

# Penerapan Algoritma K-Medoids dan K-Means untuk Pemetaan Penyebaran Guru Tingkat SMP seluruh Kabupaten/Kota di Indonesia

Lilis Kartika  
Universitas Buana Perjuangan  
Karawang, Indonesia  
if16.liliskartika @mhs.ubpkarawang.ac.id

Amril Mutoi Siregar  
Universitas Buana Perjuangan  
Karawang, Indonesia  
amrilmutoi@ubpkarawang.ac.id

Dwi Sulistya Kusumaningrum  
Universitas Buana Perjuangan  
Karawang, Indonesia  
dwi.sulistya@ubpkarawang.ac.id

## Abstract—

Guru memiliki peran penting dalam pendidikan, jasanya mampu menciptakan generasi-generasi yang berkualitas, baik secara intelektual dan akhlaknya. Tenaga pengajar Indonesia belum tersebar dengan baik di pelosok negeri, sesuai dengan Roadmap Pengelolaan Aparatur Sipil Negara (ASN) dan Perencanaan Formasi tahun 2014. Oleh karena itu, perlu dilakukan pemetaan persebaran tenaga pengajar di berbagai wilayah dan kota di Indonesia. Dalam riset ini pengelompokan data memakai Algoritma K-Medoids dan K-Means dengan dataset yaitu jumlah guru, jumlah peserta didik dan jumlah sekolah jenjang SMP. Algoritma K-Medoids menghasilkan *cluster* 1 yang memiliki kekurangan guru sebanyak 302 Kabupaten/Kota, pada *cluster* 2 yang memiliki kelebihan guru sebanyak 77 Kab/Kota, sedangkan *cluster* 3 yang memiliki cukup guru sebanyak 135 Kab/Kota. Sedangkan Algoritma K-Means menghasilkan *cluster* 1 yang memiliki kekurangan guru sebanyak 363 Kabupaten/Kota, *cluster* 2 yang memiliki cukup guru sebanyak 125 Kabupaten/Kota sedangkan *cluster* 3 yang memiliki kelebihan guru sebanyak 26 Kabupaten/Kota. Manfaat dari penelitian ini sebagai penunjang keputusan pemerataan guru seluruh Kabupaten/Kota di Indonesia yang masih kekurangan atau kelebihan guru.

**Kata kunci** — Penyebaran Guru, Data Mining, Clustering, K-Medoids, K-Means

## I. PENDAHULUAN

Pendidikan bagi manusia sangatlah penting, karena dengan pendidikan manusia mampu mengembangkan potensi diri untuk meraih kehidupan yang lebih baik. Pendidikan memberikan manfaat untuk seluruh aspek kehidupan salah satunya dalam berbangsa, pendidikan mampu memberikan kemajuan dan kualitas bangsa. Kualitas guru yang baik akan memperoleh pendidikan yang baik [1]. Sekolah dasar dan menengah (SD dan SMP) serta jenjang pendidikan lain yang sederajat menjadi landasan bagi pendidikan dasar seseorang. Data Roadmap Manajemen ASN dan Perencanaan Formasi tahun 2014 menunjukkan bahwa instruktur Indonesia tidak tersebar merata di seluruh wilayah tanah air. Dari TK hingga SMA, bakat dan kekurangan instruktur sudah terlihat [2]. Pada bulan April Tahun 2018, Dewan Perwakilan Daerah (DPD) RI meresmikan perubahan Undang-Undang Guru dan Dosen Nomor 14 Tahun 2005. Terbentuklah lima tim ahli yaitu dua dari Persatuan Guru Republik Indonesia (PGRI) dan tiga dari Asosiasi Dosen Indonesia (ADI) yang bekerja selama enam bulan. Mewakili PGRI bersama Prof Supardi. Dari proses kerja tim ahli dan anggota DPD RI itu ditemukan masalah-masalah guru, diantaranya: kompetensi, perlindungan, pemerataan dan kesejahteraan [1].

Riset lain yang telah dilaksanakan dengan teknik *clustering* diantaranya oleh Priambodo dan Prasetyo [3] dengan menggunakan Algoritma K-Means untuk pemetaan penyebaran guru provinsi Banten dalam penelitian tersebut menghasilkan pemetaan wilayah Provinsi Banten berlandaskan tingkat pendidikan yang mempunyai kekurangan, kecukupan dan kelebihan guru berlandaskan kabupaten/kota. Dengan menggunakan teknik K-Means, penelitian lain oleh Pratomo, Irawati, dan Saptomo [4] menarik distribusi Pamsimas menjadi dua cluster. Klaster 1 terdiri dari enam pemukiman yang terletak di dekat sumber air tetapi tidak dekat dengan pabrik. Terdapat 22 permukiman dalam klaster dua potensi masyarakat tertinggal. Algoritma K-Medoids digunakan oleh Riyanto B [5] dalam studi lanjutannya untuk mengklasifikasikan penyebaran diare di Medan. Karena kurangnya penyakit diare, generasi cluster awal, atau C0, kurang mendapat perhatian dibandingkan generasi cluster berikutnya. Cluster kedua Medan, atau C1, harus menjadi perhatian utama pemerintah karena merupakan pusat wabah diare terbesar di kota itu. Penelitian selanjutnya oleh Dewi, Siregar dan Kusumaningrum [6] yaitu dengan menggunakan algoritma K-Means yang berjudul pengelompokan SDM kesehatan Puskesmas pada Provinsi Jawa Tengah. Penelitian tersebut menghasilkan *cluster* 1 dengan nilai tinggi (kelebihan SDM) sebanyak 4 Kab/Kota, *cluster* 2 dengan nilai sedang (kecukupan SDM) sebanyak 25 Kab/Kota dan *cluster* 3 dengan nilai rendah (kekurangan SDM) sebanyak 6 Kab/Kota. Dan penelitian lainnya oleh Murniasih, Siregar dan Wahidin [7] dengan menggunakan Algoritma K-Means dalam mengurangi tingkat buta aksara di Indonesia. Hasil dari penelitian tersebut yaitu *cluster* tertinggi memiliki jumlah anggota 2 provinsi, *cluster* sedang memiliki jumlah anggota 10 provinsi dan cluster rendah memiliki jumlah anggota 22 provinsi.

Berdasarkan penelitian [3,4,5,6,7] alhasil dalam riset ini penulis bakal menerapkan metode *clustering* algoritma K-Medoids dan K-Means untuk pemetaan penyebaran guru tingkat SMP seluruh Kab/Kota di Indonesia sebagai salah satu data pendukung yang dapat digunakan sebagai saran untuk pemerintah dalam upaya mensukseskan program pemerintah wajib belajar 9 Tahun dan mengatasi daerah-daerah yang memiliki kelebihan dan kekurangan Guru.

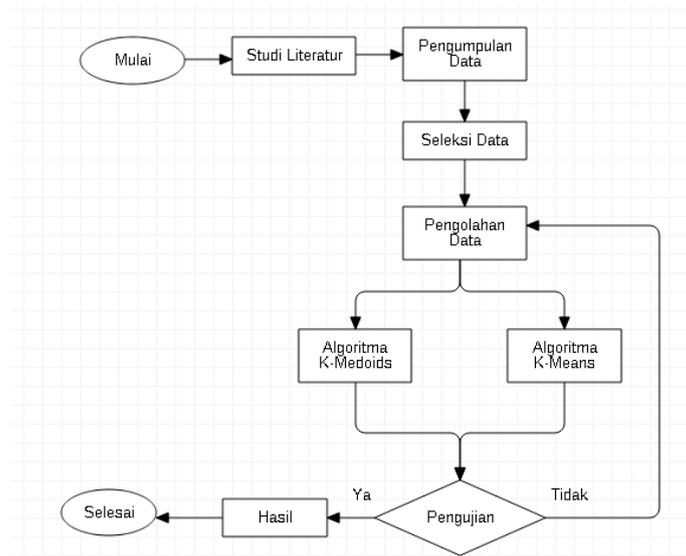
II. DATA DAN METODE

A. Pengumpulan Data

Dataset penelitian ini diperoleh dari *website* resmi Kementerian Pendidikan dan Kebudayaan (KEMDIKBUD) dengan alamat <http://dapo.dikdasmen.kemdikbud.go.id>. *Attribute* dataset terdairi dari wilayah, data guru, data sekolah dan data peserta didik. Data yang diambil hanya tingkat SMP tahun ajaran 2019/2020 dan berjumlah 514 Kab/Kota.

B. Gambaran Umum Penelitian

Terdapat juga gambaran umum riset yang dilaksanakan ada dalam Gambar 1



Gambar 1 Gambaran Umum Penelitian

C. Guru

Kedudukan guru sangat penting bagi mutu pendidikan dan kemajuan bangsa. Penduduk yang terdidik adalah aset berharga di setiap ekonomi maju. Guru membuat atau menghancurkan pengalaman pendidikan siswa [1].

D. Clustering

*Clustering* yaitu proses pengelompokkan objek-objek ke dalam beberapa group yang berlainan, atau tepatnya mempartisi dataset menjadi *subsets (cluster)* [8]. *Clustering* juga disebut analisis *cluster* yaitu proses pengelompokan antar suatu set data benda-benda ataupun abstrak kedalam kelas yang objeknya sama [9]. Untuk memaksimalkan jumlah data di setiap cluster, strategi clustering membagi data/objek menjadi pengelompokan yang lebih kecil. Penting untuk menjaga cluster sejauh mungkin sambil mengatur objek dalam cluster [10].

E. Normalisasi Data

Sebagai bagian dari transformasi data, nilai data diubah sebagai bagian dari proses normalisasi data dalam penambangan data. Properti data dinormalisasi ke rentang terbatas, seperti nilai antara -1 dan 1, atau dari 0 hingga 1 [11]. Pendekatan Min-max adalah salah satu dari berbagai cara untuk menormalkan data. Pendekatan min-max adalah transformasi linier dari data asli yang digunakan untuk menormalkan data. Perubahan linier dari atribut data asli untuk membuat rentang nilai yang sama dikenal sebagai normalisasi min-max [11]. Menggunakan pendekatan min-max, dapat melihat rumus pada Persamaan 1.

$$Newdata = \frac{(data - min) \times (new\ max - new\ min)}{(max - min) + (new\ min)} \tag{1}$$

Keterangan:

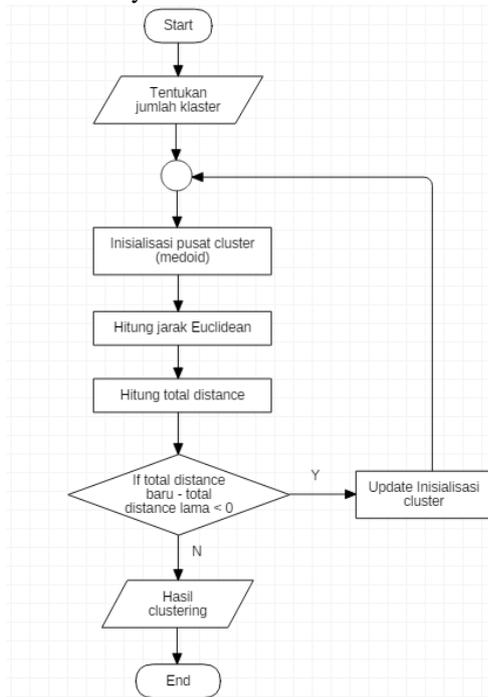
- Newdata* = Data hasil normalisasi
- Min* = Nilai minimum dari data perkolom
- Max* = Nilai maksimal dari data perkolom
- Newmin* = Batas minimum yang diberikan
- Newmax* = Batas maksimum yang diberikan

F. Algoritma K-Medoids Clustering

Menurut penelitian Hae-Sang Park dan ChiHyuck Jun, membahas tentang sebuah algoritma baru untuk K-medoids dalam pengelompokan yang berjalan seperti K-means dan tes sebagian teknik guna memilih medoids awal [12]. Mirip dengan K-Means, algoritma K-Medoids adalah algoritma clustering. Algoritma K-Medoids atau PAM menggunakan objek sebagai representasi (medoid) sebagai pusat cluster untuk setiap cluster, sedangkan K-Means menggunakan mean (mean) sebagai pusat cluster [11].

Oleh karena itu, pendekatan K-Medoids memiliki keunggulan dibandingkan K-Means, yang memiliki masalah dengan noise dan outlier, seperti objek dengan nilai besar yang memungkinkan mereka menyimpang dari distribusi nilai. Manfaat lebih lanjut adalah bahwa hasil dari prosedur pengelompokan tidak terpengaruh oleh urutan di mana kumpulan data input disajikan [11].

Terdapat juga alur dari algoritma dari K-Medoids yaitu:



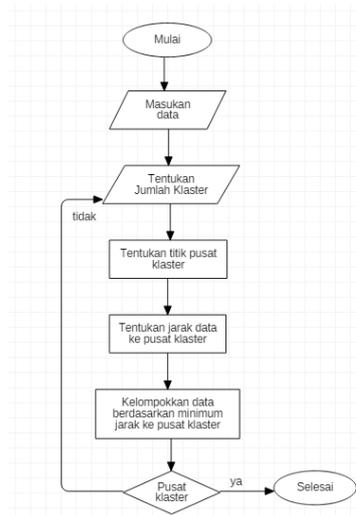
Gambar 2. 1 Alur Algoritma K-Medoids [13].

Langkah-langkah algoritma K-Medoids:

1. Inisialisasi pusat *cluster* sebanyak k (jumlah *cluster*)
2. Alokasikan setiap data (objek) ke *cluster* terdekat menggunakan persamaan ukuran jarak *Euclidian Distance*.
3. Alokasikan setiap data (objek) ke *cluster* terdekat menggunakan persamaan ukuran jarak *Euclidian Distance*.
4. Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat medoids baru.
5. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat medoids baru.
6. Hitung total simpangan (S) dengan menghitung nilai total *distance* baru – total *distance* lama. Jika  $S < 0$ , maka tukar objek dengan data *cluster* untuk membentuk sekumpulan k objek baru sebagai medoid.
7. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan medoids, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing.

### G. Algoritma K-Means Clustering

Masalah dengan pengelompokan dataset dapat diselesaikan dengan menggunakan algoritma K-Means yang banyak digunakan dan sangat mudah. Seperti namanya, algoritma K-Means adalah algoritma evolusioner, dan operasinya identik dengan namanya [14]. Bisa dikatakan memiliki kumpulan data yang penuh dengan data yang akan dikelompokkan menjadi beberapa grup. Dengan teknik K-Means, semua cluster memiliki kesalahan kuadratnya dikurangi menjadi angka serendah mungkin [15]. Namun, strategi ini memiliki kelemahan dalam menilai distribusi data dan mengandalkan centroid. Jarak centroid setiap cluster adalah semua yang dimiliki metode K-Means [16].



Gambar 2. 2 Alur Algoritma K-Means [13].

Data *clustering* memakai Algoritma K-Means menurut Ramadhan, Efendi dan Mustakim [16] secara umum dilaksanakan dengan algoritma dasar yaitu:

1. “Tentukan jumlah *cluster*
2. Alokasikan data ke dalam *cluster* secara random
3. Hitung centroid dari data yang ada dimasing-masing *cluster* dengan menggunakan persamaan 2.

$$C_i = \frac{1}{M} \sum_{j=1}^M x_j \tag{2}$$

4. Alokasikan masing-masing data ke *centroid*/ rata-rata terdekat
5. Kembali menentukan titik pusat *cluster*, apabila titik pusat *cluster* berubah, jika tidak berubah maka proses *cluster* selesai”.

### III. HASIL DAN PEMBAHASAN

Dalam komponen ini bakal diuraikan pengolahan datadengan menggunakan algoritma algoritma K-Medoids dan K-Means untuk pemetaan penyebaran guru seluruh Kab/Kota di Indonesia. Data yang diperoleh dari *website* resmi Kementerian Pendidikan dan Kebudayaan (KEMDIKBUD) dengan alamat <http://dapo.dikdasmen.kemdikbud.go.id>. Atribut data yang digunakan yaitu data guru, data sekolah dan data siswa dengan jumlah setiap data ialah 514 data. Sebelummelanjutkan perhitungan data dengan algoritma K-Medoids dan K-Means, data dilakukan proses dinormalisasi. Terdapat juga hasil dari proses normalisasi data penyebaran guru bisa diamati dalam tabel 2.

Tabel 2 Hasil Normalisasi Data

No.	Kabupaten/Kota	Jumlah Guru	Jumlah Sekolah	Jumlah Peserta Didik
1	Kab. Bandung	0,6873	0,4650	0,6510
2	Kab. Bandung Barat	0,3338	0,2381	0,2948
3	Kab. Bekasi	0,6050	0,4762	0,5659
4	Kab. Bogor	1	1	1
5	Kab. Ciamis	0,2591	0,1667	0,1774
6	Kab. Cianjur	0,5003	0,4566	0,4561
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
514	Kota Tarakan	0,0633	0,0294	0,0484

#### A. Hasil Perhitungan Algoritma K-Medoids

Algoritma K-Medoids langkah pertama menentukan jumlah *cluster*, dalam riset ini jumlah *cluster* yang di ambil yaitu 3 *cluste r* yaitu kelebihan, kecukupan dan kekurangan. Selanjutnya menentukan secara acak diantara objek dari dataset untuk dijadikan sebagai awal pusat klaster. Jumlah objek yang dipilih sesuaikan dengan jumlah klaster yang diinginkan. Dalam penelitian ini klaster yang diambil yaitu 3 klaster. Medoid awal dapat dilihat pada tabel 3.

Tabel 3 Medoid awal

Centroid 1	0,0576	0,0532	0,0327
Centroid 2	0,4588	0,3782	0,3744
Centroid 3	0,1635	0,1176	0,1004

Kemudian menghitung objek data ke pusat *cluster* dengan menggunakan metode jarak *Euclidean Distance*. Berikut perhitungan jarak data ke pusat *cluster*.

Iterasi Pertama

Jarak pusat *cluster* 1 (d1)

$$d_1 = \sqrt{(0,6873 - 0,0576)^2 + (0,4650 - 0,0532)^2 + (0,6510 - 0,0327)^2} = 0,9739$$

Ulangi langkah yang sama sampai ke data 514

$$d_{514} = \sqrt{(0,0633 - 0,0576)^2 + (0,0294 - 0,0532)^2 + (0,0484 - 0,0327)^2} = 0,0291$$

Jarak pusat *cluster* 2 (d2)

$$d_1 = \sqrt{(0,6873 - 0,4588)^2 + (0,4650 - 0,3782)^2 + (0,6510 - 0,3744)^2} = 0,3692$$

...

$$d_{514} = \sqrt{(0,0633 - 0,4588)^2 + (0,4650 - 0,3782)^2 + (0,6510 - 0,3744)^2} = 0,6199$$

Jarak pusat *cluster* 3 (d3)

$$d_1 = \sqrt{(0,6873 - 0,1635)^2 + (0,4650 - 0,1176)^2 + (0,6510 - 0,1004)^2} = 0,8356$$

...

$$d_{514} = \sqrt{(0,0633 - 0,1635)^2 + (0,4650 - 0,1176)^2 + (0,6510 - 0,1004)^2} = 0,1432$$

Setelah menghitung jarak objek ke pusat *cluster* selanjutnya mengelompokan dengan cara membandingkan objek mana yang paling dekat dengan pusat *cluster*. Hasil pengelompokan iterasi 1 dapat dilihat dalam tabel 4.

Tabel 4 Iterasi 1 algoritma K-Medoids

No.	Kab/Kota	d1	d2	d3	Jarak Terdekat	Cluster
1	Kab. Bandung	0,9739	0,3692	0,8356	0,3692	2
2	Kab. Bandung Barat	0,4233	0,2039	0,2852	0,2039	2
3	Kab. Bekasi	0,8734	0,2601	0,7350	0,2601	2
4	Kab. Bogor	1,6493	1,0349	1,5125	1,0349	2
5	Kab. Ciamis	0,2728	0,3513	0,1322	0,1322	3
6	Kab. Cianjur	0,7335	0,1206	0,5957	0,1206	2
7	Kab. Cirebon	0,5993	0,1170	0,4599	0,1170	2
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...
514	Kota Tarakan	0,0291	0,6199	0,1432	0,0291	1
<b>Total Jarak Terdekat</b>					<b>35,1517</b>	

Berdasarkan jarak terdekat atau nilai minimum yang telah ditentukan pada Tabel 4, maka diperoleh hasil pengelompokan terdapat pada tabel sebagai berikut:

Tabel 5 Jumlah Anggota Iterasi 1

Cluster	Jumlah Anggota
1	285
2	38
3	191

Setelah mendapatkan nilai hasil iterasi pertama, pada perhitungan kedua memilih kembali secara acak medoid baru pada data untuk dilakukan *clustering* data. Ulangi langkah yang sama sampai hasil terasi lebih besar dari hasil iterasi sebelumnya sehingga proses *clustering* dihentikan.

Tabel 6 Iterasi 2 algoritma K-Medoids

No.	Kab/Kota	d1	d2	d3	Jarak Terdekat	Cluster
1	Kab. Bandung	0,9707	0,6068	0,8685	0,6068	2
2	Kab. Bandung Barat	0,4212	0,0623	0,3225	0,0623	2
3	Kab. Bekasi	0,8723	0,5039	0,7626	0,5039	2
4	Kab. Bogor	1,6507	1,2812	1,5295	1,2812	2
5	Kab. Ciamis	0,2717	0,1100	0,1832	0,1100	2
6	Kab. Cianjur	0,7346	0,3646	0,6173	0,3646	2
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...
514	Kota Tarakan	0,0000	0,3700	0,1505	0,0000	1
Total Jarak Terdekat					<b>42,3641</b>	

Selanjutnya hitung total simpangan (S) dengan menghitung nilai total jarak terdekat pada iterasi kedua (b) – total jarak terdekat iterasi pertama (a).

$$S = 42,3641 - 35,1517 = 7,2124$$

Hasil dari perhitungan total simpangan (S) dapat disimpulkan bahwa *clustering* dihentikan pada iterasi ke-2 karena  $b > a$ . Berikut tabel 5 jumlah anggota dari masing-masing *cluster*.

Tabel 5 Jumlah Anggota Iterasi 2

Cluster	Jumlah Anggota
1	302
2	77
3	135

Setelah mendapatkan hasil pengelompokan masing-masing anggota *cluster* selanjutnya menentukan *cluster* mana yang memiliki nilai presentase Tinggi (kelebihan guru), presentase Sedang (cukup guru) dan presentase Rendah (kurang guru) yaitu dengan cara menjumlahkan data guru, data sekolah dan data peserta didik dari masing-masing *cluster* kemudian hitung *range cluster* 1, *cluster* 2 dan *cluster* 3. Berikut dari setiap *cluster* bisa diamati dalam tabel 6.

Tabel 6 Hasil Presentase *cluster* 1, 2 dan 3 algoritma K-medoids

Cluster	Range	Jumlah Data	Presentase
1	Max	0,3439	Rendah
	Min	0,0022	
2	Max	0,3000	Tinggi
	Min	0,5077	
3	Max	0,5737	Sedang
	Min	0,2326	

Berdasarkan Tabel 6 hasil perhitungan algoritma K-Medoids *clustering* dalam pemetaan penyebaran guru yaitu *Cluster* 1 yang memiliki kekurangan guru sebanyak 302 Kabupaten/Kota, *Cluster* 2 yang memiliki kelebihan guru sebanyak 77 Kabupaten/Kota sedangkan *Cluster* 3 yang memiliki cukup guru sebanyak 135 Kabupaten/Kota. Berikut tabel *cluster* beserta anggotanya dapat dilihat pada tabel 7.

Tabel 7 Hasil Algoritma K-Medoids

Cluster 1 Kekurangan Guru		Cluster 2 Kelebihan Guru		Cluster 3 Kecukupan Guru	
1	Kab. Pangandaran	1	Kab. Bandung	1	Kab. Pangandaran
2	Kota Cirebon	2	Kab. Bandung Barat	2	Kota Bekasi
3	Kota Tasikmalaya	3	Kab. Bekasi	3	Kota Depok
4	Kota Banjar	4	Kab. Bogor	4	Kota Blitar
5	Kab. Madiun	6	Kab. Ciamis	5	Kab. Bojonegoro
6	Kab. Magetan	7	Kab. Cianjur	6	Kab. Bondowoso
...	...	...	...	...	...
...	...	...	...	...	...

Cluster 1 Kekurangan Guru		Cluster 2 Kelebihan Guru		Cluster 3 Kecukupan Guru	
...	...	...	...	...	...
302	Kab. Tarakan	77	Kota Batam	135	Kab. Gorontalo

B. Hasil Algoritma K-Means

Selanjutnya perhitungan pula dilaksanakan dengan menggunakan algoritma K-Means guna mengidentifikasi pola tertentu, tahapannya hampir sama dengan algoritma K-Medoids perbedaannya hanya terletak pada pemilihan titik centroid baru. Pada algoritma K-Medoids pemilihan titik centroid baru dipilih secara random atau berurutan pada dataset sedangkan dalam menentukan titik centroid baru pada algoritma K-Means yaitu mencari nilai mean dengan menjumlahkan seluruh nilai anggota setiap cluster lalu dibagi jumlah anggotanya.

Dalam perhitungan algoritma K-Means dengan menggunakan metode jarak *Euclidean Distance* menghasilkan 4 iterasi dimana iterasi ketiga menghasilkan pola iterasi yang sama dengan iterasi ke-4. Berikut hasil perhitungan jarak iterasi ke-4 bisa diamati dalam tabel 9.

Tabel 9 Hasil Iterasi ke-4 Algoritma K-Means

No.	Kab/Kota	d1	d2	d3	Cluster
1	Kab. Bandung	0,9568	0,7378	0,2910	3
2	Kab. Bandung Barat	0,4063	0,1888	0,2835	2
3	Kab. Bekasi	0,8567	0,6368	0,1855	3
4	Kab. Bogor	1,6334	1,4149	0,9602	3
5	Kab. Ciamis	0,2548	0,0364	0,4307	2
6	Kab. Cianjur	0,7172	0,4980	0,0792	3
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...
514	Kota Tarakan	0,0312	0,2415	0,7013	1

Tabel 10 Jumlah Anggota Cluster Iterasi Ke-4

Cluster	Jumlah Anggota
1	363
2	125
3	26

Tabel 11 Presentase Cluster 1,2,3 Algoritma K-Means

Cluster	Range	Jumlah Data	Presentase
1	Max	0,3674	Rendah
	Min	0,0022	
2	Max	0,9322	Sedang
	Min	0,3597	
3	Max	0,3000	Tinggi
	Min	0,9715	

Tabel 12 Hasil Keseluruhan Algoritma K-Means

Cluster 1 Kekurangan Guru		Cluster 2 Kecukupan Guru		Cluster 3 Kelebihan Guru	
1	Kab. Pangandaran	1	Kab. Bandung Barat	1	Kab. Bandung
2	Kota Cirebon	2	Kab. Ciamis	2	Kab. Bekasi
3	Kota Tasikmalaya	3	Kab. Kuningan	3	Kab. Bogor
4	Kota Banjar	4	Kab. Majalengka	4	Kab. Cianjur
5	Kab. Madiun	5	Kab. Purwakarta	5	Kab. Cirebon
6	Kab. Magetan	6	Kab. Subang	6	Kab. Garut
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...
363	Kota Tarakan	125	Kota Batam	26	Kota Jakarta Timur

C. Hasil Pengujian Menggunakan *Tool* Rapid Miner

Hasil Pengujian dengan menggunakan *tool* Rapid Miner pada algoritma K-medoids ialah *cluster* 1 terdiri dari 302 *items*, *cluster* 2 terdiri dari 77 *items* dan *cluster* 3 terdiri dari 135 *items*. Sedangkan hasil dari Algoritma K-Means yaitu *cluster* 1 terdiri dari 363 *items*, *cluster* 2 terdiri dari 125 *items* dan *cluster* 3 terdiri 26 *items*. Total dari masing-masing *output* yaitu 514 *items*.

Tabel 13 Hasil Pengujian Rapid Miner

<i>Cluster</i>	Algoritma K-Medoids	Algoritma K-Means
<i>Cluster</i> 1	302 <i>Items</i>	363 <i>Items</i>
<i>Cluster</i> 2	77 <i>Items</i>	125 <i>Items</i>
<i>Cluster</i> 3	135 <i>Items</i>	26 <i>Items</i>
Total number of <i>items</i>	514	514

## IV. KESIMPULAN DAN SARAN

Sesuai dengan perhitungan dan pengujian Algoritma K-Medoids dan K-Means berhasil diimplementasikan untuk pengelompokan pemetaan penyebaran guru tingkat SMP diseluruh Kab/Kota di Indonesia dengan jumlah klaster 3 yaitu kelebihan, cukup dan kurang. Algoritma K-medoids menghasilkan *cluster* 1 kelompok anggota yang memiliki kelebihan guru sebanyak 302 Kab/Kota, *cluster* 2 kelompok anggota yang memiliki cukup guru sebanyak 135 Kab/Kota dan *cluster* 3 kelompok anggota yang memiliki kekurangan guru sebanyak 77 Kab/Kota. Sedangkan Algoritma K-means kelompok yang memiliki kelebihan guru sebanyak sebanyak 363 Kab/Kota, yang memiliki cukup guru sebanyak 125 Kab/Kota dan yang memiliki kekurangan guru sebanyak 26 Kab/Kota.

## PENGAKUAN

Naskah ilmiah ini yaitu sebagian dari riset Tugas Akhir milik Lilis Kartika dengan judul “Penerapan Algoritma K-Medoids dan K-Means untuk Pemetaan Penyebaran Guru tingkat SMP seluruh Kabupaten/Kota di Indonesia, yang dibimbing oleh Amril Mutoi Siregar dan Dwi Sulistya Kusumaningrum”.

## REFERENCES

- [1] Jilan, B. 2018. Permasalahan Guru di Indonesia. Universitas Islam Negeri Syarif Hidayatullah Jakarta. <https://www.uinjkt.ac.id/id/permasalahan-guru-di-indonesia/>. 09 Januari 2020 (15:30).
- [2] CPNS Pusat. 2014. Data Kelebihan dan Kekurangan Guru di Indonesia. [www.asncpns.com/2014/04/data-kelebihan-dan-kekurangan-guru-di.html](http://www.asncpns.com/2014/04/data-kelebihan-dan-kekurangan-guru-di.html). 12 Januari 2020 (15:40).
- [3] Priambodo, Y.A. dan S.Y.J. Prasetyo. 2018. Pemetaan Penyebaran Guru di Provinsi Banten dengan Menggunakan Metode Spatial Clustering K-Means (Studi Kasus : Wilayah Provinsi Banten). *Indonesian Journal of Modeling and Computing*. 18-27.
- [4] Pratomo, E. D., I. Tri, dan W. L. Y. Saptomo. 2019. Metode K-Means Dalam Pemetaan Penyebaran Pamsimas. *Jurnal TIKomSiN*. 7(2): 1-6.
- [5] Riyanto, B. 2019. Penerapan Algoritma K-Medoids Clustering Untuk Pengelompokan Penyebaran Diare Di Kota Medan ( Studi Kasus : Kantor Dinas Kesehatan Kota Medan. *Konferensi Nasional Teknologi Informasi dan Komputer*. 3(1): 562–568.
- [6] Dewi, S. C., A. M. Siregar, dan D. S. Kusumaningrum. 2020. Pengelompokan Jumlah Sumber Daya Manusia Kesehatan Puskesmas untuk Menunjang Pemerataan pada Provinsi Jawa Tengah Menggunakan Algoritma K-Means. *Scientific Student Journal for Information, Technology and Science*. 1(2): 86-94.
- [7]
- [8] Primartha, R. 2018. *Belajar Machine Learning Teori Dan Praktek*. Informatika Bandung. Bandung
- [9] Siregar, A. M. dan A. Puspabhuana. 2017. *DATA MINING: Pengolahan Data Menjadi Informasi dengan RapidMiner*. Penerbit CV. Kekata Group.
- [10] Santosa, B. 2013. *Data Mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis*. 1<sup>st</sup> ed. Penebar Media Pustaka. Yogyakarta.
- [11] Pramesti, D. F, M. T. Furqon, dan C. Dewi. 2017. Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer* 1(9): 723-732.
- [12] Mustofa, Z., dan I. S. Suasana. 2018. Algoritma Clustering K-Medoids pada E-Government Bidang Information and Communication. *Jurnal Teknologi Informasi Dan Komunikasi*. 9(1): 1–10.
- [13] Marlina, D., N. F. Putri, A. Fernando, dan A. Ramadhan. 2018. Implementasi Algoritma K-Medoids dan K-Means untuk

Pengelompokan Wilayah Sebaran Cacat pada Anak. *Jurnal CoreIT*. 4(2): 64-71.

- [14] Kamila, I., U. Khairunnisa, dan Mustakim. 2019. Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*. 5(1): 119–125.
- [15] Khairati, A. F., A. A. Adlina, G. F. Hertono, dan B. D. Handari. 2019. Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA. *Prosiding Seminar Nasional Matematika*. 2:161–170.
- [16] Ramadhan, A., Z. Efendi, dan Mustakim .2017. Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling. *Seminar Nasional Teknologi Informasi, Komunikasi Dan Industri (SNTIKI)*. 9: 18–19.