

## PERBANDINGAN METODE *K-MEANS* DAN *SELF ORGANIZING MAP* (STUDI KASUS: PENGELOMPOKAN KABUPATEN/KOTA DI JAWA TENGAH BERDASARKAN INDIKATOR INDEKS PEMBANGUNAN MANUSIA 2015)

Rachmah Dewi Kusumah<sup>1</sup>, Budi Warsito<sup>2</sup>, Moch. Abdul Mukid<sup>3</sup>

<sup>1</sup>Mahasiswa Departemen Statistika FSM Universitas Diponegoro

<sup>2,3</sup>Staff Pengajar Departemen Statistika FSM Universitas Diponegoro

### ABSTRACT

Cluster analysis is a process of separating the objects into groups, so that the objects that belong to the same group are similar to each other and different from the other objects in another group. In this study used two method to classify data of district / city in Central Java based on indicators of Human Development Index (HDI) 2015 are *K-Means* and *Self Organizing Map* (SOM) with the number of groups as much as two to seven. Furthermore, the results of both methods were compared using the *Davies-Bouldin* Index (DBI) values to determine which method is better. Based on the research that has been conducted found that the *K-Means* (K=4) method works better than SOM (K=2) to classify district / city in Central Java based on indicators of Human Development Index (HDI) as evidenced by the value of the *Davies-Bouldin* Index (DBI) on *K-Means* (K=4) of 0.786 is smaller than the value at SOM (K=2) *Davies-Bouldin* Index (DBI) which is equal to 0.893.

**Keywords:** clustering, HDI, *K-Means*, SOM, DBI

### 1. PENDAHULUAN

Manusia adalah kekayaan bangsa yang sesungguhnya sehingga tujuan akhir pembangunan harus difokuskan pada manusia. Kondisi ini akan menciptakan lingkungan yang memungkinkan masyarakat untuk dapat menikmati umur panjang, sehat, dan menjalankan kehidupan yang produktif. Konsep ini menjadi cikal bakal munculnya Indeks Pembangunan Manusia (IPM)<sup>[1]</sup>.

Pengelompokan wilayah kabupaten/kota di Jawa Tengah perlu dilakukan sebagai bahan perencanaan dan evaluasi sasaran program pemerintah untuk meningkatkan angka pembangunan manusia berdasarkan indikator pembentuk IPM. Algoritma pengelompokan yang dapat digunakan diantaranya adalah *K-Means* dan SOM.

*K-Means* adalah pengelompokan data non hirarki (sekatan) yang mempartisi data ke dalam bentuk dua atau lebih kelompok sehingga data yang berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama. *Self Organizing Map* (SOM) adalah salah satu bentuk topologi *Unsupervised Artificial Neural Network* (*Unsupervised ANN*) dimana dalam proses pembelajaran tidak memerlukan pengawasan (*target output*)<sup>[10]</sup>.

### 2. TINJAUAN PUSTAKA

Indeks Pembangunan Manusia (IPM) merupakan tolok ukur keberhasilan pembangunan manusia yang mencakup tiga dimensi dasar kehidupan manusia, yaitu umur panjang dan sehat (diukur dengan indikator Angka Harapan Hidup dalam tahun), pengetahuan (diukur dengan indikator Harapan Lama Sekolah dalam tahun dan Rata-rata

Lama Sekolah dalam tahun), dan standar hidup layak (diukur dengan indikator Pengeluaran Per Kapita dalam ribuan rupiah)<sup>[1]</sup>.

Indikator indeks pembangunan manusia ada empat, yaitu<sup>[1]</sup>:

1. Angka Harapan Hidup (AHH) merupakan rata-rata perkiraan banyak tahun yang dapat ditempuh seseorang selama hidupnya.
2. Angka Harapan Lama Sekolah (HLS) adalah lamanya sekolah (dalam tahun) yang dihitung untuk penduduk berusia 7 tahun ke atas.
3. Rata-rata Lama Sekolah (RLS) adalah lama sekolah (tahun) penduduk usia 25 tahun ke atas.
4. Kemampuan Daya Beli Penduduk atau *Purchasing Power Parity* (PPP) mencerminkan kemampuan masyarakat secara ekonomi dalam memenuhi kebutuhan konsumsinya.

Data mining adalah proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Data mining merupakan teknologi yang menggabungkan metode analisis tradisional dengan algoritma yang canggih untuk memproses data dengan volume besar<sup>[10]</sup>. Data mining dapat digunakan untuk menyelesaikan masalah dalam bidang intelektual, ekonomi, dan bisnis dibagi menjadi enam tugas yaitu klasifikasi, estimasi, prediksi, afinitasi, *clustering*, deskripsi dan penentuan profil<sup>[2]</sup>.

Kumpulan data yang akan diproses dengan metode-metode dalam data mining sering kali harus melalui pekerjaan awal dikarenakan memiliki masalah seperti populasi data yang terlalu besar, dimensi data yang terlalu tinggi, banyaknya fitur yang tidak berkontribusi besar, dan adanya perbedaan skala pada variabel<sup>[8]</sup>. Data *transformation* (transformasi data) adalah pekerjaan mengubah data ke dalam bentuk yang paling tepat atau cocok untuk proses data mining, untuk masalah data yang variabelnya memiliki skala yang berbeda dapat dilakukan normalisasi. Normalisasi adalah mengubah data sebuah variabel diskalakan ke dalam rentang (kecil) yang ditentukan. Salah satu metode atau teknik yang diterapkan untuk normalisasi data, misalnya Z-Score dan dinyatakan dalam bentuk persamaan (1)<sup>[5]</sup>.

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (1)$$

Analisis Kelompok (*cluster analysis*) adalah pekerjaan mengelompokkan data (objek) yang didasarkan hanya pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan diantaranya<sup>[10]</sup>. Dalam pengelompokan data, data yang akan dikelompokkan tidak mempunyai label kelas, tetapi dikelompokkan menurut karakteristiknya, barulah kelompok tersebut diberi label sesuai hasil karakteristik kelompok masing-masing. Karena alasan tersebut analisis kelompok sering disebut *segmentation* atau *partitioning*<sup>[8]</sup>.

Tujuan pengelompokan (*clustering*) data dapat dibedakan menjadi dua, yaitu pengelompokan untuk pemahaman dan pengelompokan untuk penggunaan. Jika tujuannya untuk pemahaman, kelompok yang terbentuk harus menangkap struktur alami data, biasanya proses pengelompokan dalam tujuan ini hanya sebagai proses awal untuk kemudian dilanjutkan dengan pekerjaan inti. Sementara jika untuk penggunaan, tujuan utama pengelompokan biasanya adalah mencari prototipe kelompok dimana sebuah data terletak di dalamnya<sup>[8]</sup>.

Asumsi klaster tidak difokuskan pada kualitas data, seperti normalitas dan linearitas. Akan tetapi, difokuskan pada isu-isu desain penelitian. Asumsi dalam analisis klaster tersebut sebagai berikut<sup>[4]</sup>:

- a. Sampel Mewakili Populasi

b. Tidak Terjadi Multikolinearitas

Multikolinearitas adalah adanya hubungan linear yang sempurna atau pasti di antara beberapa atau semua variabel<sup>[3]</sup>. Untuk mengetahui adanya multikolinearitas salah satunya adalah dengan menghitung nilai *Variance Inflation Factor* (VIF) dengan rumus:

$$VIF_j = \frac{1}{1-R_j^2} \quad (2)$$

Dengan  $R_j^2$  adalah nilai koefisien determinasi yang sering diartikan sebagai seberapa besar kemampuan semua variabel independen ke- $j$  dalam menjelaskan varian dari variabel dependennya.

Dalam analisis kluster terdapat banyak metode yang dapat digunakan untuk mengelompokkan sekumpulan objek ke dalam kluster. Secara umum terdapat dua metode dasar yang biasa digunakan, yaitu Metode Hirarki dan Metode Non Hirarki. Metode hirarki terdiri dari dua metode pengelompokan, yaitu metode *Divisive* dan *Agglomerative*. Sedangkan metode non hirarki terdiri dari tiga metode, yaitu *Sequential Threshold Procedure*, *Parallel Threshold Procedure*, dan *Optimizing Partitioning*<sup>[9]</sup>.

### 2.1 Metode K-Means

*K-Means* merupakan salah satu metode pengelompokan data nonhierarki (sekatan) yang berusaha membagi data ke dalam bentuk dua atau lebih kelompok. Metode ini membagi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama dan data yang berbeda karakteristik dikelompokkan ke dalam kelompok yang lain<sup>[8]</sup>.

Pengelompokan data dengan metode *K-Means* secara umum dilakukan dengan algoritma sebagai berikut<sup>[6]</sup>:

1. Menentukan banyaknya kelompok ( $k$ ).
2. Membagi data ke dalam  $k$  kelompok.
3. Menghitung pusat kelompok (sentroid atau rata-rata) dari data yang ada di masing-masing kelompok.

Lokasi sentroid (titik pusat) setiap kelompok dapat dicari dengan rumus:

$$C_i = \frac{1}{M} \sum_{j=1}^M x_j$$

dengan:

$C_i$  = Nilai sentroid pada setiap kelompok.

$i$  = banyaknya kelompok.  $i = 1, 2, \dots, k$

$M$  = banyaknya data dalam sebuah kelompok

4. Masing-masing data dialokasikan ke sentroid atau rata-rata terdekat.
5. Menghitung jarak data ke setiap sentroid kemudian dipilih jarak yang paling minimal untuk menentukan kelompok berikutnya. Jarak yang digunakan adalah jarak Euclidean yaitu:

$$D(x_i, C_i) = \sqrt{\sum_{j=1}^q (x_{ij} - C_{ij})^2} \quad (3)$$

dengan:

$x_i$  = data yang akan dihitung jaraknya

$C_i$  = Nilai sentroid pada kelompok ke  $i$ , dimana  $i = 1, 2, \dots, k$

$x_{ij}$  = data pada masing-masing dimensi

$C_{ij}$  = sentroid pada masing-masing dimensi

$q$  = banyaknya dimensi data

maka pada masing-masing data diperoleh sebanyak  $k$  jarak. Kemudian dipilih jarak yang paling minimal untuk menentukan kelompok berikutnya.

6. Kembali ke langkah 3, apabila masih ada data yang berpindah kelompok.

Karakteristik *K-Means* dapat diringkas sebagai berikut<sup>[8]</sup>:

1. *K-Means* merupakan metode pengelompokan yang sederhana dan dapat digunakan dengan mudah.
2. Pada jenis kumpulan data tertentu, *K-Means* tidak dapat melakukan segmentasi data dengan baik di mana hasil segmentasinya tidak dapat memberikan pola kelompok yang mewakili karakteristik bentuk alami data.

## 2.2 Metode *Self Organizing Map*

*Self Organizing Map* (SOM) pertama kali diperkenalkan dengan teknik pelatihan ANN yang menggunakan basis *winner takes all*, di mana hanya neuron yang menjadi pemenang yang akan diperbarui bobotnya. Meskipun menggunakan basis ANN, SOM tidak menggunakan nilai target kelas, tidak ada kelas yang ditetapkan untuk setiap data. Karakteristik seperti inilah yang kemudian membuat SOM dapat digunakan untuk keperluan pengelompokan (berbasis ANN)<sup>[8]</sup>.

Pengelompokan data dengan metode SOM secara umum dilakukan dengan algoritma sebagai berikut<sup>[8]</sup>:

1. Tetapkan:
  - a. Banyaknya variabel :  $m$
  - b. Banyaknya data :  $n$
  - c. Banyaknya cluster [d1 d2] :  $j \times i$
  - d. Jenis topologi.
2. Inisialisasi:
  - a. Bobot input ( $w_{ij}$ ).

$$w_{ij} = \frac{\text{Min}X_i + \text{Max}X_i}{2} \quad (4)$$

Dengan  $w_{ij}$  adalah bobot antara variabel input ke- $i$  dengan neuron pada lapisan output ke- $j \times i$  ( $i=1, 2, 3, \dots, m; j \times i=1, 1, 2, 1, 3, \dots, d1, d2$ );  $\text{Min} X_i$  dan  $\text{Max} X_i$  masing-masing adalah nilai terkecil pada variabel input ke- $i$ , dan nilai terbesar pada variabel input ke- $i$ .

- b. Set parameter adaptasi pembelajaran (*learning rate*).
  - c. Set maksimum epoch (Max epoh).
3. Set Epoch = 0.
4. Kerjakan jika Epoch < Max Epoch.
  - a. Epoch = Epoch + 1.
  - b. Sebuah vektor dipilih secara *random*.
  - c. Perhitungan jarak antar vektor input terhadap bobot ( $D_j$ ) dengan masing-masing neuron output menggunakan rumus jarak *Euclidean*.

$$D_j = \sqrt{\sum_{i=1}^m (w_{ij} - x_i)^2}$$

Keterangan:

- $w_{ij}$  : bobot koneksi antara neuron input dan output  
 $x_i$  : neuron pada lapisan input ke- $i$   
 $m$  : banyaknya variabel

- d. Dari seluruh bobot ( $D_j$ ) dicari yang paling kecil. *Index* dari bobot ( $D_j$ ) dikatakan *winning neuron* (neuron pemenang).

- e. Untuk setiap bobot  $w_{ij}$  diperbarui bobot koneksinya.

$$w_{ij}(\text{baru}) = w_{ij}(\text{lama}) + \eta(x_i - w_{ij}(\text{lama}))$$

Keterangan:

$\eta$  : adaptasi pembelajaran (*learning rate*)

### 2.3 Validasi Kelompok

Jika proses pengklasteran untuk masing-masing jumlah kluster selesai, maka untuk menentukan jumlah kluster yang paling optimal dapat dilakukan penilaian menggunakan *Davies-Bouldin Index* (DBI). Pengklasteran dengan jumlah kluster yang optimal adalah pengklasteran yang memiliki nilai DBI minimum<sup>[7]</sup>.

$$DBI = \frac{1}{k} \sum_{i=1}^k R_i \quad (5)$$

dengan:

$$R_i = \max_{j=1, \dots, k, i \neq j} R_{ij}, \quad R_{ij} = \frac{s_i + s_j}{d_{ij}}$$

$$d_{ij} = d(v_i + v_j)$$

dan

$$s_i = \left[ \frac{1}{n_i} \sum_{l=1}^{n_i} d^2(x_{i,l}, v_i) \right]^{\frac{1}{2}}$$

dimana:

- $k$  = Jumlah kluster
- $R_{ij}$  = Ukuran kemiripan antara  $n_i$  dan  $n_j$
- $s_i$  = Ukuran dispersi kluster ke- $i$ ,  $i = 1, 2, \dots, k$
- $d_{ij}$  = Jarak antara *centroid* kluster ke- $i$  dan *centroid* kluster ke- $j$
- $n_i$  = Banyaknya anggota kluster ke- $i$ ,  $i = 1, 2, \dots, k$
- $v_i$  = *Centroid* kluster dari  $n_i$
- $x_{i,l}$  = Data pada kluster tersebut

## 3. METODE PENELITIAN

Data yang digunakan sebagai studi kasus pada penelitian ini berupa data sekunder yang berasal dari BPS Provinsi Jateng. Variabel yang digunakan dalam pengelompokan adalah indikator-indikator Indeks Pembangunan Manusia (IPM) kabupaten/kota di Provinsi Jateng tahun 2015. Variabel-variabel tersebut adalah:

- $X_1$  = Angka Harapan Hidup saat Lahir (AHH)
- $X_2$  = Harapan Lama Sekolah (HLS)
- $X_3$  = Rata-rata Lama Sekolah (RLS)
- $X_4$  = Pengeluaran per Kapita yang disesuaikan (PPP)

Program komputer yang digunakan untuk mendukung proses penelitian adalah *software Microsoft Excel 2013*, *SPSS*, dan *Matlab R2014a*. Adapun metode analisis yang digunakan untuk mencapai tujuan penelitian ini diuraikan sebagai berikut:

1. Memilih variabel dan melakukan *Preprocessing Data* (Standardisasi data menggunakan *Z-Score*).  
Hal paling penting di dalam perumusan masalah analisis kluster adalah pemilihan variabel yang akan digunakan pada proses pengklasteran (pembentukan kluster).
2. Melakukan uji asumsi analisis *cluster*, yaitu uji asumsi multikolinearitas. Jika asumsi multikolinearitas tidak terpenuhi dilakukan tindakan perbaikan multikolinearitas, jika tidak terjadi maka dapat dilakukan proses selanjutnya.

3. Melakukan pengelompokan terhadap data hasil Standardisasi dengan metode *K-Means* dan *Self Organizing Map* (SOM).
4. Melakukan pengelompokan kabupaten/kota di Jateng berdasarkan indikator IPM tahun 2015 yang telah melalui Standardisasi menggunakan metode *K-Means* dengan cara sebagai berikut:
  - a. Menentukan banyaknya klaster yang akan dibuat ( $k$ ).
  - b. Membagi data menjadi  $k$  kelompok.
  - c. Menghitung sentroid (rata-rata) kelompok dari data masing-masing kelompok.
  - d. Menghitung jarak data ke semua sentroid (rata-rata) kelompok menggunakan rumus jarak *Euclidean*.
  - e. Data dialokasikan kembali ke kelompok dengan sentroid yang memiliki jarak terdekat dengan data tersebut.
  - f. Melakukan kembali langkah b hingga d apabila masih ada data yang berpindah kelompok.
5. Melakukan pengelompokan data kabupaten/kota di Jateng berdasarkan indikator IPM tahun 2015 yang telah melalui Standardisasi menggunakan metode *Self Organizing Map* (SOM) dengan cara berikut:
  - a. Menginput data.
 

Banyaknya variabel	: m
Banyaknya data	: n
Banyaknya cluster [d1 d2]	
Fungsi topologi	: Gridtop
Learning rate	: $\eta=0,5$
Maksimum iterasi (epoch)	: 100
  - b. Menghitung bobot ( $w_{ij}$ ).
  - c. Menghitung jarak untuk setiap neuron  $j$ .
  - d. Mencari nilai terkecil dari indeks sejumlah neuron ( $D_j$ ).
  - e. Memperbarui bobot untuk neuron  $j$  dan semua neuron yang menjadi tetangga  $j$ .
  - f. Memperbarui nilai laju pembelajaran.
  - g. Melakukan kembali langkah b hingga f apabila neuron-neuron *output* belum berada pada posisis stabil (konvergen).
6. Melakukan pengujian menggunakan uji *Davies-Bouldin Index* (DBI) pada beberapa kelompok yang telah dibentuk menggunakan metode *K-Means* dan *Self Organizing Map* (SOM).
7. Melakukan analisis dari *output* yang dihasilkan.

#### 4. HASIL DAN PEMBAHASAN

Pada penelitian ini, teknik *preprocessing* yang digunakan yaitu transformasi data. Transformasi bertujuan mendapatkan data baru dalam ukuran yang standar. Pada proses pengelompokan ini dilakukan transformasi dengan cara normalisasi *Z-Score*.

Selanjutnya data yang digunakan diperiksa apakah mengalami multikolinearitas atau tidak dengan cara menghitung nilai VIF. Nilai VIF dapat diketahui melalui persamaan (2) dengan  $R_i^2$  merupakan koefisien determinasi dan indeks  $i$  menyatakan variabel bebas ke  $i$ .

**Tabel 1.** Nilai VIF dari Empat Variabel Indikator IPM 2015

Variabel	R <sup>2</sup>	VIF
AHH (tahun)	0,479	1,919
MYS (tahun)	0,809	5,236
EYS (tahun)	0,801	5,025
PPP (ribu Rupiah)	0,670	3,030

**Tabel 1.** menunjukkan bahwa dari keempat variabel penelitian tersebut tidak terdapat nilai VIF dari variabel-variabel tersebut yang memiliki nilai VIF lebih dari 10. Oleh karena itu, dapat disimpulkan tidak terdapat pengaruh multikolinieritas antar variabel bebas.

Penelitian ini bertujuan untuk membandingkan kelompok mana yang paling optimal diantara metode *K-Means* dan SOM. Hasil pengelompokan data indikator Indeks Pembangunan Manusia (IPM) Provinsi Jawa Tengah pada tahun 2015 menggunakan metode *K-Means* dengan jumlah kelompok dua sampai dengan tujuh tersaji dalam **Tabel 2.**

**Tabel 2.** Hasil Pengelompokan *K-Means* untuk K=2 sampai dengan K=7

Kelompok	1	2	3	4	5	6	7
Jumlah Data untuk K=2	24	11	-	-	-	-	-
Jumlah Data untuk K=3	17	14	4	-	-	-	-
Jumlah Data untuk K=4	12	12	8	3	-	-	-
Jumlah Data untuk K=5	7	15	5	5	3	-	-
Jumlah Data untuk K=6	3	9	10	5	5	3	-
Jumlah Data untuk K=7	3	9	10	5	4	1	3

Hasil pengelompokan data indikator Indeks Pembangunan Manusia (IPM) Provinsi Jawa Tengah pada tahun 2015 menggunakan metode SOM dengan jumlah kelompok dua sampai tujuh tersaji dalam **Tabel 3.**

**Tabel 3.** Hasil Pengelompokan SOM untuk K=2 sampai dengan K=7

Kelompok	1	2	3	4	5	6	7
Jumlah Data untuk K=2	21	14	-	-	-	-	-
Jumlah Data untuk K=3	8	10	17	-	-	-	-
Jumlah Data untuk K=4	8	7	5	15	-	-	-
Jumlah Data untuk K=5	4	7	9	8	7	-	-
Jumlah Data untuk K=6	4	4	8	5	7	7	-
Jumlah Data untuk K=7	3	4	5	6	5	8	4

Proses selanjutnya yaitu menentukan jumlah kluster yang paling optimal dengan menggunakan *Davies-Bouldin Index* (DBI). Pada pengamatan DBI untuk mengetahui *clustering* terbaik adalah dengan mengetahui nilai DBI terkecil dari pilihan yang ada. Dengan menggunakan persamaan (5), diperoleh hasil perhitungan nilai DBI untuk *K-Means* dan SOM menggunakan program Matlab pada **Tabel 4**.

**Tabel 4.** Nilai DBI untuk Masing-masing *Cluster K-Means*

<i>Cluster Ke-</i>	Nilai DBI untuk <i>K-Means</i>	Nilai DBI untuk SOM
2	0,850	0,893
3	0,876	1,062
4	0,786	1,110
5	0,860	1,267
6	0,998	1,125
7	0,896	1,267
<b>Nilai DBI Terkecil</b>	<b>0,786 (K=4)</b>	<b>0,893 (K=2)</b>

Berdasarkan **Tabel 4**, nilai DBI untuk *K-Means* terkecil terletak pada *cluster K-Means* ke-4 (K=4) dengan nilai DBI sebesar 0,786, sedangkan nilai DBI untuk SOM terkecil terletak pada *cluster SOM* ke-2 (K=2) dengan nilai DBI sebesar 0,893. Karena nilai DBI pada *cluster K-Means* (K=4) lebih kecil dibandingkan pada *cluster SOM* (K=2), maka diketahui bahwa pengelompokan menggunakan metode *K-Means* lebih baik dibandingkan dengan metode SOM. Berdasarkan kelompok yang paling optimal kemudian dilakukan interpretasi *clustering* terhadap data indikator Indeks Pembangunan Manusia (IPM) kabupaten/kota di Provinsi Jateng tahun 2015 yang dapat dilihat pada **Tabel 5**.

**Tabel 5.** Hasil Pengelompokan Optimal untuk *Cluster* dengan Metode *K-Means* (K=4)

Kelompok	Jumlah Anggota	Anggota Kelompok
1	12	Kab. Cilacap, Kab. Purbalingga, Kab. Banjarnegara, Kab. Kebumen, Kab. Wonosobo, Kab. Magelang, Kab. Blora, Kab. Batang, Kab. Pekalongan, Kab. Pemalang, Kab. Tegal, dan Kab. Brebes
2	12	Kab. Banyumas, Kab. Purworejo, Kab. Boyolali, Kab. Wonogiri, Kab. Sragen, Kab. Grobogan, Kab. Rembang, Kab. Pati, Kab. Jepara, Kab. Demak, Kab. Temanggung, dan Kab. Kendal
3	8	Kab. Klaten, Kab. Sukoharjo, Kab. Karanganyar, Kab. Kudus, Kab. Semarang, Kota Magelang, Kota Pekalongan, dan Kota Tegal
4	3	Kota Surakarta, Kota Salatiga, dan Kota Semarang

## 5. KESIMPULAN

Kesimpulan yang diperoleh dari penelitian ini adalah:

1. Berdasarkan nilai DBI untuk pengelompokan menggunakan metode *K-Means* dan SOM dengan banyaknya kelompok masing-masing sebanyak enam (K=2, 3, 4, 5, 6, dan 7), didapat bahwa pengelompokan menggunakan *K-Means* ke-4 (K=4) adalah metode yang paling optimal. Hal ini diketahui karena memiliki nilai DBI lebih kecil yaitu sebesar 0,786 dibanding nilai DBI pada SOM ke-2 (K=2) sebesar 0,893.

2. Hasil pengelompokan menggunakan metode *K-Means* (K=4) menunjukkan bahwa terdapat 12 kabupaten/kota pada kelompok 1, 12 kabupaten/kota pada kelompok 2, 8 kabupaten/kota pada pada kelompok 3, dan 3 kabupaten/kota pada pada kelompok 4 untuk penerapan *clustering* data indikator Indeks Pembangunan Manusia (IPM) Provinsi Jawa Tengah pada tahun 2015.

#### DAFTAR PUSTAKA

- [1] Badan Pusat Statistik. 2016. *Indeks Pembangunan Manusia Provinsi Jawa Tengah dan Kabupaten/Kota se-Jawa Tengah*. Semarang.
- [2] Berry, M.J.A. & Linoff, G.S. 2004. *Data Mining Techniques For Marketing, Sales, and Customer Relationship Management Second Edition*. Indianapolis, Indiana: Wiley Publishing, Inc.
- [3] Gujarati, D. 1978. *Basic Econometrica*. McGraw-Hill. New York
- [4] Hair, J.F. JR., Anderson, R.E, Tatham, R.L & Black W.C. 2006. *Multivariate Data Analysis 6th Edition*. New Jersey: Pearson Prentice Hill. Inc.
- [5] Han, J., Kamber, M. & Pei, J. 2006. *DATA MINING Concepts and Techniques*. USA: Morgan Kaufmann Publisher.
- [6] Jhonson, R.A. & Wichern, D.W. 2007. *Appyislied Multivariate Statitital Analysis Sixth Edition*. USA: Pearson Education Inc.
- [7] Permatadevi, M. A., Hendrawan, R. A., dan Hafidz, I. 2013. *Karakteristik Pelanggan Telepon Kabel Menggunakan Clustering SOM dan K-Means untuk Mengurangi kesalahan Klasifikasi Pelanggan perusahaan Telekomunikasi*. Jurnal Teknik Pomits Vol. 1, No. 1 : Hal. 1-6. <http://digilib.its.ac.id/public/ITS-paper-34583-5209100025-Paper.pdf>. (23 November 2016)
- [8] Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi menggunakan MATLAB*. Yogyakarta: Andi.
- [9] Supranto, J. 2004. *Analisis Multivariat: Arti dan Interpretasi*. Jakarta: PT. Rineka Cipta.
- [10] Tan, P.N., Steinbach, M. & Kumar, V. 2006. *Introduction to Data Mining*. New York: Pearson Addison Wesley.