Jurnal Matematika UNAND Vol. **IX** No. **2** Hal. 192 – 198

Edisi April 2020 ISSN : 2303–291X e-ISSN : 2721–9410

©Jurusan Matematika FMIPA UNAND

# PENGKLASTERAN KABUPATEN/KOTA DI PROVINSI SUMATERA BARAT BERDASARKAN INDIKATOR KESEJAHTERAAN MASYARAKAT DENGAN VALIDITAS KOEFISIEN *SILHOUETTE*

#### ULLYA IZZATY, IZZATI RAHMI HG, DODI DEVIANTO

Program Studi S1 Matematika,
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,
Kampus UNAND Limau Manis Padang, Indonesia.
email: ullyaizzaty48@gmail.com, izzatirahmihg@sci.unand.ac.id, ddevianto@sci.unand.ac.id

Diterima 17 Februari 2020 - Direvisi 7 Maret 2020 - Dipublikasikan 29 April 2020

Abstrak. Penelitian indikator kesejahteraan masyarakat dilakukan pengelompokan 19 kabupaten/kota di Provinsi Sumatera Barat dengan menggunakan metode analisis klaster. Tujuan penelitian ini adalah menentukan pengelompokan kabupaten/kota di Provinsi Sumatera Barat berdasarkan indikator kesejahteraan dengan menggunakan validitas Koefisien Silhouette. Indikator kesejahteraan masyarakat yang digunakan yaitu, pendidikan, kesehatan, kriminalitas, kemiskinan, sosial, dan ketenagakerjaan. Metode analisis yang digunakan yaitu, metode berhirarki dan metode tak berhirarki. Metode berhirarki yang digunakan pada penelitian ini yaitu, metode pautan tunggal, pautan lengkap, pautan rata-rata, dan Ward. Metode tak berhirarki yang digunakan yaitu, metode K-Means. Anggota klaster yang diperoleh dengan menggunakan metode pautan tunggal, pautan lengkap, pautan rata-rata, Ward dan K-Means dilakukan validitas dengan Koefisien Silhouette. Data yang digunakan pada penelitian ini merupakan data sekunder yang diperoleh dari Badan Pusat Statistik tahun 2017. Hasil penelitian ini didapatkan bahwa pengklasteran terbaik yang diperoleh dengan menggunakan validitas Koefisien Silhouette yaitu metode pautan rata-rata dan banyaknya klaster optimal diperoleh sebanyak tujuh klaster. Pada klaster 2 dan 4 diperoleh masing-masing tujuh kabupaten/kota dan pada klaster 1,3,5,6, dan 7 masing-masing terdiri dari satu kota. Klaster 6 yang terdiri dari Kota Padang memiliki nilai yang tertinggi hampir disetiap indikator, yaitu indikator kriminalitas, kemiskinan, dan sosial.

 $\it Kata\ Kunci$ : Analisis Klaster, Koefisien  $\it Silhouette,$  Indikator Kesejahteraan Masyarakat,Sumatera Barat

#### 1. Pendahuluan

Kesejahteraan masyarakat merupakan tujuan setiap masyarakat Indonesia yang diperjuangkan sejak sebelum kemerdekaan Indonesia. Kesejahteraan masyarakat masih menjadi permasalahan di Indonesia yang belum tuntas hingga saat ini, tidak

<sup>\*</sup>penulis korespondensi

terkecuali di Provinsi Sumatera Barat yang memiliki daerah yang luas. Telah banyak upaya yang dilakukan pemerintah untuk meningkatkan kesejahteraan masyarakat, baik pembangunan fisik maupun non fisik. Namun, setiap daerah memiliki jumlah penduduk, kondisi geografi, dan tingkat ekonomi yang berbeda-beda sehingga sulit untuk menetapkan kebijakan pada setiap kabupaten/kota di Provinsi Sumatera Barat. Berdasarkan hal ini, pemerintah perlu mengetahui gambaran umum kondisi setiap kabupaten/kota agar dapat mengelompokkan kabupaten/kota yang memiliki kondisi yang sama. Dengan pengelompokkan kabupaten/kota yang memiliki kondisi yang sama, pemerintah dapat menentukan kebijakan untuk tercapainya kesejahteraan masyarakat.

Terdapat metode analisis dalam statistika yang dikenal dengan metode analisis klaster untuk pengelompokkan objek-objek yang memiliki persamaan karakteristik. Dengan metode analisis klaster akan dikelompokkan kabupaten/kota yang memiliki kondisi yang sama. Metode analisis klaster memiliki dua pendekatan yaitu, metode berhirarki (hierarchical clustering methods) dan metode tidak berhirarki (non hierarchical clustering methods) [1,6]. Metode berhirarki yang digunakan yaitu, pautan tunggal, pautan lengkap, pautan rata-rata, dan metode Ward. Sedangkan metode tidak berhirarki yang akan digunakan yaitu metode K-Means. Untuk mengevaluasi hasil algoritma clustering dalam konsep kuantitatif yang didapat dari data dari kelima metode ini, selanjutnya akan ditentukan validitas klaster untuk mengetahui metode terbaik dan banyaknya klaster yang optimal dengan menggunakan Koefisien Silhouette.

Tujuan penelitian ini adalah menentukan pengelompokan kabupaten/kota di Provinsi Sumatera Barat berdasarkan indikator kesejahteraan dengan menggunakan validitas Koefisien Silhouette.

#### 2. Landasan Teori

#### 2.1. Analisis Klaster

Analisis klaster merupakan metode analisis peubah ganda untuk mengelompokkan n objek ke dalam m klaster r (m < n) berdasarkan karakteristiknya [1,3]. Hasil pengklasteran tergantung pada objek yang digunakan, peubah yang diamati, ukuran kemiripan dan ketakmiripan yang digunakan, serta metode pengklasteran yang digunakan [6]. Dalam [6], Ciri-ciri klaster adalah:

- (a) Homogenitas (kesamaan) yang tinggi antar anggota dalam satu klaster (within-cluster),
- (b) Heterogeneritas (perbedaan) yang tinggi antar klaster yang satu dengan klaster yang lain (between cluster).

### 2.2. Metode Pengklasteran

#### 2.2.1. Metode Berhirarki

Metode ini bisa digunakan jika peneliti belum mengetahui banyak klaster yang akan dibentuk dan ukuran contoh relatif kecil. Terdapat beberapa metode perbaikan

jarak berhirarki dan tak berhirarki yang dapat digunakan, beberapa diantaranya adalah:

(a) Metode Pautan Tunggal (Single Linkage).

Metode ini merupakan metode yang paling sederhana diantara metode perbaikan jarak yang lain [3]. Metode ini mengambil jarak terkecil, dirumuskan sebagai berikut:

$$d((ij)k) = \min\{d(ik), d(jk)\},\tag{2.1}$$

dengan:

 $d(ik) = jarak \ antara \ klaster \ i \ dan \ k$  $d(jk) = jarak \ antara \ klaster \ j \ dan \ k$ 

(b) Metode Pautan Lengkap (Complete Linkage). Metode ini hampir sama dengan metode pautan tunggal [3]. Metode ini mengambil jarak terbesar, dirumuskan sebagai berikut:

$$d((ij)k) = \max\{d(ik), d(jk)\},\tag{2.2}$$

dengan:

 $d(ik) = jarak \ antara \ klaster \ i \ dan \ k$  $d(jk) = jarak \ antara \ klaster \ j \ dan \ k$ 

(c) Metode Pautan Rataan (Average Linkage).

Pautan rataan memperlakukan jarak antara dua klaster sebagai jarak setiap anggota klaster dengan klaster lainnya, dirumuskan sebagai berikut:

$$d(ij)k = \frac{\sum a \sum b(d_a b)}{n((i,j)nk}$$
(2.3)

dengan:

 $(d_ab)$  = jarak antara objek a pada klaster (i,j) dan objek b pada klaster k n((i,j)) = banyaknya objek pada klaster (i,j) = banyaknya objek pada klaster k

(d) Metode Ward.

Metode Ward merupakan metode analisis klaster yang didasari karena adanya informasi yang hilang akibat penggabungan objek menjadi klaster. Fungsi objektif yang digunakan adalah *error sum of square* (ESS) dengan rumus sebagai berikut:

$$ESS = \frac{1}{n} \sum_{i} (i < j) d_i^2 j = \frac{1}{n} \sum_{i} (i < j) \sum_{k=1}^{p} (x_i k - x_j k)^2, \qquad (2.4)$$

dengan:

 $(d_i^2 j) = jarak kuadrat Euclid objek ke-i dan objek ke-k$ 

 $x_i k$  = nilai pada objek ke-i pada peubah ke-k  $x_i k$  = nilai pada objek ke-i pada peubah ke-k

p = banyaknya peubah

 $n = banyaknya \ objek \ dalam \ suatu \ klaster$ 

# 2.2.2. Metode Tak Berhirarki

Metode tak berhirarki digunakan untuk mengelompokkan objek dimana banyaknya klaster yang akan dibentuk dapat ditentukan terlebih dahulu sebagai bagian dari prosedur pengklasteran. Metode tak berhirarki yang umum digunakan adalah krataan (k-means) [3]. Rumus metode K-Means sebgai berikut:

$$c_i j = \frac{\sum_{k=1}^{N_i} (x_k j)}{N_i},\tag{2.5}$$

dengan:

 $x_k j = nilai pengamatan objek ke-k untuk peubak ke-j dalam klaster ke-i$  $<math>N_i = banyaknya anggota klaster ke-i$ 

### 2.2.3. Jarak Euclid

Jarak Euclid digunakan bila peubah-peubah yang digunakan tidak berkorelasi dan memiliki satuan yang sama. Jarak Euclid dirumuskan sebagai berikut:

$$d_i j = \sqrt{\sum_{k=1}^{p} (x_i k - x_j k)^2},$$
(2.6)

atau

$$d_i j = \sqrt{(x_i 1 - x_j 1)^2 + (x_i 2 - x_j 2)^2 + \dots + (x_i n - x_j n)^2}.$$
 (2.7)

#### 2.3. Silhouette Coefficient

Silhouette Coefficient digunakan untuk melihat kualitas dan kekuatan klaster, seberapa baik suatu objek ditempatkan dalam suatu klaster. Metode ini merupakan gabungan dari metode kohesi (cohesion) dan pemisahan (separation). Kohesi digunakan untuk mengukur kedekatan objek yang berada pada satu klaster, sedangkan pemisahan digunakan untuk mengukur kedekatan antar klaster yang terbentuk. Nilai Koefisien Silhouette suatu objek i sebagai berikut [3]:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}},$$
(2.8)

dengan:

 $a(i) = jarak \ rata-rata \ antara \ objek \ i \ dengan \ seluruh \ objek \ yang \ berada \ dalam \ klaster \ yang \ sama$ 

b(i) = jarak rata-rata minimum antara objek i dengan seluruh objek yang berada pada klaster terdekat

Koefisien Silhouette memiliki rentang nilai antara -1 sampai 1. Nilai Koefisien Silhouette yang mendekati 1 adalah yang lebih baik. Nilai Koefisien Silhouette suatu klaster yang terdiri dari N objek dapat dihitung dengan rata-rata dari seluruh nilai Koefisien Silhouette (SLH) setiap objek i sebagai berikut [1,4]:

$$SLH = \frac{1}{N} \sum_{i=1}^{N} S(i).$$
 (2.9)

Hasil pengklasteran terbaik dicapai ketika SLH maksimum, hal ini berarti meminimumkan jarak antar objek dalam klaster, sekaligus memaksimumkan jarak antar klaster [5].

#### 3. Pembahasan

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari publikasi Badan Pusat Statistika (BPS) tahun 2017 yaitu, data kabupaten/kota di Provinsi Sumatera Barat B. Terdapat 26 peubah yang digunakan berdasarkan indikator kesejahteraan masyarakat menurut BPS yang terdiri dari 6 indikator, yaitu kriminalitas, kesehatan, pendidikan, kemiskinan, kependudukan, jumlah pendapatan, pengeluaran atau konsumsi per kapita, ketenagakerjaan, dan lingkungan.

## 3.1. Klaster Terbaik dengan Koefisien Silhouette

Setelah didapatkan anggota klaster dengan menggunakan metode pautan tunggal, pautan lengkap, pautan rata-rata, Ward dan K-Means, maka ditentukan banyaknya jumlah klaster optimal dengan validitas, yaitu uji validitas Koefisien *Silhouette*.

Banyaknya klaster yang optimal adalah 7 klaster karena memiliki nilai Koefisien *Silhouette* relatif paling besar yaitu 0,473. Sehingga dapat disimpulkan bahwa metode pautan rata-rata adalah metode yang terbaik karena menghasilkan klaster yang terbaik pada kasus ini. Selanjutnya akan dilihatkan hasil pengklasteran yang diperoleh dengan metode pautan rata-rata pada tabel berikut.

Berdasarkan Tabel 3.1.2 dapat dilihat bahwa sebagian besar kab/kota berada pada klaster 2 dan 4 dengan persentase masing-masing 36,842%. Sedangkan pada klaster 1,3,5,6 dan 7 beranggotakan satu kab/kota dengan persentase masing-masing 5,263%.

# 3.2. Karakteristik Klaster yang Terbentuk

Klaster 1 terdiri dari Kep. Mentawai dengan peubah tertinggi yaitu, jumlah penduduk yang bekerja bukan disektor pertanian, persentase penduduk miskin menurut pengeluaran perkapita, dan tingkat partisipasi angkatan kerja (TPAK). Klaster 3 terdiri dari Sijunjung dengan penyandang masalah sosial anak terlantar tertinggi. Klaster 5 terdiri dari Pasaman Barat dengan peubah tertinggi yaitu, jumlah angka penduduk buta huruf usia 15-64 tahun keatas, jumlah fasilitas kesehatan, jumlah

Klaster 2 klaster Nilai SHL Metode Pautan tunggal, pautan lengkap, pautan rata-rata, dan K-Means Ward  $0,508 \\ 0,379$ Pautan tunggal Pautan lengkap dan K-Means Pautan rata-rata  $0,169 \\ 0,307$ 3 Klaster  $0,346 \\ 0,304$ Ward Pautan tunggal Pautan lengkap dan Ward Pautan rata-rata 0,2180,2544 Klaster K-Means 0.2375 Klaster Pautan tunggal Pautan lengkap dan Ward Pautan rata-rata dan K-Means 0.3116 Klaster Pautan tunggal Pautan lengkap  $0,409 \\ 0,322$ Pautan rata-rata Ward K-Means Pautan tunggal  $0,298 \\ 0,4$ 7 Klaster  $0,321 \\ 0,409$ Pautan lengkap Pautan rata-rata Ward
K-Means
Pautan tunggal
Pautan lengkap 0,322 0,298 0,4588 Klaster 0,364 Pautan rata-rata dan K-Means Ward

Tabel 1. Nilai Koefisien Silhouette Masing-Masing Klaster

Tabel 2. Anggota Masing-Masing Metode Pautan Rata-Rata

K-Means

Klaster	Persentase	Kabupaten/Kota
Klaster 1	5,263%	Kep. Mentawai
Klaster 2	36,842%	Pesisir Selatan Kab, Solok Tanah Datar Padang Pariaman Agam
		Lima Puluh Kota Pasaman
Klaster 3	5,263%	Sijunjung
Klaster 4	36,842%	Solok Selatan, Dhamasraya Kota Solok Sawahlunto Padang Panjang
		Payakumbuh Pariaman
Klaster 5	5,263%	Pasaman Barat
Klaster 6	$5,\!263\%$	Padang
Klaster 7	5,263%	Bukittinggi

puskesmas dan fasilitas tenaga medis serta paramedis, jumlah penduduk yang bekerja di sektor pertanian dan laju petumbuhan penduduk.

Klaster 6 terdiri Padang yang hampir disetiap peubah memiliki nilai tertinggi yaitu, angka partisipasi sekolah usia 5 tahun keatas, jumlah rumah sakit dan tenaga medis serta paramedis, jumlah kelahiran yang hidup, jumlah tindak pidana umum menurut kepolisian resort, jumlah kecelakaan lalu lintas, jumlah penduduk miskin, jumlah penduduk yang tidak bekerja, penyandang masalah sosial disabilitas, penyandang masalah sosial fakir miskin, jumlah penduduk, penduduk berumur diatas 15 tahun pengangguran terbuka selama seminggu, dan angkatan kerja. Klaster 7 terdiri dari Bukittinggi dengan peubah tertinggi yaitu, jumlah tenaga kesehatan, daerah rawan banjir, dan kepadatan penduduk. Klaster 2 terdiri dari Pesisir Selatan, Kab. Solok, Tanah Datar, Padang Pariaman, Agam, Lima Puluh Kota, dan Pasaman dengan penyandang masalah wanita rawan sosial ekonomi tertinggi. Klaster 4 terdiri

dari Solok Selatan, Dhamasraya, Solok, Sawahlunto, Padang Panjang, Payakumbuh, dan Pariaman.

#### 4. Kesimpulan

Berdasarkan hasil penelitian ini dapat diambil kesimpulan hasil pengklasteran terbaik yang diperoleh dengan menggunakan validitas Koefisien *Silhouette* yaitu metode pautan rata-rata dan banyaknya klaster optimal diperoleh sebanyak tujuh klaster. Pada klaster 2 dan 4 diperoleh masing-masing tujuh kabupaten/kota dan pada klaster 1, 3, 5, 6, dan 7 masing-masing terdiri dari satu kota. Klaster 6 yang terdiri dari Kota Padang memiliki nilai yang tertinggi hampir disetiap indikator, yaitu indikator kriminalistas, kemiskinan, dan sosial.

# 4.1. Ucapan Terima Kasih

Terima kasih kepada ibu Hazmira Yozza, bapak Ahmad Iqbal Baqi, dan bapak Yudiantri Asdi yang telah memberikan kritik dan saran untuk penulisan artikel ini.

# Daftar Pustaka

- [1] Afrimayani, Yozza, H., dan Devianto, D., 2019, Pengelompokan Negara di Dunia Berdasarkan Data Runtun Waktu Realisasi Penanaman Modal Asing di Indonesia Menggunakan Analisis Cluster. *Skripsi S1*, diterbitkan. Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Andalas, Padang.
- [2] Badan Pusat Statistika. 2018. Provinsi Sumatera Barat dalam Angka 2018. BPS Provinsi Sumatera Barat. Padang
- [3] Johson R.A. and D.W Whinchern. 2007. Applied Multivariate Statistical Analysis. New Jersey. Prentice Hall.
- [4] Kaufman I. Rousseeuw. 1990. Finding Groups in Data. An Introduction to Cluster Analysis. New Jersey. John Wiley and Sons.
- [5] Vendramin L. Campello RJGB. Hruschka ER. 2009. On the Comparison of Relative Clustering Validity Criteria. Proceedings of the 2009 SIAM International Conference on Data Mining. 733 744.
- [6] Widiyanti, W. 2017. Sentroid Awal Metode K-Means dengan Menggunakan Pendekatan Metode Berhirarki dalam Pengelompokan Provinsi di Indonesia. Skripsi S1, diterbitkan. Departemen Matematika dan Ilmu Pengetahuan Alam Institut Pertanian Bogor, Bogor.