelompok *cluster*. Penelitian dengan menggunakan metode *ensemble clustering* pernah dilakukan oleh Yusfar (2020) untuk mengelompokkan kabupaten/kota yang ada di Sulawesi Selatan berdasarkan Indikator Kinerja Pembangunan Ekonomi Daerah.

Penelitian ini mengelompokkan kabupaten/kota di Provinsi Papua berdasarkan indikator daerah tertinggal. Data yang digunakan berjumlah 29 objek dengan variabel berskala numerik dan kategorik. Terdapat 10 variabel numerik yaitu rasio desa dengan permukaan jalan terluas, rasio rumah sakit terhadap jumlah penduduk, rasio SMA/MA/Sederajat terhadap penduduk usia SMA, persentase rumah tangga pengguna listrik, rasio derajat desentralisasi, persentase penduduk miskin, persentase pengeluaran penduduk perkapita, Angka Harapan Hidup, Indeks Pembangunan Manusia, dan Tingkat Pengangguran Terbuka. Untuk data berskala kategorik terdapat 3 variabel yaitu keberadaan kawasan industri, kekuatan sinyal telepon seluler, dan sumber air minum. Dalam penelitian ini, digunakan metode *ensemble clustering* karena data yang digunakan adalah data berskala campuran. Metode *ensemble clustering* berfungsi untuk menggabungkan dua hasil metode berskala campuran. Dua metode yang digabungkan yaitu metode *K-means* untuk data berskala numerik dan metode *K-modes* untuk data berskala kategorik.

## METODE

### Data

Data yang digunakan pada penelitian ini merupakan data indikator daerah tertinggal di Provinsi Papua. Data merupakan data campuran dengan variabel-variabel berskala numerik dan kategorik. Objek pengamatan berjumlah 29 kabupaten/kota dan terdapat 13 variabel yang mewakili kriteria-kriteria indikator daerah tertinggal. Pada Tabel 1 dan Tabel 2 merupakan variabel-variabel yang digunakan dalam penelitian ini.

**Tabel 1. Variabel Numerik**

Variabel	Kriteria	Keterangan
$X_1$	Aksesibilitas	Rasio Desa dengan Permukaan Jalan Terluas Aspal
$X_2$	Sarana Prasarana	Rasio Rumah Sakit terhadap Jumlah Penduduk
$X_3$		Rasio SMA/MA/Sederajat terhadap Penduduk Usia SMA Sederajat
$X_4$		Persentase Rumah Tangga Pengguna Listrik
$X_5$	Kemampuan Keuangan Daerah	Rasio Derajat Desentralisasi
$X_6$	Perekonomian	Persentase Penduduk Miskin
$X_7$	Masyarakat	Persentase Rata-Rata Pengeluaran Perkapita
$X_8$		Angka Harapan Hidup

Variabel	Kriteria	Keterangan
$X_9$	Sumber Daya	Indeks Pembangunan Manusia
$X_{10}$	Manusia	Tingkat Pengangguran Terbuka

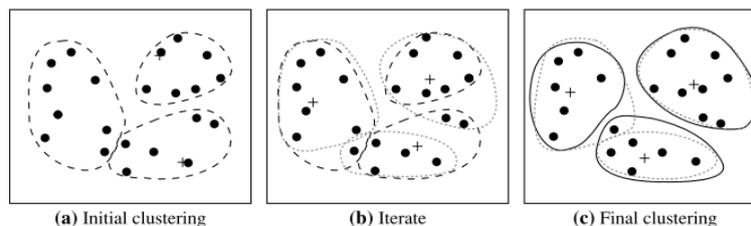
**Tabel 2. Variabel Kategorik**

Variabel	Kriteria	Keterangan	Kategori
$X_{11}$	Karakteristik Daerah	Keberadaan	1: Ada
		Kawasan Industri	2: Tidak Ada
$X_{12}$	Sarana Prasarana	Kekuatan Sinyal	1: Sangat Kuat
		Telepon Seluler	2: Kuat
$X_{13}$	Sumber Air Minum		3: Lemah
			4: Tidak Ada
			1: Air Kemasan
			2: Ledeng
			3: Sumur
		4: Mata Air	
		5: Danau/Sungai/Lainnya	

Data yang digunakan merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Papua. Data yang digunakan berupa publikasi hasil Potensi Desa (PODES) dan Survei Sosial dan Ekonomi Nasional (Susenas) periode 2018. Data diambil dari laman resmi BPS Provinsi Papua dan Kementerian Keuangan, yaitu <https://papua.bps.go.id/> dan <https://djpk.kemenkeu.go.id/>.

### ***K-means***

*K-means* merupakan salah satu metode *cluster analysis* non hierarki yang berfungsi untuk mempartisi objek yang ada kedalam satu atau lebih *cluster* atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama dikelompokkan dalam satu *cluster* yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokkan kedalam *cluster* yang lain. Metode *K-means* mendefinisikan *centroid* dari sebuah *cluster* sebagai nilai rata-rata dari titik-titik di dalam *cluster* tersebut. Prosedur *K-means* secara ringkas dapat dilihat pada Gambar 1 (Han et al., 2012).



Gambar 1. Prosedur *K-means*

Langkah pertama yang dilakukan yaitu menentukan banyak *cluster* yang digunakan. Kemudian, secara acak dipilih objek dengan banyak *k* cluster yang masing-masing awalnya mewakili nilai rata-rata atau *centroid cluster*. Objek kemudian dipindahkan ke *cluster* yang paling mirip berdasarkan nilai jarak *euclidean* antara objek dan nilai *centroid cluster*. Kemudian, menghitung nilai *centroid* baru menggunakan rata-rata objek yang berada pada masing-masing *cluster* pada iterasi sebelumnya. Semua objek kemudian dipindahkan

menggunakan cara yang sama dengan *centroid cluster* baru. Iterasi berlanjut sampai stabil, yaitu *cluster* yang terbentuk pada iterasi saat ini sama dengan yang terbentuk pada iterasi sebelumnya (Han et al., 2012:452).

Jarak *euclidian* merupakan salah satu cara dalam menghitung jarak 2 objek data. Untuk menghitung jarak *euclidian* antara objek  $i$  dan objek  $j$  digunakan sebagai berikut (Han et al., 2012):

$$d(i, j) = \sqrt{\sum_{l=1}^p (x_{il} - x_{jl})^2} \quad (1)$$

dengan  $d(i, j)$  jarak *euclidian* antara objek  $i$  dan objek  $j$  untuk  $i = 1, 2, \dots, n, j = 1, 2, \dots, n, l = 1, 2, \dots, p, x_{il}$  nilai objek ke  $i$  pada variabel ke  $l, x_{jl}$  nilai objek ke  $j$  pada variabel ke  $l, p$  banyak variabel.

### **K-modes**

Metode *K-modes* merupakan metode klusterisasi untuk data bertipe kategorik. Metode *K-modes* termasuk metode *hard clustering*, yang berarti bahwa setiap objek termasuk dalam *cluster* yang unik. Namun, jika distribusi populasi merupakan campuran dari distribusi, satu objek dapat menjadi milik lebih dari satu *cluster* (Maciel et al., 2017). Metode ini pertama kali diperkenalkan oleh Huang (1998) yang merupakan pengembangan dan pemodifikasian metode *K-means* untuk dapat digunakan pada data bertipe kategorik. Modifikasi yang dilakukan menghasilkan metode *K-modes* yang mengubah fungsi jarak *euclidean* dengan jarak *simple matching*. Selain itu, untuk menentukan *centroid cluster* digunakan nilai modus.

Langkah pengelompokkan menggunakan *K-modes* adalah sebagai berikut (Huang, 1998):

1. Menentukan banyak *cluster* dan *centroid* awal untuk masing-masing *cluster*.
2. Memproses objek dengan menempatkan pada *centroid* terdekat. Jarak *centroid* terdekat dihitung dengan menggunakan jarak *simple matching* dengan persamaan berikut:

$$d(i, j) = \sum_{l=1}^p \delta(x_{il}, x_{jl}) \quad (2)$$

dengan:

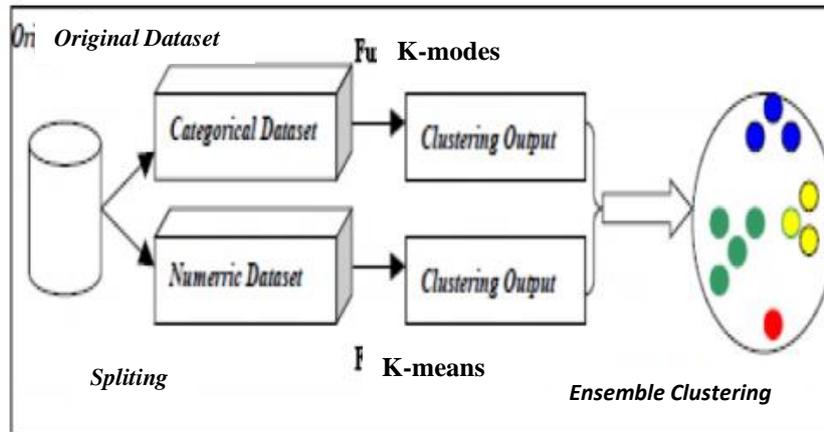
$$\delta(i, j) = \begin{cases} 0, & x_{il} = x_{jl} \\ 1, & x_{il} \neq x_{jl} \end{cases}$$

3. Setelah semua objek telah dialokasikan ke *cluster*, uji ulang ketidaksamaan objek terhadap *centroid* saat ini. Memperbarui pusat *cluster* dari masing-masing kelompok dengan nilai yang sering muncul atau modus di tiap kelompok tersebut.
4. Ulangi langkah (3) hingga tidak ada objek yang berpindah kelompok atau tidak ada objek yang berubah.

### **Ensemble Clustering**

*Ensemble clustering* adalah metode yang digunakan untuk menggabungkan atau menggabungkan sekumpulan hasil *clustering* yang berbeda. Metode ini pertama kali diperkenalkan oleh Strehl & Ghosh (2002). Gagasan ini muncul sebagai pendekatan alternatif untuk meningkatkan kualitas hasil ketahanan solusi kelompok dari algoritma *clustering* (Vega-Pons & Ruiz-Shulcloper, 2011). *Ensemble clustering* dapat digunakan untuk menggabungkan

dua atau lebih metode *clustering* yang sebelumnya telah dilakukan. Pada penelitian ini *ensemble clustering* menggabungkan hasil *K-means* dan *K-modes* yang dilakukan sebelumnya. Hasil *clustering* dari *K-means* dan *K-modes* menjadi input pada proses *ensemble clustering*, sehingga terdapat dua input bertipe kategori. Oleh karena itu proses *ensemble clustering* dilakukan dengan *K-modes*. Pada Proses *ensemble clustering* digambarkan seperti Gambar 2.



Gambar 2 Proses *Ensemble Clustering*

### Validasi Cluster

*Cluster* yang sudah terbentuk perlu dilakukan validasi, untuk dapat diketahui data tersebut layak dibagi kedalam berapa *cluster*. Terdapat dua kategori validasi *cluster* yaitu validasi internal dan validasi eksternal. Validasi eksternal merupakan validasi yang mengevaluasi ke *cluster* berdasarkan label kelas yang diberikan. Validasi eksternal menggunakan informasi yang tidak ada dalam data, sedangkan, validasi internal hanya menggunakan informasi yang ada dalam data (Liu et al., 2010). Beberapa metode validasi *cluster* yang dapat digunakan yaitu *Davies Bouldin Index* (DBI) dan *Silhouette*.

### *Davies Bouldin Index*

*Davies Bouldin Index* atau DBI pertama kali diperkenalkan oleh Davies & Bouldin (1979). DBI berfungsi untuk menentukan banyak *cluster* terbaik pada pengelompokan yang dilakukan. Pengukuran pada DBI bertujuan memaksimalkan jarak *intercluster* antara satu *cluster* dengan *cluster* yang lain. Pada dasarnya, untuk mengevaluasi perbedaan antara *cluster* ke-*k* dan ke-*k'*, yang harus bernilai sebesar mungkin untuk antar *cluster* dan sekecil mungkin untuk di dalam *cluster*. DBI memiliki korelasi positif untuk kasus dalam *cluster* dan korelasi negatif untuk kasus antar *cluster* (Xiao, 2017). Perhitungan DBI dilakukan dengan persamaan berikut:

$$DBI = \frac{1}{K} \sum_{k=1}^K \max_{k \neq k'} R_{k,k'} \tag{3}$$

dengan

$$R_{k,k'} = \frac{S_k + S_{k'}}{B_{k,k'}}$$

untuk  $k = 1, 2, \dots, K$  dan  $k' = 1, 2, \dots, K$ ,  $K$  banyak *cluster*,  $S_k$  rata rata jarak objek ke *centroid* pada *cluster*  $k$ ,  $B_{k,k'}$  jarak antar *centroid cluster* ke  $k$  dan *centroid cluster* ke  $k'$  dengan  $k \neq k'$

Semakin kecil nilai DBI yang diperoleh dan bersifat positif, maka *cluster* yang diperoleh semakin bagus. Dalam proses perhitungan nilai DBI, nilai tidak mungkin bernilai negatif karena ketika menghitung jarak objek terhadap *centroid* nilai yang diperoleh akan bersifat positif. Sehingga, untuk tahapan selanjutnya akan selalu bernilai positif.

#### *Silhouette*

Metode *silhouette* pertama kali diperkenalkan oleh Rousseeuw (1987) yang didasarkan pada kohesi dan separasi. Metode ini menunjukkan objek mana yang terletak dengan baik di dalam *cluster* mereka dan mana yang hanya berada di antara *cluster*. Untuk menghitung nilai *Silhouette* dapat digunakan persamaan berikut (Rousseeuw, 1987):

$$SI_{ik} = \frac{b_{ik} - a_{ik}}{\max\{a_{ik}, b_{ik}\}} \quad (4)$$

dengan  $SI_{ik}$  nilai *silhouette* objek ke- $i$  dalam satu *cluster* ke- $k$ ,  $a_{ik}$  rata-rata jarak objek ke- $i$  terhadap semua objek dalam satu *cluster* untuk  $i = 1, 2, \dots, n_k$ ,  $b_{ik}$  rata-rata jarak objek ke- $i$  terhadap semua objek yang tidak dalam satu *cluster* dengan objek.

#### Prosedur Analisis Daerah Tertinggal

Metode *ensemble clustering* adalah metode yang digunakan untuk menggabungkan hasil *cluster* dari proses *clustering* sebelumnya. Langkah pertama yang dilakukan yaitu mendeskripsikan karakteristik kabupaten/kota berdasarkan indikator daerah tertinggal. Kemudian memisahkan data berdasarkan jenis datanya, yaitu data berskala numerik dan data berskala kategorik. Selanjutnya, untuk data berskala numerik dilakukan dengan metode *K-means* yang menghasilkan *output* 1 dan data kategorik menggunakan *K-modes* menghasilkan *output* 2. Kedua *output* akan menghasilkan data kategorik. Langkah selanjutnya, dilakukan validasi *cluster* untuk menentukan *cluster* terbaik. Masing-masing *output* dari hasil *K-means* dan *K-modes* dijadikan sebagai data *input* pada proses *ensemble clustering*, yaitu *clustering* dengan metode *K-modes*.

Hasil clustering disajikan dalam bentuk peta dengan tingkat satus daerah tertinggal. Analisis daerah tertinggal dilakukan untuk mendapatkan tingkat status daerah tertinggal dan karakteristik dari daerah-daerah yang telah dikelompokkan dalam beberapa *cluster*. Tingkat status yang diperoleh dapat digunakan untuk mengetahui mana kelompok atau *cluster* yang paling jauh dari ketertinggalan hingga kelompok atau *cluster* yang paling parah tertinggal. Berikut prosedur analisis daerah tertinggal:

1. Untuk data berskala numerik dilakukan perhitungan rata-rata di setiap variabel pada masing-masing *cluster*. Untuk data berskala kategorik, digunakan nilai modus pada setiap *cluster*.
2. Kemudian hasil rata-rata tersebut diberi peringkat sesuai dengan banyak  $k$  *cluster*. Nilai rata-rata tertinggi akan diberi peringkat yang paling tinggi yaitu dimulai dari peringkat 1, peringkat 2, sampai peringkat  $k$ . Namun terdapat variabel yang pemberian peringkat dilakukan secara terbalik karena variabel tersebut bernilai negatif. Terdapat 3 variabel yaitu  $X_6$  (persentase penduduk miskin),  $X_7$  (rata-rata pengeluaran perkapita), dan  $X_{10}$  (Tingkat Pengangguran Terbuka).
3. Untuk variabel bertipe kategorik, pemberian peringkat berbeda disetiap variabelnya menyesuaikan banyak variabel yang ada.

**Tabel 3. Peringkat Data Kategorik**

Variabel	Keterangan	Peringkat
$X_{11}$	Keberadaan	1: Tidak Ada
	Kawasan Industri	2: Ada
$X_{12}$	Kekuatan Sinyal Telepon Seluler	1: Tidak Ada Sinyal
		2: Lemah
		3: Kuat
		4: Sangat Kuat
$X_{13}$	Sumber Air Minum	1: Danau/Sungai/Lainnya
		2: Mata Air
		3: Sumur
		4: Ledeng
		5: Air Kemasan

- Dilakukan penjumlahkan peringkat pada data numerik dan data kategorik sehingga diperoleh nilai total peringkat dari masing-masing *cluster* sehingga diperoleh tingkatan status daerah tertinggal. Tingkat pertama akan diberikan kepada *cluster* yang memiliki nilai total peringkat yang paling besar dan tingkat terakhir akan diberikan kepada *cluster* yang memiliki nilai total peringkat paling kecil. Semakin kecil nilai total semakin tertinggal daerah yang berada di *cluster* tersebut.

## HASIL DAN PEMBAHASAN

### Hasil

#### *Pembentukan Cluster Data Numerik dengan K-means*

Pembentukan *cluster* data berskala numerik dilakukan dengan menggunakan metode *K-means*. Dalam penelitian ini, terdapat 29 objek yang digunakan dan merupakan kabupaten/kota di Provinsi Papua. Terdapat 10 variabel yang digunakan, antara lain yaitu 1) rasio desa dengan permukaan jalan terluas, rasio rumah sakit terhadap jumlah penduduk, rasio SMA/MA/Sederajat terhadap penduduk usia SMA, persentase rumah tangga pengguna listrik, rasio derajat desentralisasi, persentase penduduk miskin, persentase pengeluaran penduduk perkapita, Angka Harapan Hidup, Indeks Pembangunan Manusia, dan Tingkat Pengangguran Terbuka. Untuk mendapat banyak *cluster* terbaik, digunakan metode DBI dan *Silhouette*. Dengan bantuan *software* Rstudio diperoleh hasil seperti pada Tabel 4.

**Tabel 4. Validasi Cluster Data Numerik**

Banyak <i>cluster</i>	DBI	<i>Silhouette</i>
2	0,9823	0,4435
3	1,2522	0,3294
4	1,1553	0,2954
5	1,3593	0,2024

Berdasarkan Tabel 4 diperoleh hasil bahwa 2 merupakan banyak *cluster* terbaik. Hal itu dikarenakan validasi 2 *cluster* memiliki nilai DBI terkecil dan nilai *Silhouette* terbesar.

Tabel 5 merupakan hasil pemetaan data numerik dengan menggunakan metode *K-means*. Hasil 2 *cluster* diperoleh dengan bantuan *software* Rstudio. Dari 29 kabupaten/kota di Provinsi Papua, 19 diantaranya masuk dalam *cluster* 1. Sedangkan, 10 lainnya masuk ke dalam

cluster 2.

**Tabel 5. Hasil Pemetaan Data Numerik**

Cluster	Kabupaten/Kota
Cluster 1	Kabupaten Jayawijaya, Kabupaten Paniai, Kabupaten Puncak Jaya, Kabupaten Mappi, Kabupaten Asmat, Kabupaten Yahukimo, Kabupaten Pegunungan Bintang, Kabupaten Tolikara, Kabupaten Waropen, Kabupaten Supiori, Kabupaten Mamberamo Raya, Kabupaten Dogiyai, Kabupaten Deiyai, Kabupaten Pegunungan Bintang, Kabupaten Tolikara, Kabupaten Nduga, Kabupaten Lanny Jaya, Kabupaten Mamberamo Tengah, Kabupaten Yalimo, Kabupaten Puncak, Kabupaten Intan Jaya.
Cluster 2	Kabupaten Merauke, Kabupaten Jayapura, Kabupaten Nabire, Kabupaten Kepulauan Yapen, Kabupaten Biak Numfor, Kabupaten Mimika, Kabupaten Boven Digoel, Kabupaten Sarmi, Kabupaten Keerom, Kota Jayapura.

*Pembentukan Cluster Data Numerik dengan K-modes*

Pembentukan *cluster* data berskala kategorik akan dikelompokkan menggunakan metode *K-modes*. Pada data bertipe kategorik terdapat 3 variabel yang digunakan yaitu keberadaan kawasan industri, kekuatan sinyal telepon seluler, dan sumber air minum. Sama seperti sebelumnya, dilakukan penentuan banyak *cluster* terbaik dengan metode DBI dan *Silhouette*.

**Tabel 6. Validasi Cluster Data Kategorik**

Banyak cluster	DBI	Silhouette
2	1,4634	0,4316
3	1,6277	0,4365
4	0,7429	0,6257
5	0,4958	0,6208

Berdasarkan **Tabel 6**, nilai minimum DBI terdapat pada 5 *cluster*, sedangkan nilai maksimal *Silhouette* terdapat pada 4 *cluster*. Terdapat perbedaan jumlah *cluster* terbaik diantara kedua metode validasi. Oleh karena itu, akan dilakukan dua perhitungan dengan jumlah *cluster* berbeda yaitu hasil 4 *cluster* dan 5 *cluster* dan diperoleh hasil seperti pada **Tabel 7**. Nantinya, pada hasil analisis data campuran akan dilakukan kembali validasi jumlah *cluster* terbaik diantara kedua hasil pengelompokkan data kategorik.

**Tabel 7. Hasil Pemetaan Data Kategorik**

No	Kabupaten/Kota	4 cluster	5 cluster	No	Kabupaten/Kota	4 cluster	5 cluster
1	Merauke	2	2	16	Sarmi	3	3
2	Jayawijaya	4	4	17	Keerom	2	2
3	Jayapura	3	3	18	Waropen	2	2
4	Nabire	2	2	19	Supiori	2	2
5	Kepulauan Yapen	4	4	20	Mamberamo Raya	2	2
6	Biak Numfor	3	3	21	Nduga	4	4

No	Kabupaten/Kota	4 cluster	5 cluster	No	Kabupaten/Kota	4 cluster	5 cluster
7	Paniai	4	4	22	Lanny Jaya	4	4
8	Puncak Jaya	4	4	23	Mamberamo Tengah	1	5
9	Mimika	2	2	24	Yalimo	4	4
10	Boven Digoel	2	2	25	Puncak	2	2
11	Mappi	2	2	26	Dogiyai	4	4
12	Asmat	2	2	27	Intan Jaya	1	1
13	Yahukimo	4	4	28	Deiyai	2	2
14	Pegunungan Bintang	2	2	29	Kota Jayapura	2	2
15	Tolikara	1	1				

*Pembentukan Cluster Data Numerik dengan K-modes*

Pembentukan hasil *cluster* akhir menggunakan metode *ensemble clustering*. Metode tersebut menggabungkan hasil dari metode *K-means* dan *K-modes* sehingga memperoleh hasil *cluster* akhir. Hasil dari pengelompokan data numerik dituliskan sebagai variabel baru yaitu  $X_1$ , sedangkan hasil pengelompokan data kategorik juga dianggap sebagai variabel baru yaitu  $X_2$ . Data dari perhitungan kedua metode dapat dilihat pada **Error! Reference source not found.** Karena hasil *cluster* bertipe kategorik, maka penggabungan menggunakan metode *K-modes* yang berfungsi untuk *clustering* data bertipe kategorik. Dilakukan tahapan validasi dengan menggunakan DBI dan *Silhouette* sama seperti ketika melakukan validasi pada pengelompokkan data kategorik dengan *K-modes*.

**Tabel 8. Validasi Cluster Data Campuran**

Banyak <i>cluster</i>	DBI		Silhouette	
	4 cluster	5 cluster	4 cluster	5 cluster
2	0,8466	0,8214	0,5652	0,5706
3	0,4079	0,4209	0,8161	0,8152
4	0,3232	0,3385	0,8101	0,8390
5	0,1844	0,2452	0,9240	0,8524

Dapat dilihat pada Tabel 8, berdasarkan nilai DBI dan *silhouette* diperoleh *cluster* terbaik pada 5 *cluster* dengan penggabungan 2 *cluster* data numerik dan 4 *cluster* data kategorik. Dari hasil tersebut diperoleh hasil akhir seperti pada Tabel 9.

**Tabel 9. Hasil Pemetaan Data Numerik**

<i>Cluster</i>	Kabupaten/Kota
<i>Cluster 1</i>	Kabupaten Merauke, Kabupaten Nabire, Kabupaten Kepulauan Yapen, Kabupaten Mimika, Kabupaten Boven Digoel, Kabupaten Mappi, Kabupaten Asmat, Kabupaten Keerom, Kabupaten Waropen, Kabupaten Supiori, Kabupaten Mamberamo Raya, Kota Jayapura.
<i>Cluster 2</i>	Kabupaten Jayapura, Kabupaten Biak Numfor, Kabupaten Sarmi.

Cluster	Kabupaten/Kota
Cluster 3	Kabupaten Puncak Jaya, Kabupaten Yahukimo, Kabupaten Tolikara, Kabupaten Nduga, Kabupaten Lanny Jaya, Kabupaten Yalimo, Kabupaten Mamberamo Tengah, Kabupaten Intan Jaya.
Cluster 4	Kabupaten Jayawijaya, Kabupaten Paniai, Kabupaten Dogiyai.
Cluster 5	Kabupaten Pegunungan Bintang, Kabupaten Puncak, Kabupaten Deiyai.

Dari 29 kabupaten/kota di Provinsi Papua, 12 diantaranya masuk dalam *cluster* 1. *Cluster* 2 terdiri dari 3 kabupaten/kota. *Cluster* 3 terdiri dari 8 kabupaten/kota. *Cluster* 4 dan 5 terdiri dari 3 kabupaten/kota.

*Pemetaan dan Analisis Hasil Clustering*

Pembentukan *cluster* dengan *ensemble clustering* menghasilkan banyak *cluster* terbaik yaitu 5 *cluster*. Jumlah tersebut sudah melalui tahap validasi dengan menggunakan metode DBI dan *Silhouette*. Dari 29 kabupaten/kota di Provinsi Papua terdapat 11 kabupaten yang berada di *cluster* 1. Sedangkan untuk jumlah anggota *cluster* lainnya berturut-turut yaitu 3, 8, 3, dan 3. Tingkat daerah tertinggal dapat diukur dari nilai rata-rata untuk data bertipe numerik dan modus untuk data bertipe kategorik dari masing-masing *cluster*. *Cluster* yang memiliki nilai rata-rata *cluster* yang rendah termasuk dalam daerah yang sangat tertinggal dan akan diberi peringkat 1 untuk penentuan status tingkat.

**Tabel 10. Nilai Rata-Rata Masing-Masing Cluster**

Variabel	Cluster				
	1	2	3	4	5
$X_1$	0.3339	0.6209	0.0376	0.1873	0.1232
$X_2$	0.0010	0.0009	0.0004	0.0003	0.0005
$X_3$	0.0023	0.0027	0.0006	0.0007	0.0006
$X_4$	0.1614	0.2787	0.0033	0.0348	0.0000
$X_5$	4.694	3.150	1.161	2.982	0.431
$X_6$	23.25	17.89	37.75	35.50	37.46
$X_7$	54.57	53.61	64.77	56.71	64.87
$X_8$	65.10	66.89	63.70	63.42	64.75
$X_9$	64.03	68.74	45.20	55.70	45.19
$X_{10}$	5.56	9.58	1.09	0.57	1.54

**Tabel 11. Nilai Modus dari Masing-Masing Cluster**

Variabel	Cluster				
	1	2	3	4	5
$X_{11}$	1	1	1	1	1
$X_{12}$	4	3	4	4	4
$X_{13}$	5	3	4	4	5

Tabel 10. merupakan hasil dari perhitungan rata-rata *cluster* dari variabel-variabel

bertipe numerik. Sedangkan Tabel 11. merupakan tabel yang menunjukkan nilai modus pada *cluster* di setiap variabel. Langkah selanjutnya adalah memberi peringkat seperti yang sudah ditentukan pada prosedur analisis daerah tertinggal.

**Tabel 12. Peringkat Cluster Data Numerik**

Variabel	Cluster				
	1	2	3	4	5
$X_1$	4	5	1	3	2
$X_2$	5	4	2	1	3
$X_3$	4	5	1	3	2
$X_4$	4	5	2	3	1
$X_5$	5	4	2	3	1
$X_6$	4	5	1	3	2
$X_7$	4	5	1	3	2
$X_8$	4	5	2	1	3
$X_9$	4	5	2	3	1
$X_{10}$	2	1	4	5	3
Total	<b>40</b>	<b>44</b>	<b>18</b>	<b>28</b>	<b>20</b>

Pada Tabel 12 dapat dilihat bahwa *cluster* 3 rata-rata mendapatkan peringkat yang tinggi dan total dari peringkat dari *cluster* 3 bernilai 18. Nilai tersebut dapat diartikan bahwa *cluster* 3 merupakan *cluster* dengan daerah yang sangat tertinggal. Kemudian untuk nilai total tertinggi terdapat pada *cluster* 2 dengan nilai total yaitu 44.

**Tabel 13 Peringkat Cluster Data Kategorik**

Variabel	Cluster				
	1	2	3	4	5
$X_{11}$	2	2	2	2	2
$X_{12}$	1	2	1	1	1
$X_{13}$	1	3	2	2	1
Total	<b>4</b>	<b>7</b>	<b>5</b>	<b>5</b>	<b>4</b>

Pada Tabel 13Error! Reference source not found. dapat dilihat bahwa *cluster* 1 dan 5 memiliki nilai total yang rendah. Artinya, *cluster* tersebut memiliki daerah yang sangat tertinggal dibandingkan dengan daerah lainnya pada variabel-variabel yang bertipe kategorik. Nilai total paling tinggi dimiliki *cluster* 2 dengan nilai total yaitu 7. Selanjutnya dilakukan penggabungan antara total peringkat dari data numerik dan data kategorik.

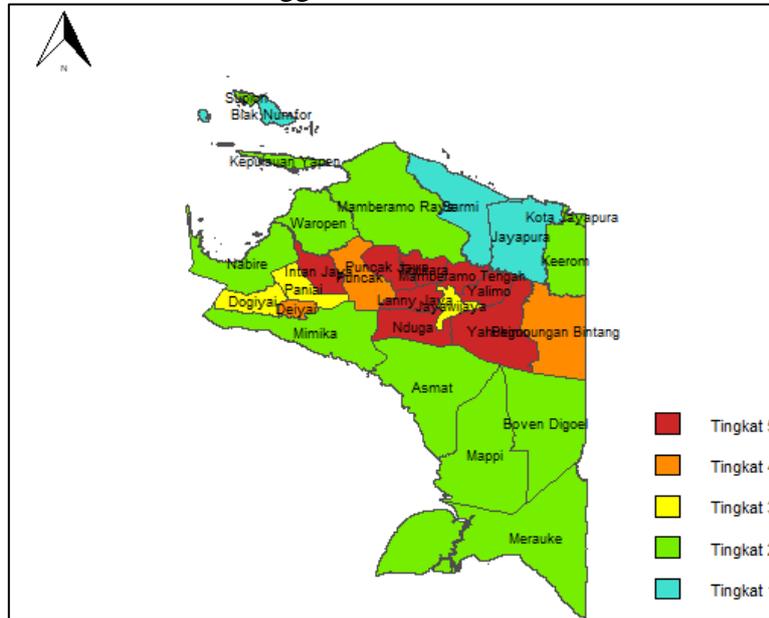
**Tabel 14. Hasil Gabungan Total Peringkat**

Tipe Data	Cluster				
	1	2	3	4	5
Numerik	40	44	18	28	20
Kategorik	4	7	5	5	4
Total	<b>44</b>	<b>51</b>	<b>23</b>	<b>33</b>	<b>24</b>

Berdasarkan hasil gabungan peringkat yang dapat dilihat pada Tabel 14, diperoleh tingkatan status daerah tertinggal yaitu tingkat 1 untuk *cluster* 2, tingkat 2 untuk *cluster* 1,

tingkat 3 untuk *cluster* 4, tingkat 4 untuk *cluster* 5, dan tingkat 5 untuk *cluster* 3. Tingkat status menunjukkan bahwa semakin rendah tingkatan status maka daerah yang berada di tingkat tersebut semakin jauh dari ketertinggalan.

Setelah diperoleh tingkatan status ketertinggalan daerah, dibuat peta dengan setiap tingkat diberikan warna yang berbeda. Berikut **Error! Reference source not found.** merupakan hasil pemetaan daerah tertinggal.



Gambar 3 Peta Daerah Tertinggal di Provinsi Papua

*Analisis Karakteristik Cluster*

Analisis karakteristik dilakukan untuk mengetahui karakteristik disetiap *cluster*nya berdasarkan indikator daerah tertinggal. Analisis dilakukan dengan melihat hasil perhitungan rata-rata disetiap variabel pada masing-masing *cluster* untuk data bertipe numerik. Untuk data bertipe kategorik, karakteristik dari *cluster* dilihat dari nilai modus di setiap variabel.

**Tabel 15. Karakteristik Ketertinggalan Setiap Cluster**

Cluster	Karakteristik Ketertinggalan	Status Ketertinggalan
2	Tingkat pengangguran terbuka tertinggi	Tingkat 1
1	Sebagian daerah menggunakan sumber air minum dari sungai, danau, dll	Tingkat 2
4	Rasio rumah sakit terhadap jumlah penduduk terendah, Angka Harapan Hidup terendah	Tingkat 3
5	Persentase rumah tangga pengguna listrik terendah, rasio derajat desentralisasi terendah, mayoritas daerah menggunakan sumber air minum dari sungai, danau, dll	Tingkat 4
3	Rasio desa dengan permukaan jalan terluas aspal terendah, rasio SMA/MA/ sederajat terhadap penduduk usia SMA sederajat terendah, persentase penduduk miskin	Tingkat 5

<i>Cluster</i>	Karakteristik Ketertinggalan	Status Ketertinggalan
	tertinggi, persentase rata-rata pengeluaran perkapita tertinggi, Indeks Pembangunan Manusia terendah, terdapat daerah yang tidak memiliki kawasan industri	

*Cluster 2* merupakan daerah tingkat 1 yang merupakan *cluster* dengan variabel mayoritas bernilai tinggi. Pada tingkat 1 hampir semua memiliki nilai yang tinggi kecuali tingkat pengangguran terbuka. Menurut perhitungan yang dilakukan sebelumnya, tingkat 1 merupakan *cluster* yang paling tidak tertinggal dibandingkan dengan daerah di *cluster* lainnya.

*Cluster 1* merupakan daerah tertinggal tingkat 2 dengan sebagian daerah mayoritas menggunakan sumber air minum dari sungai, danau, dll. Namun, sebagian daerah lainnya juga menggunakan sumur sebagai sumber air minumannya. Pada tingkat ini, nilai rasio rumah sakit terhadap jumlah penduduk yang lebih tinggi dibandingkan dengan tingkat lainnya, yaitu terdapat 1 rumah sakit yang dapat menampung sekitar 1000 penduduk. Selain itu, nilai rasio derajat desentralisasi pada tingkat 2 merupakan nilai tertinggi dibandingkan tingkat lainnya.

*Cluster 4* merupakan daerah tertinggal tingkat 3 dengan nilai rasio rumah sakit terhadap jumlah penduduk terendah dan Angka Harapan Hidup terendah di Provinsi Papua. Nilai rata-rata dari rasio rumah sakit terhadap jumlah penduduk yaitu bernilai 0,00026 atau terdapat 2 rumah sakit yang dapat menampung 10.000 penduduk. Nilai rata-rata Angka Harapan Hidup di tingkat 3 merupakan nilai yang paling rendah atau dapat dikatakan bahwa rata-rata perkiraan banyak tahun yang dapat ditempuh oleh penduduk di daerah pada tingkat 3 bernilai rendah dibandingkan dengan daerah di tingkat lainnya.

*Cluster 5* merupakan daerah tertinggal tingkat 4 dengan persentase rumah tangga pengguna listrik terendah, rasio derajat desentralisasi terendah, dan mayoritas daerah di tingkat 4 menggunakan sungai, danau, dll untuk sumber air minumannya. Persentase rumah tangga pengguna listrik terendah berarti bahwa banyak keluarga di daerah tingkat 4 yang belum menggunakan listrik. Selain itu, kemampuan keuangan daerah di daerah tingkat 4 termasuk yang buruk karena memiliki nilai rasio derajat desentralisasi terendah.

*Cluster 3* menempati tingkat 5 dalam status ketertinggalan daerah di Provinsi Papua. Daerah-daerah di tingkat 5 memiliki nilai indikator daerah tertinggal yang lebih rendah dibandingkan tingkat lainnya. Tingkat 5 hampir memiliki nilai yang rendah di semua variabel. Diantaranya yaitu nilai rasio desa dengan permukaan jalan terluas aspal terendah, rasio SMA/MA/ sederajat terhadap penduduk usia sma sederajat terendah, persentase penduduk miskin tertinggi, persentase rata-rata pengeluaran perkapita tertinggi, Indeks Pembangunan Manusia terendah, dan pada tingkat tersebut terdapat daerah yang tidak memiliki kawasan industri.

## Pembahasan

Pada penelitian ini bertujuan untuk memetakan kabupaten/kota di Provinsi Papua berdasarkan indikator daerah tertinggal dengan metode *ensemble clustering*. Selain itu, dari hasil pemetaan dapat diketahui karakteristik dari setiap *cluster* yang telah dikelompokkan. Data yang digunakan yaitu data dari variabel-variabel yang mewakili indikator daerah tertinggal. Terdapat 13 variabel yang digunakan, 10 variabel merupakan data bertipe numerik dan 3 variabel merupakan data bertipe kategorik.

Hasil menunjukkan bahwa terbentuk 5 *cluster* sebagai banyak *cluster* terbaik. Dari 5 *cluster* yang terbentuk dipetakan menjadi 5 tingkat daerah tertinggal. Semakin tinggi tingkat

semakin tertinggal parah daerah yang berada di tingkat tersebut. Tingkat 1 beranggotakan 3 kabupaten/kota. Tingkat 2 beranggotakan 12 kabupaten/kota. Tingkat 3 beranggotakan 3 kabupaten/kota. Tingkat 4 beranggotakan 3 kabupaten/kota. Tingkat 5 beranggotakan 8 kabupaten/kota. Setiap tingkatan status daerah tertinggal memiliki karakteristik yang berbeda-beda. Tingkat 1 merupakan tingkat dengan daerah yang mayoritas daerahnya memiliki karakteristik jauh dari indikator-indikator daerah tertinggal. Atau, daerah yang berada di tingkat 1 ini adalah daerah yang jauh dari ketertinggalan daerah. Namun, tingkat 1 ini memiliki tingkat pengangguran yang tinggi. Tingkat 2 memiliki karakteristik dengan ketertinggalan yaitu pada sumber air minumannya. Sebagian daerah yang ada di tingkat 2 mengkonsumsi air minum bersumber dari sungai, danau, dll. Tingkat 3 merupakan tingkat dengan daerah yang memiliki rasio rumah sakit terhadap jumlah penduduk yang rendah dan Angka Harapan Hidup yang rendah. Tingkat 4 merupakan tingkat dengan daerah yang persentase rumah tangga pengguna listrik terendah, rasio derajat desentralisasi terendah, dan juga mayoritas daerah menggunakan sumber air minum dari sungai, danau, dll. Tingkat 5 merupakan daerah tertinggal sangat parah dibandingkan dengan daerah yang ada di tingkat lain dengan rasio desa dengan permukaan jalan terluas aspal terendah, rasio SMA/MA/ sederajat terhadap penduduk usia SMA sederajat terendah, persentase penduduk miskin tertinggi, persentase rata-rata pengeluaran perkapita tertinggi, Indeks Pembangunan Manusia terendah, juga pada tingkat ini terdapat daerah yang tidak memiliki kawasan industri.

## SIMPULAN

Penelitian ini mengaplikasikan metode *ensemble clustering* untuk pemetaan daerah tertinggal di Provinsi Papua. Variabel yang digunakan yaitu variabel-variabel yang bertipe numerik dan kategorik. Pembentukan *cluster* dilakukan dengan menggunakan metode *K-means* untuk variabel bertipe numerik dan *K-modes* untuk variabel bertipe kategorik. Pembentukan *cluster* diawali dengan menentukan banyak *cluster* terbaik dengan menggunakan Davies Index Bouldin (DBI) dan Silhouette. Hasil dari proses pembentukan *cluster* menggunakan *ensemble clustering* yaitu terbentuknya 5 *cluster*. Dari hasil tersebut, dapat disimpulkan bahwa metode *ensemble* dapat digunakan untuk pemetaan daerah tertinggal.

## DAFTAR PUSTAKA

- Badan Perencanaan Pembangunan Daerah Provinsi Papua. (2020). *Analisis Kerangka Pembangunan Daerah Provinsi Papua Tahun 2020*. Badan Perencanaan Pembangunan Daerah Provinsi Papua. <https://bappeda.papua.go.id/file/456181782.pdf>
- Budianta, A. (2010). Pengembangan wilayah perbatasan sebagai upaya pemerataan pembangunan wilayah di Indonesia. *Jurnal SMARTek*, 8(1), 72–82. <http://jurnal.untad.ac.id/jurnal/index.php/SMARTEK/article/view/628/546>
- Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (Third). Elsevier. <https://www.sciencedirect.com/book/9780123814791/data-mining-concepts-and-techniques>
- Hasugian, P. M., Hutahaean, H. De., Sinaga, B., Sriadhi, & Silaban, S. (2020). Villages Status Classification Analysis Involving K- Means Algorithm To Support Kementerian Desa Pembangunan Daerah Tertinggal dan Villages Status Classification Analysis Involving

- K-Means Algorithm To Support Kementerian Desa Pembangunan Daerah Tertinggal. *Journal of Physics: Conference Series*. <https://doi.org/10.1088/1742-6596/1641/1/012058>
- Huang, Z. (1998). Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. *Data Mining and Knowledge Discovery* 2, 283–304. <https://doi.org/10.3923/ajbmb.2011.284.290>
- Hunt, L., & Jorgensen, M. (2011). Clustering mixed data. *WIREs Data Mining Knowl Discov*, 1(August), 352–361. <https://doi.org/10.1002/widm.33>
- Liu, Y., Li, Z., Xiong, H., Gao, X., & Wu, J. (2010). Understanding of internal clustering validation measures. *Proceedings - IEEE International Conference on Data Mining, ICDM*, 911–916. <https://doi.org/10.1109/ICDM.2010.35>
- Maciel, D. B. M., Amaral, G. J. A., de Souza, R. M. C. R., & Pimentel, B. A. (2017). Multivariate fuzzy k-modes algorithm. *Pattern Analysis and Applications*, 20(1), 59–71. <https://doi.org/10.1007/s10044-015-0465-3>
- Peraturan Pemerintah RI. (2014). *Peraturan Pemerintah RI tentang Percepatan Pembangunan Daerah Tertinggal (Nomor 78 Tahun 2014 Pasal 1 Ayat 3)*.
- Peraturan Pemerintah RI. (2020). *Peraturan Pemerintah RI tentang Penetapan Daerah Tertinggal Tahun 2020-2024 (No 63 Tahun 2020 Pasal 2 Ayat 1)*.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(C), 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Strehl, A., & Ghosh, J. (2002). Cluster ensembles - A knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3(3), 583–617. <https://doi.org/10.1162/153244303321897735>
- Vega-Pons, S., & Ruiz-Shulcloper, J. (2011). A survey of clustering ensemble algorithms. *International Journal of Pattern Recognition and Artificial Intelligence*, 25(3), 337–372. <https://doi.org/10.1142/S0218001411008683>
- Dihni, V. A., & Annissa Mutia, A. (2021, October 15). *Papua Jadi Provinsi yang Miliki Desa Sangat Tertinggal Terbanyak pada 2021 | Databoks*. Databoks. <https://databoks.katadata.co.id/datapublish/2021/10/15/papua-jadi-provinsi-yang-miliki-desa-sangat-tertinggal-terbanyak-pada-2021>
- Xiao, J., Lu, J., & Li, X. (2017). *Davies Bouldin Index based hierarchical initialization K-means*. 21, 1327–1338. <https://doi.org/10.3233/IDA-163129>
- Yusfar, A. A., Tiro, M. A., & Sudarmin, S. (2020). Analisis Cluster Ensemble dalam Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Indikator Kinerja Pembangunan Ekonomi Daerah. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 3(1), 31. <https://doi.org/10.35580/variansiunm14626>