ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 8, Nomor 2, Tahun 2019, Halaman 161 - 170
Online di: http://ejournal3.undip.ac.id/index.php/gaussian

METODE *k-MEDOIDS CLUSTERING* DENGAN VALIDASI *SILHOUETTE INDEX* DAN *C-INDEX*

(Studi Kasus Jumlah Kriminalitas Kabupaten/Kota di Jawa Tengah Tahun 2018)

Milla Alifatun Nahdliyah[1], Tatik Widiharih [2], Alan Prahutama[3]
[1,2,3] Departemen Statistika FSM Universitas Diponegoro
widiharih@gmail.com

ABSTRACT

The k-medoids method is a non-hierarchical clustering to classify n object into k clusters that have the same characteristics. This clustering algorithm uses the medoid as its cluster center. Medoid is the most centrally located object in a cluster, so it's robust to outliers. In cluster analysis the objects are grouped by the similarity. To measure the similarity, it can be used distance measures, euclidean distance and cityblock distance. The distance that is used in cluster analysis can affect the clustering results. Then, to determine the quality of the clustering results can be used the internal criteria with silhouette width and C-index. In this research the k-medoids method to classify of regencies/cities in Central Java based on type and number of crimes. The optimal cluster at k= 4 use euclidean distance, where the silhouette index= 0,3862593 and C-index= 0,043893.

Keywords: Clustering, k-Medoids, Euclidean distance, Cityblock distance, Silhouette index, C-index, Crime

1. Pendahuluan

Jawa Tengah merupakan salah satu provinsi di Indonesia yang cukup besar dan padat penduduk. Banyaknya jumlah penduduk dapat menimbulkan berbagai masalah sosial di dalam masyarakat seperti kemiskinan, pengangguran, dan kesenjangan sosial ekonomi. Masalah sosial juga dapat mendorong beberapa orang untuk melakukan tindak kriminalitas atau kejahatan. Kriminalitas atau kejahatan adalah suatu perbuatan yang dapat mengakibatkan timbulnya masalah-masalah dan keresahan bagi kehidupan masyarakat (Abdulsyani, 1987). Kasus kriminalitas yang masih terjadi di Jawa Tengah menjadi tugas bersama antara penegak hukum dan masyarakat dalam meminimalisir kasus kriminalitas yang terjadi. Salah satu langkah awal yang dapat dilakukan yaitu mengelompokkan daerah-daerah mana yang memiliki karakteristik yang sama berdasarkan kasus kriminalitas. Pengelompokan daerah kriminalitas dapat dilakukan dengan menggunakan analisis kelompok atau klaster.

Analisis klaster adalah teknik pengklasteran untuk mengelompokkan objek berdasarkan karakteristik yang dimiliki dari objek tersebut. Salah satu metode analisis klaster yang sering digunakan adalah metode non hierarki *partitioning*. Metode tersebut objek-objek pada data dikelompokkan ke dalam *k* klaster, dengan banyak klaster ditentukan oleh peneliti. Metode partitioning yang sering digunakan yaitu metode *k-means* dan *k-medoids*. Menurut Han dan Kamber (2006), algoritma *k-means* sensitif terhadap pencilan, karena menggunakan nilai rata-rata (*mean*) sebagai pusat kelompoknya. Untuk mengatasi hal tersebut, digunakan metode *k-medoids* untuk mengelompokkan objek-objek pada suatu data yang mengandung pencilan. Algoritma pengklasteran ini menggunakan *medoid* sebagai pusat klasternya. *Medoid* merupakan objek yang letaknya terpusat di dalam suatu klaster.

Pada analisis klaster, objek-objek dikelompokkan berdasarkan kemiripannya, untuk mengukur tingkat kemiripan dapat digunakan ukuran jarak. Semakin besar nilai jarak yang diperoleh, maka semakin jauh letak objek dengan pusat klaster yang terbentuk. Pada

penelitian ini akan digunakan dua ukuran jarak yaitu jarak *euclidean* dan jarak *cityblock*. Hal yang perlu juga diperhatikan dalam analisis klaster yaitu validasi hasil pengklasteran. Validasi hasil pengklasteran dilakukan untuk memperoleh partisi yang paling sesuai dengan data. Jika tidak divalidasi, maka akan berpengaruh pada hasil analisis. Pada penelitian ini digunakan dua validasi dengan pendekatan kriteria internal yaitu validasi *silhouette index* dan *C-index*.

Kabupaten/kota di Jawa Tengah akan dikelompokkan menjadi 4 klaster berdasarkan jumlah kriminalitas yang terjadi pada tahun 2018, sehingga dapat diketahui tinggi/rendahnya kriminalitas pada masing-masing kabupaten/kota di Jawa tengah. Pengelompokan tersebut dapat dilakukan dengan menggunakan analisis klaster *k-medoids*.

2. Tinjauan Pustaka

2.1. Pencilan/*Outlier*

Pencilan merupakan pengamatan yang tidak mengikuti sebagian besar pola dan terletak jauh dari pusat data. Menurut Hair *et al*. (2010), adanya pencilan dapat mengakibatkan kurang tepatnya hasil analisis yang diperoleh dan tidak mewakili keadaan populasi. Pada kasus multivariat, metode yang dapat digunakan untuk mendeteksi pencilan adalah pengukuran jarak kuadrat mahalanobis (Folzmiser, 2005). Pengukuran jarak kuadrat mahalanobis objek ke- i dapat dihitung dengan rumus sebagai berikut :

$$d_{MD}^2(i) = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})$$

dengan, $d_{MD}^2(i)$: jarak kuadrat mahalanobis objek pada pengamatan ke- i, $\boldsymbol{x}_i$: vektor data objek pada pengamatan ke- i berukuran p x1, $\overline{\boldsymbol{x}}$: vektor rata-rata dari tiap variabel berukuran p x 1, dan $\boldsymbol{\Sigma}$**:** matriks kovariansi berukuran p x p, dimana p banyaknya variabel. Pengamatan ke- i teridentifikasi pencilan jika,

$$d_{MD}^2(i) > \chi_{p,1-\alpha}^2$$

dimana $\chi_{p,1-\alpha}^2$ merupakan batas pencilan dengan probabilitas $1-\alpha$.

2.2. Analisis Klaster

Analisis klaster adalah teknik pengklasteran untuk mengelompokkan objek berdasarkan karakteristik yang dimiliki oleh objek tersebut. Objek diklasifikasikan ke dalam satu atau lebih klaster sehingga objek-objek yang berada di dalam klaster akan mempunyai kemiripan atau kesamaan karakter (Hair *et al*., 2010).

Menurut Hair *et al*., (2010) terdapat dua asumsi dalam analisis klaster yaitu sampel yang representatif dan tidak terjadi multikolinieritas.

a. Sampel Representatif

Sampel representatif adalah sampel yang diambil dapat mewakili populasi yang ada. Pengujian sampel yang mewakili populasi dapat dilakukan dengan melihat syarat kecukupan suatu sampel menggunakan uji KMO (*Kaiser Mayer Olkin*).

Berikut uji hipotesis untuk melihat apakah sampel dapat mewakili populasi:

Hipotesis

H_0 : Sampel mewakili populasi atau sampel representatif

H_1 : Sampel tidak mewakili populasi atau sampel tidak representatif

Statistik Uji

Menurut Widarjono (2010), rumus KMO adalah sebagai berikut :

$$KMO = \frac{\sum_{j=1}^{p} \sum_{l=1, l \neq j}^{p} r_{x_j x_l}^2}{\sum_{j=1}^{p} \sum_{l=1, l \neq j}^{p} r_{x_j x_l}^2 + \sum_{j=1}^{p} \sum_{l=1, k \neq j}^{p} \rho_{x_j x_k, x_m}^2}$$

dengan, p: banyaknya variabel, n: banyaknya objek, x_i: objek pada pengamatan ke-i, $r_{x_j x_l}$: korelasi antara variabel x_j dan x_l, $\bar{x}_j$: rata-rata variabel x_j, $\bar{x}_l$: rata-rata variabel x_l, dan $\rho_{x_j x_l, x_m}$: korelasi parsial antara variabel x_j dan x_l dengan menjaga agar x_m konstan

Kriteria Uji

Sampel dikatakan dapat mewakili populasi atau sampel representatif apabila diperoleh nilai KMO berkisar antara 0,5 sampai dengan 1

b. Tidak Terjadi Multikolinieritas

Multikolinieritas adalah suatu peristiwa dimana terjadi korelasi yang kuat antara dua atau lebih variabel. Multikolinieritas merupakan masalah yang perlu diperhatikan dalam analisis klaster, karena dapat mempersulit dalam menentukan pengaruh/efek dari maring-masing variabel dan mempengaruhi hasil pengklasteran akhir. Pada analisis klaster sebaiknya variabel-variabel tidak terindikasi multikolinieritas (Hair *et al*., 2010). Berikut uji hipotesis untuk melihat multikolinieritas data:

Hipotesis

H_0 : Tidak ada hubungan linier antar variabel (Tidak terjadi multikolinieritas)

H_1 : Ada hubungan linier antar variabel (Terjadi multikolinieritas)

Statistik Uji

Menurut Gujarati (2009), salah satu cara indentifikasi adanya multikolinieritas adalah menghitung nilai *Variance Inflation Factor* (VIF) yang dirumuskan sebagai berikut:

$$VIF_i = \frac{1}{(1 - R_i^2)}$$

dengan, R_i^2: koefisien determinasi yang diperoleh bila nilai variabel ke - *i* diregresikan dengan variabel lainnya

Kriteria Uji

Jika nilai VIF>10 maka H_0 ditolak sehingga terjadi multikolinieritas antar variabel.

Salah satu solusi untuk menangani data yang terjadi multikolinearitas adalah dengan *Principal Component Analysis* (PCA) atau disebut juga dengan analisis komponen utama. Menurut Johnson dan Wichern (2007), analisis komponen utama adalah suatu analisis yang bertujuan untuk mentransformasikan p variabel asal yang masih berkorelasi satu dengan yang lain menjadi satu set variabel baru yang tidak berkorelasi lagi. Variabel-variabel baru ini saling ortogonal dan merupakan kombinasi linear dari variabel asal tanpa menghilangkan sebagian besar informasi yang terkandung dalam variabel asal.

Pada penelitian ini digunakan dua jarak untuk membandingkan hasil klaster dengan jarak yang berbeda yaitu jarak *euclidean* dan jarak *cityblock.*

a. Jarak *Euclidean*

Jarak *euclidean* merupakan akar dari jumlah kuadrat selisih antar objek yang dikuadratkan (Supranto, 2004). Adapun persamaan untuk menghitung jarak *euclidean* adalah sebagai berikut:

$$d_{euc}(x_i, c_k) = \sqrt{\sum_{j=1}^{p}(x_{ij} - c_{kj})^2} \text{ , } i = 1,2,3,\ldots,n \text{ dan } k = 1,2,3,\ldots,k$$

dengan, x_{ij} : Objek pada pengamatan ke-*i* pada variabel ke-*j*

c_{kj} : Pusat kelompok ke-*k* pada variabel ke-*j*

b. Jarak *Cityblock*

Jarak *cityblock* atau manhattan jarak merupakan jumlah selisih mutlak/absolut pada setiap objek (Supranto, 2004). Adapun persamaan untuk jarak *cityblock* adalah sebagai berikut:

$$d_{cb}(x_i, c_k) = \sum_{j=1}^{p}|x_{ij} - c_{kj}| \text{ , } i = 1,2,3,\ldots,n \text{ dan } k = 1,2,3,\ldots,k$$

Metode *k-medoids* atau dikenal pula dengan *PAM* (*Partitioning Around Medoids*) menggunakan metode partisi *clustering* untuk mengelompokkan sekumpulan *n* objek ke dalam *k* klaster. Algoritma pengelompokan ini menggunakan *medoid* sebagai pusat klasternya. *Medoid* merupakan objek yang letaknya terpusat di dalam suatu klaster. Berikut tahapan-tahapan algoritma *k-medoids*, yaitu:

1. Menentukan *k* sebagai banyaknya klaster yang ingin dibentuk
2. Membangkitkan *k* pusat klaster (*medoid*) secara acak
3. Menghitung jarak objek *non-medoid* dengan *medoid* pada tiap klaster dan menempatkan tiap objek *non-medoid* tersebut ke *medoid* terdekat, kemudian hitung total jaraknya
4. Memilih secara acak objek *non-medoid* pada masing-masing klaster sebagai kandidat *medoid* baru
5. Menghitung jarak setiap objek *non-medoid* dengan *medoid* baru dan menempatkan tiap objek *non-medoid* tersebut ke kandidat *medoid* terdekat, kemudian hitung total jaraknya
6. Menghitung selisih total jarak ($S_{\text{total jarak}}$), dengan $S_{\text{total jarak}}$ = total jarak pada kandidat *medoid* baru – total jarak pada *medoid* lama
7. Jika diperoleh nilai $S_{\text{total jarak}} < 0$, maka kandidat *medoid* baru tersebut menjadi *medoid* baru dan jika diperoleh S_{total} jarak > 0 iterasi berhenti
8. Kembali ke langkah (4) sampai (7) sampai tidak terjadi perubahan *medoid* atau $S_{\text{total jarak}} > 0$

Validasi hasil analisis klaster dilakukan untuk memperoleh partisi yang paling sesuai dengan data. Jika klaster tidak divalidasi, maka akan berpengaruh pada hasil analisis. Pada penelitian ini digunakan dua validasi untuk memilih jarak dan validasi terbaik dalam pengklasteran *k-medoids* yaitu kriteria internal *silhouette index* dan *c-index*.

a. Metode Validasi *Silhouette Index*

Metode validasi *silhouette index* merupakan salah satu ukuran validasi yang berbasis kriteria internal. *Silhouette index* akan mengevaluasi penempatan setiap objek dalam setiap klaster dengan membandingkan jarak rata-rata objek dalam satu klaster dan jarak antara objek dengan klaster yang berbeda (Aini *et al*., 2014). Menghitung koefisien *silhouette* yang didefinisikan sebagai rata-rata s(i) yaitu,

$$SC = \frac{1}{n}\sum_{i=1}^{n} s(i)$$

dengan, $s(i) = \frac{b(i)-a(i)}{max(a(i),b(i))}$, $b(i) = min\, d(i,C)$, $dan\ a(i) = \frac{1}{|A|-1}\sum_{j\in A, j\neq i} d(i,j)$

$b(i)$: nilai minimum dari jarak rata-rata objek *i* dengan semua objek pada klaster lain C

$a(i)$: rata -rata jarak objek ke-*i* dengan semua objek yang berada di dalam satu klaster A

Hasil perhitungan nilai koefisien *silhouette* berada pada *range* -1 sampai 1. Semakin besar nilai koefisien *silhouette* akan semakin baik kualitas suatu kelompok

b. Metode Validasi *C-Index*

Menurut Charrad *et al*. (2014), metode validasi *C-index* merupakan validasi klaster internal yang menunjukan ukuran jarak antar klaster dan jarak dalam klaster dengan membandingkan selisih jumlah jarak antar objek di dalam tiap klaster dan jumlah minimum jarak antar objek dengan selisih jumlah maksimum dan minimum jarak antar objek. Nilai CI ditunjukan rumus sebagai berikut:

$$C - Index = \frac{S_W - S_{min}}{S_{max} - S_{min}}\ , S_{max} \neq S_{max}$$

dengan, S_{min} : Jumlah jarak nilai minimum dalam klaster dan antar klaster

S_{max} : Hitung jumlah jarak nilai maksimum dalam klaster dan antar klaster

Nilai CI berada pada *range* 0 sampai 1, nilai minimum dari indeks ini digunakan untuk menunjukkan pengklasteran yang optimal.

3. Metodologi Penelitian

Data yang digunakan merupakan data sekunder yang diperoleh dari Direktorat Reserse Kriminal Umum Polda Jawa Tengah. Data tersebut merupakan data jumlah dan jenis kriminalitas tiap Kabupaten/Kota di Provinsi Jawa Tengah tahun 2018.

Variabel yang digunakan dalam penelitian ini adalah Jumlah kejahatan terhadap nyawa atau pembunuhan (X_1), Jumlah kejahatan terhadap kesusilaan (X_2), Jumlah kejahatan penganiayaan (X_3), Jumlah kejahatan perjudian (X_4), Jumlah kejahatan pengeroyokan (X_5), Jumlah kejahatan pencurian (X_6), Jumlah kejahatan penggelapan (X_7), Jumlah kejahatan penipuan (X_8), Jumlah kejahatan perlindungan anak (X_9), Jumlah kejahatan pemerasan (X_{10}), Jumlah kejahatan kekerasan dalam rumah tangga/KDRT (X_{11}), Jumlah kejahatan penghancuran dan perusakan barang (X_{12})

Tahapan analisis data dengan *k-medoids* adalah sebagai berikut:

1. Melakukan deteksi *outlier* dengan jarak kuadrat mahalanobis.
2. Uji asumsi dalam analisis klaster yaitu
   a. Melakukan uji asumsi sampel mewakili populasi (representatif), dengan menggunakan uji *Kaiser Mayer Olkin* (KMO)
   b. Melakukan uji asumsi non-multikolinieritas, dengan menggunakan nilai *variance inflation factor* (VIF). Apabila terjadi multikolinieritas pada salah satu variabel maka dilakukan analisis komponen utama, skor komponen utama yang diperoleh akan digunakan sebagai input dalam analisis selanjutnya sebagai pengganti nilai data variabel awal.
3. Penentuan banyaknya klaster yang akan dibuat (*k*), nilai *k* yang digunakan dalam penelitian ini yaitu *k* = 3, 4 dan 5
4. Melakukan analisis klaster dengan algoritma *k-medoids*
   a. Tentukan *medoids* iterasi ke *t* dengan memilih sebanyak *k* objek secara acak dari objek yang akan dikelompokan/$C_{(t,k)}$, *(t)* adalah proses iterasi ke *t* = 1, 2, 3, …, N pada kelompok ke (*k)* = 1, 2, 3, …, *k*
   b. Hitung jarak objek *non-medoids* dengan $C_{(t,k)}$ iterasi ke *t* pada tiap klaster dengan perhitungan jarak menggunakan jarak *euclidean* dan jarak *cityblock.*

      Rumus ukuran jarak yang digunakan yaitu :

      - Jarak *euclidean* : $d_{euc}(x_i, c_k) = \sqrt{\sum_{j=1}^{p}(x_{ij} - c_{kj})^2}$
      - Jarak *cityblock* : $d_{cb}(x_i, c_k) = \sum_{j=1}^{p}|x_{ij} - c_{kj}|$

   c. Menentukan anggota dari masing-masing klaster berdasarkan jarak terdekat dengan $C_{(t.k),}$ dengan jarak terdekat = min {$d(x_i,c_1)$, $d(x_i,c_2)$, …, $d(x_i,c_k)$}
   d. Hitung total jarak objek *non-medoids* terdekat dengan $C_{(t,k)}$ pada iterasi ke *t*
   e. Memilih secara acak satu objek *non-medoids* pada masing-masing klaster sebagai kandidat *medoids* baru/$C_{(t+1,k)}$, dengan objek yang sudah pernah menjadi *medoids* tidak boleh dipilih lagi
   f. Hitung jarak objek *non-medoids* dengan $C_{(t+1,k)}$ iterasi ke *t* pada tiap klaster dengan perhitungan jarak menggunakan jarak *euclidean* dan jarak *cityblock*

g. Menentukan anggota dari masing-masing klaster berdasarkan jarak terdekat dengan $C_{(t+1,k)}$ tersebut, dengan jarak terdekat = min{$d(x_i,c_1)$, $d(x_i,c_2)$, …, $d(x_i,c_k)$}
h. Hitung total jarak objek *non-medoids* terdekat dengan kandidat *medoids* baru/$C_{(t+1,k)}$ pada iterasi ke *t*
i. Hitung nilai $S_{\text{total jarak}}$, dimana $S_{\text{total jarak}}$ adalah selisih nilai dari total jarak dengan *medoids* $C_{(t+1,k)}$/kandidat *medoids* baru dan total jarak dengan *medoids* $C_{(t,k)}$ pada iterasi ke *t*
j. Jika nilai $S_{\text{total jarak}} < 0$, maka *medoids* $C_{(t+1,k)}$ tersebut menjadi *medoids* baru pada iterasi berikutnya
k. Mengulangi langkah (e) sampai langkah (j) hingga tidak terjadi perubahan *medoids*. Proses iterasi akan berhenti apabila diperoleh $S_{\text{total jarak}} > 0$ dan pada langkah ini diperoleh klaster beserta anggota masing-masing klaster

5. Validasi hasil pengklasteran
   a. Hitung nilai koefisien *silhouette* dan *C-index* pada masing-masing klaster yang terbentuk dengan jarak *euclidean* dan jarak *cityblock*
   b. Bandingkan nilai koefisien *silhouette* dan *C-index* dari k = 3, 4, dan 5 dengan jarak *euclidean* dan jarak *cityblock*, pengklasteran dikatakan terbaik jika nilai validasi untuk koefisien *silhouette* mendekati angka 1 atau untuk nilai *C-index* mendekati angka nol.
6. Interpretasi dan profilisasi karakteristik wilayah dari hasil pengklasteran terbaik

4. Hasil dan Pembahasan

4.1. Pendeteksian Pencilan

Pada pendeteksian pencilan digunakan metode jarak kuadrat mahalanobis, pengamatan ke- i teridentifikasi pencilan jika $d_{MD}^2(i) > \chi_{p,1-\alpha}^2$. Berdasarkan kasus ini p merupakan banyaknya variabel yang diteliti yaitu 12 variabel dan nilai α yang digunakan sebesar 5%. Maka diperoleh nilai $\chi_{p=12,(1-0.05)}^2$ sebesar 21,02607. Berdasarkan pendeteksian pencilan dengan membandingkan hasil jarak kuadrat mahalanobis tiap objek dan nilai $\chi_{12,0,95}^2$, diketahui bahwa terdapat 6 kabupaten/kota yang merupakan pencilan, yaitu kota Semarang, kota Surakarta, kabupaten Pati, kabupaten Magelang, kabupaten Banyumas, dan kabupaten Cilacap. Pada metode Pada kasus ini data pencilan tetap disertakan dalam analisis berikutnya dengan metode *k-medoids* karena metode tersebut merupakan metode pengelompokan yang *robust* terhadap pencilan atau adanya pencilan tidak akan mempengaruhi hasil.

4.2. Analisis Klaster

a. Representatif

Berdasarkan output pengujian KMO, diperoleh nilai KMO sebesar 0,6973411 dimana nilai tersebut berkisar diantara 0,5 sampai 1. Jadi, dapat disimpulkan bahwa sampel mewakili populasi atau sampel representatif terpenuhi.

b. Tidak Terjadi Multikolinieritas

Berdasarkan hasil pengujian VIF dalam studi kasus ini, diperoleh hasil *output* pada Tabel 1.

Tabel 1. Nilai VIF dari 12 Variabel

Variabel	Nilai VIF	Variabel	Nilai VIF
X_1	2,168486	X_7	6,848555
X_2	1,462609	X_8	5,816459
X_3	7,757228	X_9	2,349945
X_4	2,154906	X_{10}	4,475394
X_5	3,745394	X_{11}	2,995203
X_6	14,392131	X_{12}	2,695189

Menurut Tabel 1 di atas, diketahui 11 variabel dengan nilai VIF≤10, maka variabel tersebut tidak terjadi multikolinieritas. Ada satu variabel dengan nilai VIF>10 yaitu kejahatan pencurian sebesar 14,392131, dapat disimpulkan variabel tersebut terjadi multikolinieritas. Karena masih terdapat variabel yang terjadi multikolinieritas maka asumsi ini belum terpenuhi sehingga perlu dilakukan penanganan multikoliniertas dengan menggunakan analisis komponen utama.

Jumlah komponen utama yang harus dibentuk ditentukan melalui kriteria berdasarkan nilai *eigenvalue*, diperoleh pada Tabel 3

Tabel 2. Nilai *Eigen*, Proporsi Varian dan Proporsi Kumulatif

Komponen Utama	Nilai *Eigen*	Proporsi Varian	Proporsi Kumulatif
1	7486,6530084	0,9010093	0,9010093
2	335,1741081	0,0403378	0,9413471
3	181,9607639	0,02189875	0,9632459
4	120,3525158	0,01448427	0,9777301
5	84,4599759	0,01016465	0,9878948
6	39,6051368	0,00476643	0,9926612
7	31,3292254	0,00377043	0,9964316
8	19,5511484	0,00235296	0,9987846
9	5,1838821	0,00062387	0,9994085
10	2,0973976	0,00025242	0,9996609
11	1,8287202	0,00022008	0,9998809
12	0,9889913	0,00011902	1

Berdasarkan Tabel 2 di atas, dapat diketahui nilai *eigenvalue* masing-masing komponen. Nilai *eigenvalue* tersebut digunakan untuk menentukan jumlah komponen yang akan dipilih dengan ketentuan dimana nilai *eigenvalue* tiap variabel yang lebih dari satu ($\lambda i \geq 1$). Tabel di atas memperlihatkan bahwa ada 11 komponen yang memiliki nilai *eigenvalue* lebih besar dari satu. Berdasarkan ketentuan $\lambda i \geq 1$, maka dipilih 11 komponen utama dengan proporsi komponen masing-masing ditunjukan pada Tabel 3. Adapun proporsi kumulatif dari 11 komponen tersebut sebesar 0,99988, artinya sebesar 99,988% varians dari 12 variabel yang mempengaruhi kriminalitas dapat dijelaskan oleh 11 skor komponen utama.

Berdasarkan hasil analisis pengklasteran kriminalitas dengan metode *k-medoids* dan uji validasi hasil klaster dengan *silhouette index* dan *c-index* untuk $k = 3$, 4, dan 5 dan jarak *euclidean* dan *manhattan*. Maka dilakukan perbandingan nilai validasi untuk mendapatkan pengklasteran terbaik, dapat dilihat pada tabel 3.

Tabel 3. Perbandingan Pengklasteran *k-Medoids*

Banyak Klaster	Jarak	*Silhouette index*	*C–index*
3	*Euclidean*	0,3862593	0,057396
	Manhattan	0,2748682	0,073448
4	*Euclidean*	0,3898083	0,043893
	Manhattan	0,2774014	0,065621
5	*Euclidean*	0,2261416	0,048417
	Manhattan	0,2585274	0,064034
Silhouette index (maksimum)		0,389809	
C – index (minimum)		0,043893	

Berdasarkan tabel di atas, dapat diketahui bahwa pada pengelompokan kabupaten/kota di Jawa Tengah dengan metode *k-medoids* dengan dua validasi, diperoleh pengklasteran terbaik yaitu

- Untuk validasi *silhouette index,* diperoleh nilai validasi maksimum pada *k*= 4 menggunakan jarak *euclidean* yaitu sebesar 0,3862593.
- Untuk meyakinkan hasil validasi *silhouette index,* digunakan validasi *C-index* diperoleh nilai validasi minimum pada *k*= 4 klaster dengan jarak *euclidean* yaitu sebesar 0,043893.

Oleh karena itu, berdasarkan pengklasteran dengan dua validasi *silhouette index* dan *c-index* dapat disimpulkan bahwa hasil pengklasteran terbaik dengan kedua validasi tersebut berdasarkan data jumlah kriminalitas di Jawa tengah tahun 2018 dengan metode *k-medoids* yaitu pengklasteran pada *k*= 4 dan pengukuran jarak *euclidean*, karena memiliki nilai indeks validasi yang paling optimum pada masing-masing validasi.

Profilisasi hasil analisis klaster dilakukan pada hasil pengklasteran terbaik, yaitu pengklasteran kabupaten/kota di Jawa Tengah berdasarkan data jumlah kriminalitas di Jawa tengah tahun 2018 ke dalam 4 klaster dengan metode *k-medoids* pengukuran jarak *euclidean.* Pada tahap profilisasi akan dilihat karakteristik dari tiap klaster yang terbentuk, sehingga dapat dilihat kecenderungan dari tiap klaster. Pada metode *k-medoids*, karakteristik dari klaster-klaster yang terbentuk dapat direpresentasikan dengan nilai rata-rata dari setiap variabel. Variabel yang digunakan adalah variabel awal, yaitu variabel yang bukan merupakan hasil dari analisis komponen utama. Hasil perhitungan nilai *centroids*/rata-rata tiap variabel pada setiap klaster ditunjukkan pada Tabel 4.

Tabel 4. Nilai *Centroid*/Rata-rata Setiap Variabel pada Setiap Klaster

Kota	**Klaster 1**	**Klaster 2**	**Klaster 3**	**Klaster 4**
x1	7,000	1,250	0,632	0,667
x2	11,000	3,500	3,474	5,667
x3	42,000	14,500	6,684	43,333
x4	7,000	18,333	12,789	40,000
x5	46,000	8,167	5,421	14,000
x6	422,000	130,833	63,789	259,000
x7	132,000	18,417	10,947	43,667
x8	59,000	17,250	11,000	50,667
x9	12,000	9,667	9,684	33,000
x10	14,000	2,833	1,474	2,667
x11	11,000	2,833	2,421	10,000
x12	3,000	1,500	1,105	6,333

Berdasarkan Tabel 4 di atas diperoleh informasi sebagai berikut:

a. Klaster 1

Anggota klaster satu terdiri dari kota Semarang. Klaster ini memiliki jenis kejahatan dengan *centroid*/rata-rata paling tinggi di antara klaster lain, dengan jenis kejahatan yang menonjol yaitu kejahatan pembunuhan, kejahatan kesusilaan, kejahatan pencurian, kejahatan penggelapan, kejahatan penipuan, kejahatan pemerasan, dan kejahatan KDRT. Akan tetapi dalam klaster ini kejahatan perjudian memiliki *centroid*/rata-rata paling rendah diantara klaster lain.

b. Klaster 2

Anggota klaster satu terdiri dari kab. Sragen, kab. Semarang, kab. Boyolali, kab. Klaten, kab. Cilacap, kota Salatiga, kab. Pati, kab. Sukoharjo, kab. Kebumen, kab. Jepara, kab. Tegal, dan kab. Kudus. Klaster ini tidak ada kejahatan yang paling menonjol, karena pada klaster ini jenis kejahatan yang terjadi didominasi dengan jenis kejahatan dengan *centroid*/rata-rata cukup rendah diantara klaster lain. Akan tetapi dalam klaster ini kejahatan pembunuhan, kejahatan perjudian, dan kejahatan pemerasan memiliki *centroid*/rata-rata yang cukup tinggi di bawah *centroid*/rata-rata klaster 1.

c. Klaster 3

Anggota klaster satu terdiri dari kab. Kendal, kota Magelang, kab. Temanggung, kab. Blora, kota Pekalongan, kab. Wonosobo, kab. Wonogiri, kab. Rembang, kab. Banjarnegara, kab. Brebes, kab. Demak, kab. Grobogan, kab. Purworejo, Kab. Pemalang, kota Tegal, kab. Pekalongan, kab. Karanganyar, kab. Purbalingga, dan kab. Batang. Klaster ini tidak ada kejahatan yang paling menonjol, karena pada klaster ini jenis kejahatan yang terjadi didominasi dengan jenis kejahatan dengan *centroid*/rata-rata paling rendah diantara klaster lain.

d. Klaster 4

Anggota klaster satu terdiri dari kota Surakarta, kab. Banyumas, dan kab. Magelang. Pada klaster ini memiliki jenis kejahatan dengan *centroid*/rata-rata paling tinggi diantara klaster lain dengan jenis kejahatan yang menonjol yaitu kejahatan penganiayaan, kejahatan perjudian, kejahatan terhadap perlindungan anak, dan kejahatan penghancuran dan perusakan barang.

5. Penutup

5.1. Kesimpulan

1. Pengklasteran menggunakan metode *k-medoids* dengan jarak *euclidean* dan *manhattan* untuk k = 3, 4, dan 5 diperoleh klaster yang optimal pada k = 4 dengan jarak *euclidean* dimana nilai SI = 0,3862593 dan CI = 0,043893. Berdasarkan hasil pengklasteran pada metode ini didapatkan bahwa jarak pengukuran yang digunakan akan berpengaruh terhadap hasil pengklasteran.
2. Berdasarkan profilisasi hasil analisis klaster, diketahui bahwa klaster 1 jenis kejahatan yang menonjol yaitu kejahatan pembunuhan, kesusilaan, pencurian, penggelapan, penipuan, pemerasan, dan kejahatan KDRT. Untuk klaster 4 jenis kejahatan yang menonjol yaitu kejahatan penganiayaan, perjudian, penghancuran dan perusakan barang, dan kejahatan terhadap perlindungan anak. Sedangkan untuk klaster 3 dan klaster 4 didominasi jenis kejahatan dengan *centroid*/rata-rata rendah diantara klaster lain.

5.2. Saran

1. Penelitian selanjutnya dapat dilakukan pengklasteran dengan metode *k-medoids* dan *k-means* untuk mengetahui pengklasteran mana yang lebih *robust* terhadap data yang

mengandung pencilan. Metode *k-medoids* menggunakan *medoids* sebagai pusat klasternya dan *k-means* menggunakan *mean* sebagai pusat klasternya.

2. Karena masih terdapat daerah-daerah dengan kejahatan yang cukup tinggi, maka dari itu harus adanya kerjasama antara penegak hukum dan masyarakat dalam meminimalisir terjadinya kriminalitas, mobilitas penduduk lebih diperketat dengan dilakukannya pendataan/pengecekan data penduduk dari luar daerah yang menetap di daerah tersebut dan lebih ditingkatkan lagi pos kampling di setiap daerah.

DAFTAR PUSTAKA

Abdulsyani. 1987. *Sosiologi Kriminalitas*. Bandung: Remadja Karya CV.

Aini, F.N., Palgunadi. S., dan Anggrainingsih. R. 2014. *Clustering Business Rocess Model Petri Net dengan Complete Linkage*. Jurnal ITSMART Vol 3. No. 2: Hal. 47-51.

Charrad, M., Ghazzali, N., Boiteau, V., dan Niknafs, A. 2014. *NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set*. Journal of Statistical Software Vol 61. No. 6: Hal. 1-36.

Folzmiser, P. 2005. *Identification of Multivariate Outliers: A Performance Study*. Australian Journal of Statistics Vol 34. No. 2: Hal. 127-138.

Gujarati, D. 2009. *Dasar-dasar Ekonometrika Jilid 2*. Jakarta: Erlangga.

Hair, J.F., Anderson, R.E., Thatham, R.L., dan Black, W.C. 2010. *Multivariate Data Analysis Seventh Edition*. New Jersey: Pearson Education. Inc.

Han, J. dan Kamber, M. 2006. Data Mining Concepts and Techniques Second Edition. San Francisco: Elsever.

Johnson, R.A. dan Wichern, D.W. 2007. *Applied Multivariate Scatistical Analysis Six Edition*. New Jersey: Prentice Hill. Inc.

Polda Jawa Tengah. 2018. *Data Kriminalitas Tiap Kabupaten/Kota di Jawa Tengah Tahun 2018*. Semarang: Dir. Reserse Kriminal Umum.

Supranto, J. 2004. *Analisis Multivariat : Arti dan Interpretasi*. Jakarta: PT. Rineka Cipta.

Widarjono, A. 2010. *Analisis Statistika Multivariat Terapan. Edisi Pertama*. Yogyakarta: UPP STIM YKPN.