

**PENERAPAN METODE *AGGLOMERATIVE NESTING*  
(AGNES) PADA PENGELOMPOKAN WILAYAH DESA  
ATAU KELURAHAN DI KABUPATEN KUTAI  
KARTANEGARA  
(Studi Kasus: Data Potensi Desa (PODES) Tahun 2018)**

**Muhammad Hajar<sup>1\*</sup>, Yuki Novia Nasution<sup>2</sup>, Surya Prangga<sup>3</sup>**

<sup>1,2,3</sup>Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas  
Mulawarman, Indonesia

*Corresponding author:* muhammadhajat@gmail.com

**Abstrak.** Kabupaten Kutai Kartanegara (Kukar) merupakan salah satu penyumbang penduduk miskin terbesar di Provinsi Kalimantan Timur (Kaltim) pada Tahun 2017, sehingga diperlukan strategi yang tepat untuk mengatasi permasalahan kemiskinan di daerah ini. Salah satu strategi pembangunan ini dilakukan melalui proses identifikasi untuk memaksimalkan pemberantasan kemiskinan. Identifikasi dilakukan berdasarkan sarana, prasarana, akses, sosial, penduduk serta ekonomi yang disediakan dalam data Potensi Desa. Dalam melakukan identifikasi ini, digunakan Analisis *Cluster*. Analisis *Cluster* yang digunakan pada penelitian kali ini adalah analisis *cluster* untuk data numerik yaitu algoritma *Agglomerative Nesting* (AGNES). Metode *clustering* terbaik dan jumlah *cluster* optimum pada algoritma AGNES dipilih berdasarkan nilai maksimum *Pseudo-F* dan nilai minimum *icdrate*. Tujuan dari penelitian ini adalah untuk mengetahui *cluster* optimal dari pengelompokan 232 desa atau kelurahan di Kabupaten Kutai Kartanegara Tahun 2018. Berdasarkan hasil analisis, didapatkan 2 kelompok terbaik dengan nilai minimum *icdrate* sebesar 0,5965. *Cluster* satu beranggotakan 228 desa atau kelurahan dan *cluster* dua beranggotakan 4 desa atau kelurahan.

**Kata Kunci:** AGNES, *Icdrate*, *Podes*, *Pseudo-F*

## 1 PENDAHULUAN

*Data mining* adalah suatu proses untuk mendapatkan informasi yang berguna dari gudang basis data berskala besar yang membantu dalam proses pengambilan keputusan. *Data mining* merupakan gabungan dari sejumlah disiplin ilmu, yang didefinisikan sebagai proses penemuan suatu pola baru dari kumpulan data berskala besar, yaitu *artificial intelligence* (kecerdasan buatan), *machine learning* (pembelajaran mesin), *statistics* (statistika) dan *database system* (teknologi basis data) [1].

Salah satu teknik dalam *data mining* adalah menelusuri data yang ada untuk membangun sebuah model. Kemudian model tersebut digunakan untuk mengenali pola data lain yang tidak berada dalam basis data yang tersimpan. *Data mining* juga dapat digunakan untuk melakukan proses pengelompokan atau klasterisasi (*clustering*), dengan tujuan untuk mengetahui pola universal dari data-data yang ada [1]. Menurut [2], klasterisasi (*clustering*) adalah proses pengelompokan himpunan data ke dalam beberapa kelompok atau klaster sedemikian sehingga objek-objek dalam suatu klaster memiliki kemiripan yang tinggi (*homogen*), namun sangat berbeda (memiliki ketidakmiripan) dengan objek-objek di klaster lainnya. Kemiripan (*similarity*) dan ketidakmiripan (*dissimilarity*) dihitung berdasarkan nilai-nilai atribut yang menggambarkan objek tersebut.

Tujuan klasterisasi (*clustering*) data dapat dibedakan menjadi dua, yaitu pengelompokan untuk pemahaman dan pengelompokan untuk penggunaan. Jika tujuannya untuk pemahaman, maka kelompok yang terbentuk harus dapat menangkap struktur alami data. Pada umumnya proses pengelompokan dengan tujuan ini hanya sebagai proses awal untuk kemudian dilanjutkan dengan pekerjaan inti seperti peringkasan atau *summarization* (rata-rata, standar deviasi), pelabelan kelas pada setiap kelompok untuk kemudian digunakan sebagai data latih klasifikasi dan sebagainya. Sementara jika tujuannya adalah penggunaan, maka pada umumnya proses pengelompokan bertujuan mencari sebuah bentuk dasar kelompok yang paling representatif terhadap data, serta memberikan gambaran dari setiap objek data dalam kelompok di mana sebuah data terletak didalamnya [1].

Analisis klaster ini sudah banyak dikembangkan oleh para ahli baik untuk pemahaman maupun untuk penggunaan yang mempunyai karakteristik, kelebihan serta kekurangan. Analisis klaster dikelompokkan menjadi empat kategori yaitu metode berbasis hirarki (*hierarchical methods*), metode berbasis kepadatan (*density-based methods*), metode berbasis kisi (*grid-based methods*) dan metode berbasis partisi (*partition methods*). Pada metode yang berbasis hirarki (*hierarchical methods*), salah satu algoritma yang termasuk dalam metode ini adalah pemusatan (*agglomerative*) dan penyebaran (*divisive*) [3].

Algoritma *Agglomerative Nesting* (AGNES) pertama diperkenalkan oleh Kaufmann dan Rousseeuw pada tahun 1990. Algoritma AGNES merupakan algoritma *agglomerative hierarchical clustering* yang cukup populer bekerja secara numerik. Algoritma AGNES yang termasuk dalam metode pemusatan (*agglomerative*) dimulai dengan menghitung matriks jarak antar objek. Setiap objek berfungsi sebagai *cluster*. Kemudian secara bertahap penggabungan setiap pasangan *cluster* terdekat berdasarkan ukuran jarak dan metode penggabungan yang digunakan, sampai semua *cluster* tergabung dalam satu *cluster* [4].

Analisis klaster dapat digunakan untuk mengetahui karakteristik wilayah yang sama berdasarkan sarana, prasarana, akses, sosial, penduduk serta ekonomi. Suatu

wilayah mempunyai karakteristik tertentu, yang meliputi aspek fisik dan aspek non fisik. Aspek fisik diantaranya sarana, prasarana dan akses. Aspek non fisik diantaranya sosial-budaya, ekonomi dan penduduk. Aspek-aspek ini merupakan salah satu cara untuk menerangkan keragaman antar wilayah, salah satunya wilayah desa atau kelurahan. Suatu wilayah desa atau kelurahan memiliki kondisi aspek-aspek yang berbeda dengan wilayah lainnya. Potensi Desa (PODES) menyediakan data tentang keberadaan, ketersediaan dan perkembangan potensi yang dimiliki setiap wilayah administrasi pemerintah. Potensi tersebut meliputi sarana dan prasarana wilayah serta potensi ekonomi, sosial, budaya dan aspek kehidupan masyarakat lainnya [5].

Potensi ekonomi merupakan fokus perhatian dalam menyusun strategi pembangunan. Pembangunan ini diprioritaskan bagi daerah-daerah dengan jumlah penduduk miskin terbesar. Kabupaten Kutai Kartanegara (Kukar) tercatat sebagai penyumbang penduduk miskin terbesar di Provinsi Kalimantan Timur (Kaltim) dengan jumlah penduduk miskin 56.570 jiwa pada tahun 2017. Kota Samarinda dan Kutai Timur berada di posisi tertinggi kedua dan ketiga dengan jumlah penduduk miskin masing-masing sebesar 40.010 jiwa dan 31.950 jiwa.

Pada penelitian sebelumnya yaitu [6] Menunjukkan bahwa metode Agglomerative Nesting (AGNES) dapat dengan baik mengelompokkan data dan dari tiga metode yang di digunakan, yang mampu menghasilkan dendogram paling seimbang adalah metode average linkage. Sementara penelitian [7] juga menunjukkan bahwa metode Agglomerative Nesting (AGNES) dapat dengan baik mengelompokkan data dan dari tiga metode yang di digunakan, yang mampu menghasilkan dendogram paling seimbang adalah metode average linkage. Oleh karena itu, akan dilakukan penelitian untuk mengelompokkan wilayah desa atau kelurahan di Kabupaten Kutai Kartanegara Provinsi Kalimantan Timur berdasarkan indikator desa tertinggal dengan menggunakan algoritma Agglomerative Nesting (AGNES) dengan tiga metode penggabungan yaitu Single Linkage, Average Linkage dan Complete Linkage yang masing-masing metodenya menggunakan 2,3,4 dan 5 jarak cluster berdasarkan data hasil pendataan PODES Tahun 2018.

## **2 TINJAUAN PUSTAKA**

### **2.1 Data Mining**

Data mining adalah suatu 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. Teknik-teknik data mining yang digunakan bertugas untuk menemukan pola baru dan bermakna di dalam basis data yang mungkin masih belum diketahui [8].

Data mining termasuk dalam proses *Knowledge Discovery in Database* (KDD), yaitu kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan. Beberapa metode yang terdapat dalam literatur data mining antara lain *clustering*, *classification*, *association rules mining*, *neural network* dan *genetic algorithm* [9].

## 2.2 Transformasi Data

Transformasi data digunakan untuk mengubah data ke dalam bentuk yang sesuai untuk suatu analisis. Transformasi data dilakukan baik terhadap data numerik maupun data kategorik. Transformasi data numerik dilakukan dengan standarisasi data. Jika rentang nilai antar variabel memiliki perbedaan skala yang cukup besar yang dapat menyebabkan bias dalam analisis *cluster*, maka data asli perlu standarisasi.

Rentang nilai yang besar akan menyebabkan perhitungan jarak menjadi tidak valid [10]. Standarisasi data dapat dilakukan dengan cara semua dimensi atau sub-variabel penyusun atau *item* ditransformasi ke dalam data standar atau data Z (nilai rata-rata sama dengan nol, variansi sama dengan satu dan data tanpa satuan/relatif). Cara menentukan nilai standarisasi adalah dengan menghitung nilai rata-rata (Persamaan 2.1) dan standar deviasi (Persamaan 2.2) [11].

$$\bar{x}_l = \frac{1}{n} \sum_{i=1}^n X_{il} \quad (2.1)$$

$$s_l = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{il} - \bar{x}_l)^2} \quad (2.2)$$

Kemudian menghitung data hasil standarisasi dengan menggunakan Persamaan (2.3).

$$Z_{il} = \frac{x_{il} - \bar{x}_l}{s_l} \quad (2.3)$$

dimana

$n$  : banyaknya data

$x_{il}$  : data ke- $i$  variabel ke- $l$

$\bar{x}_l$  : rata-rata variabel ke- $l$

$Z_{il}$  : standarisasi data untuk data ke- $i$  variabel ke- $l$

$s_l$  : standar deviasi variabel ke- $l$

Transformasi untuk data kategorik dilakukan dengan cara mengubah data kategorik menjadi kode numerik. Kode numerik yang dimiliki oleh suatu atribut tidak boleh sama dengan kode numerik yang dimiliki oleh atribut lain.

## 2.3 Clustering

*Clustering* atau analisis *cluster* adalah suatu teknik *data mining* untuk mengelompokkan himpunan objek (*dataset*) ke dalam beberapa grup/*cluster* hanya berdasarkan kemiripan karakteristik dari atribut yang dimiliki oleh data objek sedemikian 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 [4].

Analisis *cluster* berbeda dengan metode klasifikasi. Klasifikasi berkaitan dengan sejumlah kelompok yang telah diketahui sebelumnya dan tujuan operasionalnya adalah untuk menetapkan pengamatan baru. Analisis *cluster* adalah teknik yang lebih primitif karena tidak ada asumsi yang dibuat mengenai jumlah kelompok atau struktur kelompok sebelumnya. Pengelompokkan dilakukan atas

dasar kemiripan atau jarak (ketidak-miripan). Analisis *cluster* mengelompokkan objek-objek yang paling dekat kesamaannya dengan objek lain dan mempunyai kemiripan satu dengan yang lain dalam kelompok yang sama [12].

Metode dalam analisis *cluster* dibagi menjadi dua yaitu, metode hierarki dan non-hierarki. Metode hierarki memulai pengelompokkan dengan dua atau lebih objek yang mempunyai kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang mempunyai kedekatan dua tipe dasar yaitu *agglomerative* (pemusatan) dan metode *divisive*. Berbeda dengan metode hierarki, metode nonhierarki dimulai dengan terlebih dahulu menentukan jumlah *cluster* yang diinginkan. Setelah jumlah cluster diketahui, kemudian proses *cluster* dilakukan tanpa mengikuti proses hierarki [1].

Terkait dengan pengertian dan tujuan dilakukannya analisis cluster, dapat dinyatakan bahwa suatu kelompok yang baik adalah kelompok yang memiliki ciri-ciri sebagai berikut [13]:

- 1) Homogenitas (kesamaan) yang tinggi antar anggota dalam satu kelompok (*within cluster*).
- 2) Heterogenitas (perbedaan) yang tinggi antar kelompok yang satu tahap dengan kelompok yang lain (*between cluster*).

Salah satu faktor yang sangat berpengaruh terhadap hasil dari kelompok yang dibentuk antar objek pengamatan adalah jarak antar objek pengamatan [14]. Oleh karena itu, dibutuhkan suatu alat ukur untuk menentukan jarak antar objek lain pengamatan. Salah satu jenis metode pengukuran jarak yang umum digunakan adalah jarak *Euclidean*. Misalkan terdapat dua observasi dengan variabel-variabel berdimensi  $m$ , yaitu  $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{im}]^T$  dan  $\mathbf{x}_j = [x_{j1}, x_{j2}, \dots, x_{jm}]^T$ . Konsep jarak *Euclidean* yang mengukur jarak antara observasi  $\mathbf{x}_i$  dan  $\mathbf{x}_j$  dinyatakan sebagai berikut [12]:

$$d_{ij} = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j)} \quad (2.4)$$

dengan

$d_{ij}$  : jarak antara objek ke- $i$  dan ke- $j$

$\mathbf{x}_i$  : vektor objek ke- $i$

$\mathbf{x}_j$  : vektor objek ke- $j$

Tahap pengelompokkan dalam analisis *cluster* dibedakan berdasarkan jenis data yang dimiliki. Pada umumnya analisis kelompok terfokus pada data numerik, namun terdapat juga kasus pengelompokkan pada data kategorik, atau bahkan data campuran numerik dan kategorik. Analisis *cluster* pada data kategorik tidak dapat dilakukan seperti data numerik. Hal tersebut dikarenakan data kategorik memiliki sifat khusus, sehingga pengelompokkan data kategorik menjadi lebih rumit dibandingkan pengelompokkan data numerik [15].

#### 2.4 Analisis Cluster Data Numerik

Analisis *cluster* untuk data numerik dilakukan berdasarkan ukuran ketidakmiripan atau jarak untuk data numerik. Hasil pengelompokkan disajikan dalam bentuk *dendrogram* (diagram pohon) yang memungkinkan penelusuran objek-objek yang diamati menjadi lebih mudah dan informatif. Teknik yang

digunakan untuk pengelompokan meliputi metode hierarki dan metode non hierarki.

## 2.5 Metode Hierarki

Metode hierarki digunakan jika banyak *cluster* yang akan dibentuk tidak diketahui sebelumnya. Terdapat dua prosedur pada metode berhierarki yaitu *Agglomerative Hierarchical Clustering* dan *Divisive Hierarchical Clustering* [3]. Pada teknik pengelompokan hierarki *agglomerative*, setiap objek merupakan satu kelompok tersendiri. Lalu dua kelompok terdekat digabungkan dan seterusnya sehingga diperoleh satu kelompok yang berunsurkan semua objek. Untuk menggabungkan dua kelompok pada tahap awal, dimana tiap kelompok hanya terdiri dari satu objek, diperlukan ukuran ketidakmiripan antar objek. Bila suatu kelompok merupakan penggabungan dari beberapa kelompok sebelumnya, maka diperlukan ukuran ketidakmiripan antar kelompok, kelompok-kelompok dengan ukuran ketidakmiripan terkecil digabungkan menjadi kelompok yang baru. Beberapa teknik pengelompokan antar kelompok dinyatakan sebagai berikut [16]:

### 1) *Single Linkage*

Metode *Single Linkage* adalah proses *clustering* yang didasarkan pada jarak terdekat antar objeknya. Jika dua objek terpisah oleh jarak pendek, maka kedua objek tersebut akan bergabung menjadi satu *cluster* dan demikian seterusnya. Pengukuran jarak metode ini dapat ditulis dalam rumus berikut:

$$d_{w(i,j)} = \min\{d_{wi}, d_{wj}\} \quad (2.5)$$

dimana  $d_{w(i,j)}$  merupakan ukuran kemiripan antara kelompok ke- $w$  dengan kelompok  $(i,j)$  yang merupakan penggabungan antara kelompok ke- $i$  dan ke- $j$

### 2) *Complete Linkage*

Pengukuran jarak pada metode *complete linkage* dilakukan berdasarkan jarak maksimum atau terjauh. Rumus yang digunakan dalam menentukan jarak pada metode ini adalah sebagai berikut:

$$d_{w(i,j)} = \max\{d_{wi}, d_{wj}\} \quad (2.6)$$

dimana  $d_{w(i,j)}$  merupakan ukuran kemiripan antara kelompok ke- $w$  dengan kelompok  $(i,j)$  yang merupakan penggabungan antara kelompok ke- $i$  dan ke- $j$ .

### 3) *Average Linkage*

Kriteria yang digunakan dalam mengukur jarak pada metode ini adalah rata-rata jarak seluruh individu dalam satu kelompok dengan jarak seluruh individu dalam kelompok lain. Dapat dituliskan dalam rumus sebagai berikut:

$$d_{w(i,j)} = \frac{\sum_q \sum_r d_{qr}}{n_{(ij)} n_w} \quad (2.7)$$

dimana  $d_{w(i,j)}$  merupakan ukuran kemiripan antara kelompok ke- $w$  dengan kelompok  $(i,j)$  yang merupakan penggabungan antara kelompok ke- $i$  dan ke- $j$ , (disimbolkan sebagai kelompok  $ij$ ).  $n_{(ij)}$  dan  $n_w$  merupakan jumlah

pengamatan dalam kelompok  $ij$  dan  $w$ , sedangkan  $d_{qr}$  merupakan jarak antara pengamatan ke- $q$  dalam kelompok  $ij$  dengan pengamatan ke- $r$  dalam kelompok  $w$ .

## 2.6 Algoritma Agglomerative Nesting (AGNES)

Algoritma AGNES pertama kali diperkenalkan oleh Kaufmann dan Rousseeuw pada tahun 1990. Algoritma AGNES merupakan algoritma *agglomerative hierarchical clustering* yang cukup populer bekerja pada numerik [4]. Algoritma AGNES dimulai dengan menghitung matriks jarak antar objek, setiap objek berfungsi sebagai *cluster*, kemudian secara bertahap menggabungkan setiap pasangan *cluster* terdekat berdasarkan ukuran jarak dan metode penggabungan yang digunakan, sampai semua *cluster* tergabung dalam satu *cluster*. Langkah-langkah algoritma AGNES dalam [8] sebagai berikut:

- 1) Menghitung matriks jarak dengan menggunakan jarak Euclidean.
- 2) Menetapkan setiap objek adalah sebuah *cluster*.
- 3) Menggabungkan dua *cluster* berdasarkan ukuran jarak dan metode penggabungan yang digunakan.
- 4) Memperbarui matriks jarak dan menghitung jarak antar *cluster* baru dengan *cluster* awal.
- 5) Mengulangi langkah 3 dan 4 sampai terbentuk sejumlah cluster yang telah ditentukan.

## 2.7 Kinerja Hasil Pengelompokkan

Pengukuran kinerja hasil pengelompokkan merupakan langkah untuk mengetahui validitas suatu pengelompokkan. Kelompok yang baik adalah kelompok yang memiliki kehomogenan yang tinggi antar anggota kelompok dan keheterogenan yang tinggi antar kelompok [13].

Indeks *R-Squared* merupakan satu indeks yang dapat digunakan untuk menentukan jumlah kelompok optimum pada pengelompokkan hierarki [14]. Indeks tersebut melibatkan perhitungan keragaman data, baik keragaman total, keragaman dalam kelompok, maupun keragaman data antar kelompok. Indeks validitas untuk menentukan jumlah kelompok optimum pada pengelompokkan hierarki dapat dituliskan sebagai berikut:

*Sum of Square Total (SST):*

$$SST = \sum_{l=1}^m \sum_{i=1}^n (x_{il} - \bar{x}_l)^2 \quad (2.10)$$

*Sum of Square Within Group (SSW):*

$$SSW = \sum_{c=1}^C \sum_{l=1}^m \sum_{i=1}^{n_c} (x_{ilc} - \bar{x}_{lc})^2 \quad (2.11)$$

*Sum of Square Between Group (SSB):*

$$SSB = SST - SSW \quad (2.12)$$

Dengan keterangan sebagai berikut:

- $m$  : jumlah variabel dalam pengamatan
- $C$  : jumlah kelompok yang dibentuk dalam pengamatan
- $n$  : jumlah objek pengamatan

- $x_{il}$  : data ke- $i$  variabel ke- $l$
- $x_{ilc}$  : data ke- $i$  variabel ke- $l$  pada kelompok ke- $c$
- $n_c$  : jumlah anggota pada kelompok ke- $c$ , untuk  $c = 1, 2, \dots, C$
- $\bar{x}_l$  : rata-rata keseluruhan objek variabel ke- $l$  untuk  $l = 1, 2, \dots, m$
- $\bar{x}_{lc}$  : rata-rata variabel ke- $l$  pada kelompok ke- $c$  untuk  $c = 1, 2, \dots, C$

Nilai  $R^2$  kelompok baru adalah rasio dari  $SSB$  dan  $SST$ , artinya  $R^2$  dapat didefinisikan sebagai ukuran perbedaan antar kelompok dengan nilai berkisaran antara 0 sampai 1. Nilai  $R^2$  sama dengan 0 menunjukkan bahwa tidak terdapat perbedaan antar kelompok. Sementara nilai  $R^2$  sama dengan satu menunjukkan adanya perbedaan yang signifikan antar kelompok yang terbentuk. Perhitungan nilai  $R^2$  terdapat pada persamaan (2.13).

$$R^2 = \frac{SSB}{SST} = \frac{[SST - SSW]}{SST} \quad (2.13)$$

Selain menggunakan nilai  $R^2$ , penentuan jumlah kelompok optimum juga dapat dilakukan dengan melihat nilai maksimum dari *Pseudo-F*. Berikut merupakan rumus yang digunakan untuk menghitung nilai *Pseudo-F*:

$$Pseudo - F = \frac{\frac{R^2}{C-1}}{\frac{1-R^2}{n-C}} \quad (2.14)$$

Setelah mendapatkan jumlah kelompok yang optimum dalam setiap metode, proses analisis dilanjutkan dengan menentukan metode pengelompokan yang paling optimum berdasarkan nilai *internal cluster dispersion rate (icdrate)* minimum. Berikut ini merupakan rumus perhitungan nilai *icdrate*:

$$\begin{aligned} icdrate &= 1 - \left[ \frac{SST - SSW}{SST} \right] \\ &= 1 - \left[ 1 - \frac{SSW}{SST} \right] \\ &= 1 - 1 + \frac{SSW}{SST} = \frac{SSW}{SST} \end{aligned} \quad (2.15)$$

Semakin kecil nilai *icdrate*, berarti semakin kecil perbedaan di dalam anggota kelompok tersebut.

## 2.8 Potensi Desa

Penentuan desa tertinggal dilakukan dalam rangka menetapkan penyaluran bantuan pemerintah agar bantuan tersebut dapat disalurkan dengan tepat. Penetapan status desa tertinggal diharapkan menjadi identifikasi daerah kemiskinan. Daerah tertinggal umumnya adalah daerah yang kondisinya relatif kurang berkembang dibandingkan daerah lain dalam skala nasional. Hal tersebut yang dicerminkan oleh empat faktor yang diduga menjadi penyebab kemajuan atau ketertinggalan suatu desa menurut [17] yaitu faktor alam atau lingkungan, faktor kelembagaan, faktor sarana/prasarana dan akses serta faktor sosial ekonomi penduduk.



Penetapan status suatu desa dilakukan melalui suatu model yang menggambarkan tingkat kemajuan/ketertinggalan desa. Hal ini dilakukan melalui sebuah kegiatan evaluasi dan studi kasus yang dapat mengelompokkan desa tertinggal untuk mengetahui karakteristik dalam bidang kemiskinan sehingga dapat menentukan desa-desa yang diprioritaskan untuk mendapatkan bantuan dari pemerintah.

Dalam pembangunan suatu model desa tertinggal diperlukan data Potensi Desa (PODES). Data PODES adalah data kewilayahan yang menekankan pada penggambaran situasi wilayah. Pendataan PODES telah dilaksanakan sejak tahun 1980. Pengumpulan data PODES dilakukan sebanyak 3 (tiga) kali dalam kurun waktu 10 tahun, sebagai bagian dari siklus 10 tahunan kegiatan sensus yang dilakukan BPS. PODES dilakukan 2 tahun sebelum pelaksanaan sensus untuk mendukung kelancaran pelaksanaan sensus. Pada tahun berakhiran 1, PODES dilaksanakan untuk mendukung Sensus Pertanian yaitu identifikasi wilayah konsentrasi usaha pertanian menurut sektor dan subsektor. Pada tahun berakhiran 4, PODES dilaksanakan untuk mendukung Sensus Ekonomi yaitu identifikasi usaha menurut sektor dan subsektor. Pada tahun berakhiran 8, PODES dilaksanakan untuk mendukung Sensus Penduduk yaitu identifikasi wilayah permukiman baru [18].

Pelaksanaan PODES 2014 diharapkan dapat membantu perencanaan kegiatan Sensus Ekonomi pada tahun 2016. Pendataan PODES tidak hanya ditujukan untuk menghasilkan data spesifik bagi keperluan pembangunan wilayah, tetapi juga dimaksudkan untuk memberikan indikasi awal tentang potensi wilayah, ketersediaan infrastruktur/fasilitas, serta kondisi sosial ekonomi dan budaya di setiap desa/kelurahan. Secara umum tujuan PODES 2014 adalah [18]:

- 1) Menyediakan data yang diharapkan dapat mendukung perencanaan kegiatan Sensus Ekonomi 2016 dari sisi wilayah kerja, anggaran, dan alokasi petugas.
- 2) Sebagai sarana untuk *updating Master File Data (MFD)*.
- 3) Menyediakan data tentang keberadaan dan perkembangan potensi yang dimiliki desa/kelurahan yang meliputi: sosial, ekonomi, sarana, dan prasarana wilayah.
- 4) Menyediakan data untuk berbagai keperluan yang berkaitan dengan perencanaan wilayah di tingkat nasional dan tingkat daerah.
- 5) Melengkapi penyusunan kerangka *sampling (sampling frame)* untuk kegiatan statistik lain lebih lanjut.
- 6) Menyediakan data bagi keperluan *updating* klasifikasi/tipologi desa, misalnya perkotaan-perdesaan, pesisir dan nonpesisir, dan sebagainya.
- 7) Menyediakan data bagi keperluan *updating* peta wilayah kerja statistik terendah.
- 8) Menyediakan data pokok bagi penyusunan statistik wilayah kecil (*small area statistics*).
- 9) Menyediakan data bagi penyusunan berbagai analisis seperti identifikasi dan penentuan.
- 10) Menyediakan data desa tertinggal, variabel konteks dalam Permukiman Masyarakat Tertinggal (PMT), dan identifikasi desa rawan bencana.

Selanjutnya dilakukan penggabungan variabel-variabel yang diduga menjadi faktor penentu ketertinggalan desa. Keempat faktor tersebut, selanjutnya dijabarkan

berdasarkan variabel-variabel yang ada dalam Potensi Desa Sensus Pertanian 2003 (PODES ST03), yang telah diidentifikasi mencakup 45 variabel sebagai berikut:

- 1) Faktor Alam/Lingkungan
- 2) Faktor Kelembagaan
- 3) Faktor Saran/Prasarana dan Akses
- 4) Faktor Sosial Ekonomi Penduduk

### 3 DATA

Data yang digunakan dalam penelitian ini menggunakan data Potensi Desa di 232 desa/kelurahan yang ada di Kabupaten Kutai Kartanegara pada tahun 2018.

Variabel penelitian yang digunakan adalah 15 variabel yang tersedia di BPS Provinsi Kalimantan Timur. Adapun 15 variabel tersebut yaitu:

- $X_1$  : Kepadatan Penduduk
- $X_2$  : Ketersediaan sarana pendidikan/sekolah
- $X_3$  : Ketersediaan tenaga kesehatan
- $X_4$  : Ketersediaan sarana kesehatan
- $X_5$  :Jumlah berlangganan telepon kabel
- $X_6$  :Jumlah penginapan
- $X_7$  :Jumlah bangunan pasar
- $X_8$  : Jumlah supermarket/toserba
- $X_9$  : Jumlah keberadaan bank
- $X_{10}$ :Jumlah pengguna fasilitas perkreditan
- $X_{11}$ : Jumlah Koperasi Non KUD lainnya
- $X_{12}$ : Jumlah keluarga pengguna listrik PLN
- $X_{13}$ : Jumlah keluarga tinggal di bantaran sungai
- $X_{14}$ : Jumlah keluarga tinggal di permukiman kumuh
- $X_{15}$ : Jumlah penderita gizi buruk

### 4 HASIL DAN PEMBAHASAN

#### 4.1 Analisis Statistika Deskriptif

Deskripsi data disajikan pada Tabel 1 sebagai berikut.

Tabel 1: Statistika Deskriptif

<b>variabel</b>	<b>Min</b>	<b>Max</b>	<b>Mean</b>
$(X_1)$	0,111	1687,111	119,858
$(X_2)$	0	32	6,655
$(X_3)$	0	29	3,039
$(X_4)$	0	80	5,069
$(X_5)$	0	700	5,138
$(X_6)$	0	9	0,341
$(X_7)$	0	3	0,388
$(X_8)$	0	475	36,730
$(X_9)$	0	6	0,327
$(X_{10})$	0	5	0,327
$(X_{11})$	0	9	0,478

Tabel 1: Statistika Deskriptif (Lanjutan)

<b>variabel</b>	<b>Min</b>	<b>Max</b>	<b>Mean</b>
$(X_{12})$	6	7261	915
$(X_{13})$	0	1125	62,160
$(X_{14})$	0	347	4,125
$(X_{15})$	0	4	0,155

Berdasarkan Tabel 4.1 terlihat bahwa rata-rata kepadatan penduduk di Kabupaten Kutai Kartanegara tahun 2018 adalah 119,858 jiwa/km<sup>2</sup>. Desa Muara Belinau adalah desa di Kabupaten Kutai Kartanegara dengan kepadatan penduduk terkecil yaitu 0,111 jiwa/km<sup>2</sup> sedangkan Kelurahan Melayu merupakan desa di Kabupaten Kutai Kartanegara yang memiliki kepadatan penduduk terbesar yaitu 1687,111 jiwa/km<sup>2</sup>.

#### 4.2 Analisis Pengelompokkan

Dari hasil analisis pengelompokkan menggunakan metode *Single Linkage*, *Complete Linkage* dan *Average Linkage* didapatkan jumlah *cluster* sebanyak 2,3,4 dan 5 *cluster* dengan hasil masing-masing sebagai berikut:

Tabel 2: Kinerja Pengelompokkan Data

Metode	Jumlah Cluster	Pseudo-F	Icdrate
<i>Single linkage</i>	2	13,804	0,943
	3	7,398	0,939
	4	21,295	0,781
	5	<b>37,379</b>	0,603
<i>Complete linkage</i>	2	<b>155,611</b>	<b>0,596</b>
	3	78,925	0,591
	4	59,762	0,559
	5	44,747	0,559
<i>Average linkage</i>	2	<b>55,581</b>	0,805
	3	28,420	0,801
	4	21,089	0,783
	5	37,197	0,060

Tabel 2 terlihat bahwa pengelompokkan potensi desa tertinggal di Kabupaten Kutai Kartanegara Provinsi Kalimantan Timur menggunakan metode *single linkage* dengan jumlah 2, 3, 4, dan 5 cluster diperoleh nilai Pseudo-F optimum berada pada jumlah 5 cluster yaitu sebesar 37,379. Pengelompokkan menggunakan metode *complete linkage* didapatkan nilai Pseudo-F optimum berada pada jumlah 2 cluster yaitu sebesar 155,611. Pengelompokkan menggunakan metode *average linkage* didapatkan nilai Pseudo-F optimum pada jumlah 2 cluster yaitu sebesar 55,581.

Selanjutnya dari ketiga metode penggabungan tersebut dibandingkan nilai *icdrate* sehingga diperoleh metode penggabungan terbaik untuk pengelompokkan potensi desa tertinggal di Kabupaten Kutai Kartanegara Provinsi Kalimantan Timur, yaitu metode *complete linkage* dengan 2 cluster, karena nilai *icdrate* dari metode ini adalah nilai *icdrate* minimum dibandingkan dengan nilai *icdrate* kedua metode lainnya yaitu sebesar 0,596. Cluster yang terbentuk terdiri dari 228 dalam cluster 1 dan 4 dalam cluster 2.

Dari hasil analisis pengelompokkan menggunakan metode *Single Linkage*, *Complete Linkage* dan *Average Linkage* didapatkan jumlah *cluster* sebanyak 2,3,4 dan 5 *cluster* dengan hasil masing-masing sebagai berikut:

Tabel 3: Karakteristik Hasil Analisis *Cluster*

Variabel	Cluster 1			Cluster 2		
	Mean	Min	Max	Mean	Min	Max
$X_1$	103,856	0,111	1.168,151	1.032,026	350,668	1.687,111
$X_2$	6,351	0	32	24	19	29
$X_3$	2,794	0	19	17	10	29
$X_4$	4,298	0	25	49	9	80
$X_5$	5,228	0	700	0	0	0
$X_6$	0,281	0	6	3,75	0	9
$X_7$	0,373	0	3	1,25	0	2
$X_8$	32,873	0	200	256,75	47	475
$X_9$	0,267	0	3	3,75	0	6
$X_{10}$	0,320	0	5	0,75	0	3
$X_{11}$	0,438	0	9	2,75	0	7
$X_{12}$	830,513	6	4.788	5.733,5	4.250	7.261
$X_{13}$	61,853	0	1.125	79,25	0	232
$X_{14}$	4,197	0	347	0	0	0
$X_{15}$	0,149	0	4	0,5	0	2

Dari segi jumlah kepadatan penduduk di wilayah desa atau kelurahan, desa atau kelurahan di *cluster* satu memiliki kepadatan penduduk lebih rendah dibandingkan dengan desa atau kelurahan di *cluster* dua. Rata-rata desa atau kelurahan di *cluster* satu memiliki kepadatan penduduk sebesar 103,856 jiwa/km<sup>2</sup> sedangkan desa atau kelurahan pada *cluster* dua memiliki rata-rata kepadatan penduduk sebesar 1.032,026 jiwa/km<sup>2</sup>.

Dari segi jumlah bangunan pasar, pada cluster satu masih ada desa atau kelurahan yang tidak memiliki bangunan pasar yaitu sebanyak 160 dari total 228 desa atau kelurahan, sedangkan pada cluster dua terdapat 1 desa atau kelurahan yang tidak memiliki bangunan pasar dari total 4 desa atau kelurahan.

Dari segi jumlah penderita gizi buruk, pada cluster satu terdapat 20 dari total 228 desa atau kelurahan memiliki penderita gizi buruk minimal 1 orang di tiap desa atau kelurahan sedangkan pada cluster dua terdapat 1 dari total 4 desa atau kelurahan memiliki penderita gizi buruk sebanyak 2 orang.

## 5 KESIMPULAN

### 5.1 Kesimpulan

Berdasarkan analisis dan pembahasan yang telah dilakukan, maka dapat ditarik kesimpulan sebagai berikut:

- 1) Hasil pengelompokkan wilayah desa atau kelurahan di kabupaten kutai kartanegara provinsi Kalimantan timur menghasilkan *cluster* optimum pada pengelompokkan dengan metode complete linkage berdasarkan nilai Icdrate terkecil. Nilai icdrate yang dihasilkan sebesar 0,5965 dengan jumlah *cluster* yang terbentuk adalah dua *cluster*. *Cluster* satu beranggotakan 228 desa atau kelurahan dan *cluster* dua beranggotakan 4 desa atau kelurahan.
- 2) Karakteristik dari masing-masing *cluster* adalah sebagai berikut:
  - a. *Cluster* satu berisikan 228 desa atau kelurahan, memiliki rata-rata kepadatan penduduk sebesar 103,86 jiwa/km<sup>2</sup>. Mayoritas desa atau kelurahan di *cluster* satu berada di dekat dengan bantaran tepi sungai atau muara.
  - b. *Cluster* dua berisikan 4 desa atau kelurahan, memiliki rata-rata kepadatan penduduk sebesar 1.032,03 jiwa/km<sup>2</sup>. Rata-rata kepadatan penduduk di *cluster* dua terbilang sangat besar jika dibandingkan dengan rata-rata kepadatan penduduk di *cluster* satu, hal ini di karenakan pada *cluster* dua hanya berisikan 4 desa atau kelurahan.

### 5.2 Saran

Sebaiknya dalam penelitian selanjutnya dapat menambahkan ukuran jarak lainnya seperti jarak *Manhattan* Atau jarak *Chebyshev*.

## DAFTAR PUSTAKA

- [1] Prasetyo, E. (2012). *Data Mining: Konsep dan Aplikasi menggunakan Matlab*. Yogyakarta: Andi Offset.
- [2] Suyanto. (2017). *Data Mining untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika.
- [3] Mattjik, A. A., & Sumetrtajaya, I. M. (2011). *Sidik Peubah Ganda dengan Menggunakan SAS*. Bogor: IPB Press.
- [4] Han, J., Kamber, M. & Pei, J. (2012). *Data Mining Concepts and Techniques Third Edition*. USA: Elsevier.
- [5] Fathia. (2016). Analisis Klaster Kecamatan Di Kabupaten Semarang Berdasarkan Potensi Desa Menggunakan Metode Ward dan Single Linkage. *Jurnal Gaussian, Volume 5, Nomor 4, ISSN:239-2541*.
- [6] Kurniawan, Aloysius Ari. (2017). *Implementasi algoritma agglomerative hierarchical clustering untuk mengelompokkan capaian belajar siswa SD*. Skripsi. Yogyakarta: Universitas Sanata Dharma.
- [7] Ariska, Nur. (2017). *Analisis Cluster dengan Metode Ensemble Rock Untuk Data Berskala Campuran Kategorik dan Numerik (Kasus: Mahasiswa Aktif Program Studi Statistika FMIPA UNM)*. Skripsi. Makassar: Universitas Negeri Makassar.
- [8] Tan, P., Steinbach, M. & Kumar, V. (2006). *Introduction to Data Mining*. USA: Person Education, Inc.
- [9] Santosa, B. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.

- [10] Yulianto, S & Hidayatullah, K. H. (2014). Analisis Kluster untuk Pengelompokan Kabupaten/Kota di Provinsi Jateng berdasarkan Indikator Kesejahteraan Rakyat. *Jurnal Akd. Statistika Muhammadiyah Semarang*, 2(1), 58-59.
- [11] Sartono, B., Affendi, F. M., Syafitri, U. D., Sumertajaya, I. M., & Anggraeni, Y. (2003). *Analisis Peubah Ganda*. Bogor: IPB Bogor.
- [12] Johnson, R.A. and Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis 6th Edition*. USA: Pearson Education, Inc.
- [13] Hair, J. F., Black, W. C., Babin, B. J. & Anderson, R. E. (2014). *Multivariate Data Analysis 7th Edition*. USA: Pearson Education, Inc.
- [14] Sharma, S. (1996). *Applied Multivariate Technique*. New York: John Wiley & Son, Inc.
- [15] Hair, J. F., Black, W. C., Babin, B. J. & Anderson, R. E. (2010). *Multivariate Data Analysis 7th Edition*. United States : Pearson.
- [16] Johnson, R.A. and Wichern, D.W. (2002). *Applied Multivariate Statistical Analysis*. Prentice Hall, New Jersey.
- [17] Agusta, I. (2007). Desa Tertinggal di Indonesia. *Jurnal Transdisiplin Sosiologi, Komunikasi dan Ekologi Manusia*, 1(2), 233-235.
- [18] Badan Pusat Statistik. (2014). *Pedoman Pencacah PODES 2014*. Jakarta: Badan Pusat Statistik.