

PENGELOMPOKAN WILAYAH MADURA BERDASAR INDIKATOR PEMERATAAN PENDIDIKAN MENGGUNAKAN *PARTITION AROUND MEDOIDS* DAN VALIDASI *ADJUSTED RANDOM INDEX*

Budi Dwi Satoto¹⁾, Bain Khusnul khotimah²⁾, Iswati³⁾

¹⁾ *Manajemen Informatika, Fakultas Teknik, Universitas Trunojoyo*

²⁾³⁾ *Teknik Informatika, Fakultas Teknik, Universitas Trunojoyo*

¹⁾²⁾³⁾ *Jl. Telang PO Box 2 Kamal, Bangkalan, Madura*

¹⁾ *budids@yahoo.com*

²⁾ *bain_kk@yahoo.com*

Abstrak— Pemerataan pendidikan di Indonesia telah menjadi perhatian pemerintah sejak lama. Namun hingga saat ini, pendidikan di Indonesia masih belum merata. Hal tersebut dapat dilihat dari rendahnya nilai Angka Partisipasi Kasar (APK) dan Angka Partisipasi Murni (APM) di daerah-daerah tertentu serta belum meratanya sarana dan prasarana pendidikan. Adapun tujuan penelitian ini adalah memberikan informasi kepada pemerintah setempat mengenai kondisi pendidikan di wilayahnya sehingga dapat menghasilkan kebijakan yang tepat mengenai pengembangan infrastruktur pendidikan dan distribusi guru bantu. *Clustering* adalah metode data mining yang membagi data kedalam kelompok yang mempunyai objek yang karakteristik sama. Penelitian ini menggunakan metode *clustering Partition Around Medoids* (PAM) dengan 3 *distance measure*: Manhattan, Euclidean dan Canberra *distance*. Untuk mengukur kualitas hasil *clustering*, digunakan nilai *Adjusted Rand Index* (ARI). Semakin besar nilai ARI, semakin baik kualitas *cluster*. Dari 3 kali uji coba diperoleh rata-rata nilai ARI untuk Euclidean *distance* sebesar 0.799, Manhattan *distance* dengan rata-rata sebesar 0.738 dan Canberra *distance* sebesar 0.163. Sedangkan pengelompokan terbaik diperoleh menggunakan Euclidean *distance* dengan nilai ARI sebesar 0.825 dan kecocokan dengan label asli sebesar 83.33%. Dari pengelompokan terbaik menghasilkan kelompok pemerataan tinggi terdiri dari 11 kecamatan, kelompok pemerataan sedang terdiri dari 15 kecamatan dan kelompok pemerataan rendah terdiri dari 46 kecamatan.

Kata Kunci— indikator pemerataan pendidikan, *clustering*, *Partition Around Medoid*, *distance measure*, *Adjusted Random Index*

Abstract—Distribution of education in Indonesia has become government's attention for a long time. But until now, education in Indonesia is still not evenly distributed. This can be seen from the low value of Participation Rough figures and net enrollment ratio in certain areas as well as uneven educational facilities. The purpose of this research is to provide information to local authorities about the state of education in local region to produce an appropriate policy regarding development of educational infrastructure and teachers assistant distribution. Clustering is a data mining method that divides data into several groups with the same object characteristics. This research used Partition Around Medoids methods with 3 distance measure that contain Manhattan, Euclidean and Canberra distance. Adjusted Random Index used to measure the quality of clustering results. From 3 times sampling, better value of ARI Euclidean distance 0.799, Manhattan distance 0.738 and Canberra distance 0.163 while the best clustering obtained is Euclidean distance with value of ARI 0.825 and compatibility with the original label 83.33%. it is produces high equity group composed of 11 districts with equity groups are composed of 15 districts and low equity group consists of 46 sub-districts.

Keywords—Indicator of Educational Equity, Clustering, Partition Around Medoid, Distance Measure, Adjusted Random Index .

I. PENDAHULUAN

Pemerataan pendidikan dasar di Indonesia telah menjadi perhatian pemerintah sejak lama. Namun hingga saat ini pendidikan di Indonesia masih belum merata. Pemerataan pendidikan mencakup dua aspek penting yaitu persamaan kesempatan untuk memperoleh pendidikan dan keadilan dalam memperoleh pendidikan yang sama dalam masyarakat. Secara umum, kondisi pemerataan

pendidikan dapat dilihat dari angka partisipasi kasar (APK) dan angka partisipasi murni (APM). Selain APK dan APM, terdapat beberapa indikator lain yang berpengaruh untuk menentukan tingkat pemerataan pendidikan pada suatu daerah. Indikator-indikator yang dimaksud adalah indikator yang berhubungan dengan sarana dan prasarana pendidikan seperti jumlah sekolah, ruang kelas, serta tenaga pengajar.

Dengan menggunakan data dari pedesaan di Cina, Han dan Jian-yu mempublikasikan pembangunan dua model independen menggunakan koefisien Gini dan pendapatan penduduk pedesaan. Dalam publikasinya menyatakan bahwa pembangunan ekonomi suatu negara cenderung memilih langkah atau variabel kebijakan dimana terdapat kesenjangan distribusi pendapatan dan peningkatan penduduk pedesaan. Melalui model analisis dan studi lebih lanjut ditemukan bahwa proses urbanisasi, memfasilitasi aliran tenaga kerja dari desa ke kota, akan meningkatkan pendapatan penduduk pedesaan. Selain itu, terdapat kebijakan penting dan langkah lain untuk menyesuaikan kesenjangan pendapatan penduduk pedesaan yaitu dengan mempromosikan pengembangan industri non-pertanian, meningkatkan income upah dan gaji bagi warga pedesaan, serta mengintensifkan upaya pendanaan pusat untuk mendukung pertanian melalui instensifikasi pendidikan pedesaan diantaranya mewujudkan pemerataan pendidikan sehingga dapat mendorong pertumbuhan ekonomi pedesaan serta menyesuaikan kesenjangan pendapatan penduduk desa. (Han, Ding, & Wang, 2010)

Terdapat perbedaan mencolok antara sarana dan prasarana pendidikan yang terdapat di daerah perkotaan dan pedesaan. Di perkotaan, pendidikan diselenggarakan dengan sarana dan prasarana yang sangat memadai sedangkan di pedesaan, penyelenggaraan pendidikan hanya mengandalkan sarana prasarana seadanya. Hal ini menjadi bukti nyata bahwa pendidikan di Indonesia belum merata. Guru adalah aset penting bagi pendidikan di suatu negara. Mereka adalah ujung tombak keberhasilan pendidikan dan dianggap memegang peranan penting dalam mencapai tujuan pendidikan itu. Keberhasilan siswa jauh lebih dipengaruhi oleh guru-guru mereka. Oleh karena itu guru dituntut untuk memiliki kualitas yang baik, baik moral dan pengetahuan yang cukup untuk menjadi panutan bagi siswanya. Namun, dengan kondisi geografis sebagai negara kepulauan dan beberapa keterbatasan lain, sulit untuk memiliki kesetaraan dalam fasilitas dan guru pendidikan berkualitas untuk seluruh wilayah Indonesia terutama di daerah pedesaan (Galinium, Defindal, & Melissa, 2012).

Pemerataan pendidikan dalam hal sarana dan prasarana berhubungan erat dengan kualitas pendidikan. Saat sarana dan prasarana pendidikan merata, maka kualitas pendidikan akan meningkat. Sebaliknya jika pemerataan pendidikan belum tercapai maka kualitas pendidikan pun tidak akan dapat meningkat. Oleh karena itu, pemerintah hendaknya mampu membuat kebijakan yang dapat meningkatkan pemerataan pendidikan (Xing, Zhang, Wang, & Zhang, 2009). Tujuannya agar Pemerintah setempat mendapatkan gambaran kondisi pendidikan di wilayahnya agar dapat

menghasilkan kebijakan yang tepat arah dan sasaran agar data tersebut dapat menjadi referensi dalam melakukan inventarisasi perangkat yang diperlukan untuk meningkatkan pendidikan baik infrastruktur maupun distribusi tenaga kependidikan. Pengelompokan (*clustering*) adalah metode *data mining* yang membagi data ke dalam grup-grup yang mempunyai objek yang karakteristiknya sama. Salah satu hal penting dalam *clustering* adalah pemilihan *distance measure* untuk mengukur kedekatan antara dua objek. *Distance measure* akan mempengaruhi perhitungan dalam menentukan anggota suatu *cluster*. *Partition Around Medoids* (PAM) adalah salah satu metode *partitioning clustering*. PAM ditemukan oleh Kaufman dan Rousseeuw pada tahun 1990.

Berdasarkan hasil studi literatur diatas, maka dalam penelitian ini digunakan data 72 kecamatan di 4 kabupaten wilayah Madura yaitu Bangkalan, Sampang, Pamekasan dan Sumenep. Data dikelompokkan berdasarkan indikator pemerataan pendidikan menggunakan metode PAM. Indikator pemerataan pendidikan yang digunakan sebanyak 18 indikator yaitu APK SD, APM SD, rasio murid-guru SD, rasio murid-kelas SD, rasio murid-sekolah SD, angka shift SD, APK SMP, APM SMP, rasio murid-guru SMP, rasio murid-kelas SMP, rasio murid-sekolah SMP, angka shift SMP, APK SMA, APM SMA, rasio murid-guru SMA, rasio murid-kelas SMA, rasio murid-sekolah SMA dan angka shift SMA. *Distance measure* yang digunakan dalam metode PAM yaitu *Manhattan*, *Euclidean* dan *Canberra distance*. Untuk mengetahui *distance measure* mana yang lebih cocok digunakan pada metode PAM dan kasus pengelompokan ini, digunakan nilai *Adjusted Random Index* untuk membandingkan hasil pengelompokannya (Santos & Ramos, 2010).

II. TINJAUAN PUSTAKA

A. *Distance Measure*

Dalam melakukan *pattern matching* ataupun untuk melakukan berbagai jenis pengelompokan (*clustering*), *similarity measure* merupakan bagian penting yang harus diperhatikan karena mempengaruhi perhitungan dalam menentukan anggota suatu *cluster*. Ada beberapa jenis *similarity measure* yang bisa digunakan, salah satunya adalah *Distance-Based Similarity Measure*. *Distance-Based Similarity Measure* mengukur tingkat kesamaan dua buah objek dari segi jarak geometris dari variabel-variabel yang tercakup di dalam kedua objek tersebut. Dalam penelitian ini akan digunakan tiga *distance measure* yaitu *Manhattan*, *Canberra* dan *Euclidean distance*. Untuk penjelasan singkatnya, pada tiga *distance measure* yang akan digunakan fungsi $d(o,m)$ yang merupakan jarak antara objek (o)

dengan pusat *cluster medoid* (m), sedangkan n adalah jumlah fitur/parameter. (Yu-Jie, Cao, Shuang-Yan, & Jian-Ping, 2008)

B. Manhattan Distance

Manhattan *Distance* (*city block distance*) merupakan pengukuran jarak objek dan pusat *cluster* yang paling sederhana dalam perhitungannya. Manhattan *distance* dihitung menggunakan seperti ditunjukkan Eq.1.(Di, Jin-feng, Xue-ping, Lu, & Yi-fei, 2013)

$$d(o, m) = \sum_{i=1}^n |o_i - m_i| \quad (1)$$

C. Euclidean Distance

Euclidean *distance* merupakan pengukuran jarak objek dan pusat *cluster* yang banyak digunakan secara luas dalam berbagai kasus *pattern matching*, termasuk *clustering*. Euclidean *distance* dihitung menggunakan rumus ditunjukkan Eq.2.(Julazadeh, Marsousi, & Alirezaie, 2012)

$$d(o, m) = \left(\sum_{i=1}^n (o_i - m_i)^2 \right)^{\frac{1}{2}} \quad (2)$$

D. Canberra Distance

Canberra distance mencari selisih dari fitur pertama objek pertama dan objek kedua kemudian dibagi dengan jumlah fitur pertama pada objek pertama dan kedua. Dan seterusnya dilakukan penjumlahan dengan fitur kedua hingga fitur ke- k . Dengan cara ini, dapat diantisipasi jika beberapa fitur yang didapatkan memiliki rentang nilai yang jauh. *Canberra distance* dihitung dengan rumus ditunjukkan Eq.3.(Sheng-Yi, 2006)

$$d(o, m) = \sum_{i=1}^n \left(\frac{|o_i - m_i|}{|o_i + m_i|} \right) \quad (3)$$

E. Partition Around Medoids (PAM)

Algoritma *Partition Around Medoids* (PAM) atau dikenal juga dengan *k-medoids* dikembangkan oleh Leonard Kaufman dan Peter J. Rousseeuw pada tahun 1987. Algoritma PAM termasuk metode *partitioning clustering* untuk mengelompokkan sekumpulan n objek menjadi sekumpulan k *cluster*. Representasi *cluster* pada PAM adalah objek dari sekumpulan objek yang mewakili *cluster*, disebut *medoid*. *Cluster* dibangun dengan menghitung kedekatan yang dimiliki antara medoid dengan objek non-medoid menggunakan *distance measure*. Jadi metode partisi masih dapat dilakukan berdasarkan prinsip meminimalkan jumlah dari ketidakmiripan antara setiap objek dan medoid yang sesuai.(Mishra & Hiranwal, 2014) Hal ini membentuk dasar dari metode PAM. Adapun Algoritma *Partition Around Medoids* dapat dituli:

1) *Inisialisasi*: secara acak pilih sebanyak k dari n data sebagai medoid

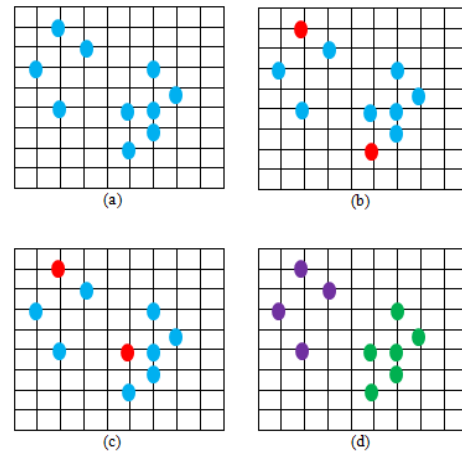
2) *Hitung jarak data ke medoid*: menggunakan rumus Manhattan, Euclidean dan Canberra Distance

3) *Ganti medoid dengan data non-medoid*

4) *Hitung jarak total*

5) *Ulangi langkah 2 hingga 4 sampai tidak ada perubahan jarak total tersebut.*

Setelah pilihan acak awal *medoid* sebanyak k , algoritma berulang kali mencoba untuk membuat pilihan yang lebih baik dari *medoid* (Ying-ting, Fu-zhang, Xing-hua, & Xiao-yan, 2014).



Gambar 1. Ilustrasi PAM

Ilustrasi PAM ditunjukkan pada gambar 1. Gambar 1 memperlihatkan simulasi dari metode PAM. Gambar (a) menunjukkan data awal sebelum penentuan *medoid*. Data akan di-*cluster* menjadi dua *cluster*. Pada gambar (b) penentuan *medoid* untuk iterasi pertama. Lingkaran berwarna merah merepresentasikan *medoid* yang terpilih secara random. Gambar (c) memperlihatkan pergantian *medoid*, untuk meminimalkan total jarak (*cost*) tiap data ke *medoid*-nya. Gambar (d) menunjukkan hasil *cluster*.

F. Adjusted Random Index (ARI)

Untuk mengetahui kualitas hasil *clustering* dari suatu algoritma, diperlukan suatu proses evaluasi. Dalam literature disebutkan bahwa prosedur validasi *cluster* dibagi menjadi 2 kategori utama : *external and internal criterion*. *External criterion* memvalidasi hasil *cluster* dengan membandingkan hasil *cluster* dan *gold standart*. *Gold standart* adalah pengelompokan yang diperoleh dari hasil pemikiran manusia.

ARI merupakan salah satu ukuran validasi yang berbasis *external criterion*. ARI populer digunakan untuk perbandingan hasil *clustering*. Metode ARI ini digunakan untuk mengukur kualitas global solusi dengan membandingkan label hasil proses *clustering* dengan *gold standart*. *Adjusted Rand Index* merupakan perkembangan dari Rand Index. Nilai *Adjusted Rand Index*

berkisar antara -1 sampai 1 $([-1, 1])$, sedang nilai *rand index* berkisar dari 0 sampai 1 $([0, 1])$. Rentang nilai *Adjusted Rand Index* lebih besar daripada *Rand index* menjadikannya sebagai ukuran kinerja yang lebih bagus (de Vargas & Bedregal, 2013). Semakin besar nilai ARI, semakin baik kualitas suatu *cluster*, ditunjukkan Eq.4.

$$RI = \frac{a + d}{a + b + c + d} \quad (4)$$

Adapun Nilai ARI adalah seperti ditunjukkan Eq.5. (Shaohong & Hau-San, 2010)

$$ARI = \frac{2(ad - bc)}{(a + b)(b + d) + (a + c)(c + d)} \quad (5)$$

Dimana:

RI = nilai *rand index*

ARI = nilai *adjusted rand index*

Seperti ditunjukkan Tabel.1, dimana :

a = jumlah data yang masuk kelompok yang sama antara gold standart dan hasil clustering

b = jumlah data yang masuk kelompok yang berbeda antara gold standart dan hasil clustering

c = jumlah data yang tidak masuk kelompok gold standart tapi masuk kelompok hasil clustering

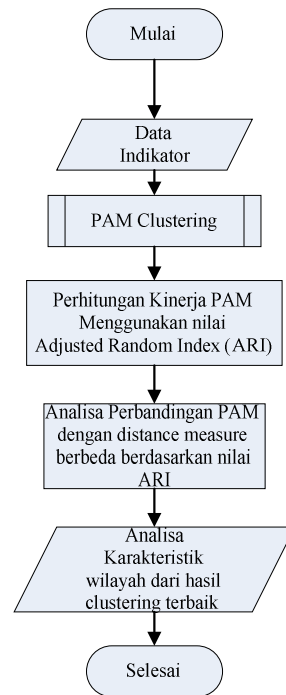
d = jumlah data yang tidak masuk kelompok gold standart dan tidak masuk dalam kelompok hasil clustering.

TABEL 1. TABEL KEMUNGKINAN

Jumlah Point	Cluster yang sama di pengelompokan	Cluster yang berbeda di pengelompokan
Kelas yang sama dan benar	<i>a</i>	<i>c</i>
Kelas yang berbeda dan benar	<i>b</i>	<i>d</i>

III. METODE PENELITIAN

Flowchart sistem ditunjukkan gambar 2 merupakan gambaran global bagaimana sistem akan berjalan. Data dari Dinas Pendidikan yang telah diolah dengan preprocessing data dengan cara melakukan normalisasi terhadap data. Proses sistem dimulai dari proses mengisi data indikator yang diperlukan, kemudian dilanjutkan proses PAM dimana hasilnya adalah *cluster* dan keanggotaannya dilanjutkan dengan menghitung nilai ARI dengan tujuan melakukan validasi terhadap *cluster* yang telah dibuat. Melakukan perbandingan hasil *clustering* dengan perbandingan jarak dalam *distance measure* dilanjutkan dengan analisa karakteristik wilayah dari hasil *cluster* terbaik. Adapun variabel data 18 variabel yang digunakan ditunjukkan Tabel 2.



Gambar 2. *Flowchart* Sistem

Data yang digunakan adalah data pendidikan 72 kecamatan yang diperoleh dari Dinas Pendidikan Kabupaten Bangkalan, Sampang, Pamekasan dan Sumenep tahun 2010 dan 2011.

TABEL 2. TABEL VARIABEL

No	Variabel
1.	APK SD
2.	APM SD
3.	Rasio Murid Guru SD
4.	Rasio Murid Kelas SD
5.	Rasio Murid Sekolah SD
6.	Angka Shift SD
7.	APK SMP
8.	APM SMP
9.	Rasio Murid Guru SMP
10.	Rasio Murid Kelas SMP
11.	Rasio Murid Sekolah SMP
12.	Angka Shift SMP
13.	APK SMA
14.	APM SMA
15.	Rasio Murid Guru SMA
16.	Rasio Murid Kelas SMA
17.	Rasio Murid Sekolah SMA
18.	Angka Shift SMA

Use case diagram merupakan gambaran fungsionalitas dari suatu sistem. Hal-hal yang digambarkan dalam *Use case* antara lain hubungan sistem dengan aktor serta proses sistem yang dilihat dari sudut pandang *user*. Komponen pembentuk *Use case diagram* terdiri dari aktor, *use case*, *relationship*, *system boundary boxes* dan *package*. *Use case diagram* sistem diperlihatkan pada gambar 3. Pada gambar tersebut terdapat 3 aktor, yaitu super admin, admin

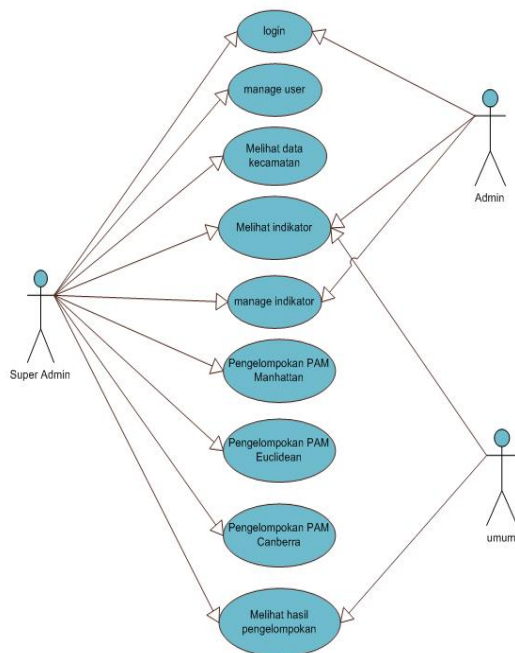
dan umum. Berikut penjelasan dari masing-masing aktor:

1) *Super Admin* : aktor yang memiliki hak akses penuh terhadap sistem. Hak akses penuh terdiri dari

- a) *Manage user*
- b) *Melihat data kecamatan*
- c) *Melihat indikator*
- d) *Manage indikator*
- e) *Pengelompokan PAM Manhattan*
- f) *Pengelompokan PAM Euclidean*
- g) *Pengelompokan PAM Canberra*
- h) *Melihat hasil pengelompokan.*

2) *Admin* : aktor yang memiliki hak untuk manage data indikator.

3) *Umum* : semua pengguna yang hanya dapat melakukan hak akses terhadap informasi yang ada tetapi tidak dapat melakukan manipulasi data. Informasi yang dapat diakses oleh pengguna umum yaitu data indikator dan hasil Clustering.



Gambar 3. Use case diagram

IV. HASIL DAN PEMBAHASAN

A. Penerapan Algoritma PAM

1) *Penentuan Medoid Awal* : Medoid awal (iterasi 1) dan seterusnya pada metode PAM ditentukan secara random ditunjukkan tabel 3.

TABEL 3. MEDOID AWAL

No	Cluster	Medoid
1	Cluster 1	Rubaru
2	Cluster 2	Arjasa
3	Cluster 3	Bangkalan

2) *Menghitung Jarak Data ke Medoid* : Ukuran kemiripan yang digunakan pada metode PAM adalah jarak minimal data dengan medoid. Pada penelitian ini, rumus jarak yang digunakan adalah Manhattan, Euclidean dan Canberra distance. Perhitungan Manhattan distance menggunakan persamaan 1, ditunjukkan tabel 4.

Perhitungan jarak data Masalembu ke medoid cluster 1 (C1) :

$$(109.3-81.18)+(107.12-69.08)+\dots+(1.2-0) = 866.4$$

TABEL 4. TABEL NILAI JARAK MANHATTAN

No	Kecamatan	C1	C2	C3
1.	Bangkalan	1736.41	498.02	0
2.	Socah	469.18	1342.21	1693.79
3.	Burneh	458.66	1274.55	1422.79
...
70.	Arjasa	1319.63	0	498.02
71.	Kangayan	841.46	1424.51	1763.01
72.	Masalembu	866.4	636.99	1027.19

3) *Menentukan medoid baru* : Langkah selanjutnya adalah menentukan medoid baru dengan cara menukar medoid dengan data non-medoid. Atau dengan kata lain melakukan random kembali untuk mendapat medoid baru ditunjukkan tabel 5.

TABEL 5. TABEL MEDOID BARU

No	Cluster	Medoid
1.	Cluster 1	Raas
2.	Cluster 2	Sampang
3.	Cluster 3	Kamal

4) *Menghitung Jarak Total* : Setelah menghitung jarak data dengan medoid, maka diperoleh jarak terkecil tiap data pada satu medoid. Jarak terkecil tiap data tersebut dijumlahkan sehingga diperoleh jarak total.

$$\text{Jarak total} = 0+469.18+458.66+\dots+0+841.46+636.99$$

$$\text{Jarak total} = 37254.049$$

5) *Cek Perubahan Jarak Total* : Lakukan langkah yang sama seperti langkah 2 hingga diperoleh jarak data ke medoid dan jarak terkecil. Kemudian hitung kembali jarak total.

$$\text{Jarak total} = 431.13 + 579.64 + 635.76 + \dots + 259.97 + 888.68 + 473.1$$

$$\text{Jarak total} = 43883.981$$

Jika jarak total baru lebih kecil dari jarak total lama, maka ulangi langkah a sampai e. Sedangkan jika jarak total baru lebih besar dari jarak total lama, maka proses pergantian medoid berhenti.

6) *Kelompokkan Data Berdasarkan Jarak Terkecil* : Aturan pengelompokan data berdasarkan jarak terkecil adalah sebagai berikut

a) Jika min jarak terdapat pada C1 maka data masuk ke cluster 1.

b) Jika min jarak terdapat pada C2 maka data masuk ke cluster 2.

c) Jika min jarak terdapat pada C3 maka data masuk ke cluster 3.

Hasil pengelompokan berdasarkan aturan tersebut ditunjukkan Tabel 6.

TABEL 6. HASIL PENGELOMPOKAN

No	Kecamatan	C1	C2	C3
1.	Bangkalan	1736.41	498.02	0
2.	Socah	469.18	1342.21	1693.79
3.	Burneh	458.66	1274.55	1422.79
...
70.	Arjasa	1319.63	0	498.02
71.	Kangayan	841.46	1424.51	1763.01
72.	Masalembu	866.4	636.99	1027.19

B. Penerapan Adjusted Rand Index(ARI)

Perhitungan *Adjusted Rand Index* (ARI) digunakan untuk mengetahui kinerja tiap metode. Untuk menghitung nilai *Adjusted Rand Index* perlu diketahui nilai *a*, *b*, *c* dan *d* dari hasil pengelompokan menggunakan metode PAM.

Berikut ini contoh perhitungan ARI metode PAM Manhattan menggunakan persamaan 5 ditunjukkan tabel 7.

TABEL 7. PERHITUNGAN ADJUSTED RAND INDEX

Cluster	a	b	c	d
Rendah	47	1	2	22
Sedang	17	2	5	48
Tinggi	1	4	0	67
Jumlah	65	7	7	137

Adapun pseudocode pencarian nilai ARI adalah sebagai berikut :

$$\$ARI_Euclidien = \text{round}(((2 * ((\$STP_euclidien * \$TN_euclidien) - (0 * \$FN_euclidien))) / (((\$STP_euclidien + 0) * (0 + \$TN_euclidien)) + ((\$STP_euclidien + \$FN_euclidien) * (\$FN_euclidien + \$TN_euclidien))))), 3);$$

$$ARI = \frac{2(65 * 137 - 7 * 7)}{(65 + 7)(7 + 137) + (65 + 7)(7 + 137)}$$

$$ARI = \frac{2(8905 - 49)}{(72)(144) + (72)(144)}$$

$$ARI = \frac{2(8905 - 49)}{10368 + 1038}$$

$$ARI = \frac{17712}{20736}$$

$$ARI = 0.854$$

C. Hasil Uji Coba

Pengujian sistem dilakukan sebanyak 4 kali. Pengujian dilakukan lebih dari satu kali karena adanya unsur *random* pada metode PAM. Ujicoba 1 sampai 3 menggunakan data indikator pemerataan pendidikan tahun 2010. Data indikator pendidikan yang digunakan untuk ujicoba telah dikelompokkan secara manual oleh pihak Dinas Pendidikan. Hasil pengelompokan tersebut akan dijadikan sebagai *gold standart* untuk menghitung nilai *adjusted rand index*.

Ujicoba 4 menggunakan dataset *Iris* yang didownload dari *UCI Machine Learning Repository*. Penggunaan dataset iris pada pengujian sistem dilakukan untuk membandingkan kinerja sistem yang menggunakan data riil dan kinerja sistem yang menggunakan dataset.

Ujicoba 5 menggunakan data riil dari dinas pendidikan kabupaten Bangkalan, yang isinya adalah data kecamatan 4 kabupaten di wilayah madura. Tujuannya untuk menguji data wilayah yang sudah diclusterkan tadi validasi riil dapat menunjukkan prosentase ditunjukkan Tabel 8.

TABEL 8. HASIL ANALISA EVALUASI BERDASARKAN ARI

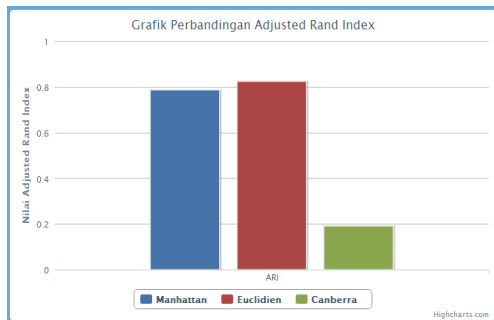
Ujicoba ke-	PAM dengan Distance	ARI
1	Manhattan	0.789
	Euclidean	0.825
	Canberra	0.193
2	Manhattan	0.678
	Euclidean	0.758
	Canberra	0.101
3	Manhattan	0.746
	Euclidean	0.815
	Canberra	0.196
4	Manhattan	0.845
	Euclidean	0.822
	Canberra	0.63

Hasil analisa di atas menunjukkan bahwa kinerja PAM Manhattan dan PAM Euclidean lebih baik dibandingkan PAM Canberra. Dari 4 ujicoba, nilai ARI PAM Canberra selalu bernilai paling kecil ditunjukkan Gambar 4. Dari pengelompokan terbaik diperoleh hasil bahwa:

1) Kelompok tinggi terdiri dari 11 kecamatan yaitu kecamatan Bangkalan, Kamal, Arosbaya, Tanjung Bumi, Sreseh, Torjun, Sampang, Galis, Kalianget, Gapura, dan Arjasa. Kecamatan-kecamatan tersebut memiliki nilai indikator pemerataan yang bagus, sehingga dapat dijadikan contoh untuk kecamatan-kecamatan yang memiliki tingkat pemerataan pendidikan rendah.

2) Kelompok sedang terdiri dari 15 kecamatan yaitu kecamatan Kwanyar, Blega, Camplong, Kedungdung, Banyuates, Pademawu, Pamekasan, Pakong, Waru, Pragaan, Bluto, Kota Sumenep, Ambunten, Gayam dan Masalembu.

3) Kelompok rendah terdiri dari 46 kecamatan yaitu kecamatan-kecamatan selain kecamatan yang termasuk dalam kelompok tinggi dan sedang. Sebagian besar kecamatan yang termasuk dalam kelompok rendah memiliki nilai indikator yang buruk. Hal ini disebabkan oleh tidak tersedianya Sekolah Menengah Atas (SMA) pada kecamatan-kecamatan tersebut.



Gambar 4. Grafik perbandingan hasil *Adjusted Random Index*

V. KESIMPULAN DAN SARAN

Hasil yang diperoleh dari penelitian ini dapat disimpulkan sebagai berikut:

1) Metode *Partition Around Medoid* (PAM) mampu diterapkan untuk memecahkan kasus pengelompokan wilayah / kecamatan berdasar indikator pemerataan pendidikan di 4 wilayah kabupaten di Madura melalui pihak Dinas Pendidikan Kabupaten Bangkalan.

2) Pemilihan algoritma *distance measure* yang tepat untuk kasus pengelompokan memiliki pengaruh significant terhadap hasil *clustering*.

3) Dari 3 kali ujicoba diperoleh rata-rata nilai ARI untuk PAM menggunakan Euclidean *distance* sebesar 0.799, diikuti oleh Manhattan *distance* dengan rata-rata sebesar 0.738 dan rata-rata ARI terendah dimiliki oleh PAM dengan Canberra *distance* sebesar 0.163.

4) Hasil analisa di atas menunjukkan bahwa kinerja PAM Manhattan dan PAM Euclidean lebih baik dibandingkan PAM Canberra.

5) Semakin besar nilai ARI maka semakin bagus kinerja suatu metode *clustering*.

6) Hasil *Clustering* adalah Dari 4 kabupaten dan 72 kecamatan di wilayah Madura, terdapat 11 kecamatan masuk ke dalam *cluster* kategori tinggi, 15 kecamatan masuk ke *cluster* sedang dan 46 kecamatan masuk ke dalam *cluster* rendah, sehingga dari proses *clustering* ini disimpulkan terdapat 46 kecamatan yang memerlukan pembinaan dari 11 kecamatan *cluster* tinggi yang berada diatas dibantu dinas pendidikan.

Adapun hasil *clustering* terbaik yang diperoleh selama ujicoba penelitian ini menghasilkan nilai ARI sebesar 0.825. Hasil tersebut diperoleh pada saat ujicoba pertama menggunakan *Euclidean distance*. Hal ini menunjukkan bahwa kombinasi terbaik dari algoritma dalam penelitian ini adalah

Partition Around Medoid menggunakan *euclidian distance* diikuti validasi *clustering* menggunakan *Adjusted Random Index*.

Adapun Saran kelanjutan penelitian ini adalah mengembangkan metode *validasi clustering* selain *Adjusted Random Index* yang lebih presisi dan terukur.

DAFTAR PUSTAKA

- de Vargas, R. R., & Bedregal, B. R. C. (2013, 15-17 Oct. 2013). *A Way to Obtain the Quality of a Partition by Adjusted Rand Index*. Paper presented at 2nd Workshop-School on the Theoretical Computer Science (WEIT).
- Di, Jia, Jin-feng, Fang, Xue-ping, He, Lu, Meng, & Yi-fei, Zhang. (2013, 23-25 July 2013). *A method of color image edge extraction based on Manhattan distance map*. Paper presented at Ninth International Conference on the Natural Computation (ICNC).
- Galiniun, M., Defindal, I. P., & Melissa, I. (2012, May 30 2012-June 1 2012). *E-learning system introduction: equality in education for teachers in rural area of Indonesia*. Paper presented at International Joint Conference on the Computer Science and Software Engineering (JCSSE).
- Han, Jian-yu, Ding, Jie-chao, & Wang, Jun. (2010, 24-26 Dec. 2010). *Income Gap of Residents, Peasant Income and Economic Growth: An Empirical Study of Rural China Data*. Paper presented at 2010 International Symposium on the Information Science and Engineering (ISISE).
- Julazadeh, A., Marsousi, M., & Alirezaie, J. (2012, 27-30 Nov. 2012). *Classification based on sparse representation and Euclidian distance*. Paper presented at IEEE the Visual Communications and Image Processing (VCIP).
- Mishra, Deepti, & Hiranwal, Saroj. (2014, 1-2 Aug. 2014). *Analysis & implementation of item based collaboration filtering using K-Medoid*. Paper presented at 2014 International Conference on the Advances in Engineering and Technology Research (ICAETR).
- Santos, J. M., & Ramos, S. (2010, Nov. 29 2010-Dec. 1 2010). *Using a clustering similarity measure for feature selection in high dimensional data sets*. Paper presented at the Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on.
- Shaohong, Zhang, & Hau-San, Wong. (2010, 23-26 Aug. 2010). *ARImp: A Generalized Adjusted Rand Index for Cluster Ensembles*. Paper presented at 20th International Conference on the Pattern Recognition (ICPR).
- Sheng-Yi, Jiang. (2006, 13-16 Aug. 2006). *Efficient Classification Method for Large Dataset*. Paper presented at International Conference on the Machine Learning and Cybernetics.
- Xing, Hui, Zhang, Litao, Wang, Yajie, & Zhang, Yanli. (2009, 11-13 Dec. 2009). *Higher Education Equity And Income Distribution: Theory Survey*. Paper presented at International Conference on the Computational Intelligence and Software Engineering CiSE 2009.

Ying-ting, Zhu, Fu-zhang, Wang, Xing-hua, Shan, & Xiao-yan, Lv. (2014, 22-24 Aug. 2014). *K-medoids clustering based on MapReduce and optimal search of medoids*. Paper presented at 9th International Conference on the Computer Science & Education (ICCSE).

Yu-Jie, Hao, Cao, Yang, Shuang-Yan, Quan, & Jian-Ping, Li. (2008, 13-15 Dec. 2008). *Distance Measure between Vague Sets*. Paper presented at International Conference on the Apperceiving Computing and Intelligence Analysis (ICACIA).