ANALISIS DAN PERBANDINGAN ALGORITMA CLUSTERING DALAM PENENTUAN ALOKASI BANTUAN DANA PENDIDIKAN PROPINSI JAWA TENGAH

Yuniaz Hezron Salulolo[1], Andeka Rocky Tanaamah[2], Alz Danny Wowor[3]
[1,2,3]Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana
Jl. Diponegoro 52-60, Salatiga, 50711
[1]682010072@student.uksw.edu, [2]atanaamah@staff.uksw.edu, [3]alzdanny.wowor@staff.uksw.edu

Abstrak

Penentuan alokasi dana pendidikan yang tepat sasaran menjadi kebutuhan dalam pengambilan keputusan oleh pemerintah Propinsi Jawa Tengah. Algoritma clustering k-means dan k-median digunakan untuk mengetahui alokasi kebutuhan dana pendidikan pada kabupaten/kota berdasarkan data penduduk, jumlah guru, dan ruang kelas. Penelitian ini melihat pengaruh nilai centroid awal terhadap banyak iterasi dan banyak anggota pada setiap cluster, dan membandingkan algoritma k-means dan k-median dalam penyelesain masalah. Hasil yang diperoleh nilai centroid awal mempengaruhi banyak iterasi dan banyak anggota pada setiap cluster. Penggunaan nilai kuartil sebagai centroid awal memberikan hasil yang sama pada algoritma k-means dan k-median. Secara keseluruhan k-median lebih baik dari k-mean terutama dalam konsisten data terhadap nilai centroid. Kabupaten Batang, Kab. Karanganyar, Kab.Purworejo, dan Kab.Temanggung menjadi daerah yang sangat membutuhkan bantuan pendidikan.

***Kata Kunci* :** *k-means, k-medians, nilai centroid awal, propinsi Jawa Tengah.*

1. Pendahuluan

Pembangunan pendidikan yang dibangun sejak Indonesia merdeka telah meningkatkan kualitas sumber daya manusia di Indonesia. Namun, masih terdapat persoalan yang belum terselesaikan di dunia pendidikan Indonesia. Salah satunya adalah ketidakmerataan pendidikan pada daerah-daerah di Indonesia yang disebabkan oleh kekurangan tenaga pengajar dan terbatasnya sarana dan prasarana (Annisa, 2014). Hal ini dirasakan besar dampaknya bagi pendidikan dikarenakan tenaga pengajar memiliki peranan penting yaitu sebagai seorang pendidik, sedangkan sarana dan prasarana dibutuhkan untuk menunjang guru dan siswa dalam melakukan proses belajar mengajar.

Terkait permasalahan tersebut maka perlu dilakukan pemberian berupa bantuan operasional atau penambahan tenaga pengajar kepada daerah-daerah yang masih mengalami kekurangan tenaga pengajar juga sarana dan prasarana. Tercatat bebarapa daerah di Propinsi Jawa Tengah yang masih mengalami kekurangan tenaga pendidikan. Salah satu contohnya yaitu Kabupaten Semarang, dimana masih terjadi kekurangan sebanyak 900 tenaga guru pada sekolah dasar (Raharjo dan Akbar, 2016). Namun yang menjadi kendala yaitu, bagaimana menentukan daerah di Propinsi Jawa Tengah yang masih mengalami kekurangan pengajar serta sarana dan prasarana. Oleh karena itu dibutuhkan sebuah teknik atau metode yang mendukung penyelesaian permasalahan yang ada. Salah satu teknik yang dapat digunakan adalah *clustering*. Dimana algoritma *clustering* dapat membantu dalam mengelompokkan data berdasarkan kemiripan karakteristik data.

Clustering merupakan proses pengelompokkan objek kedalam sebuah *cluster*, dimana objek dalam suatu *cluster* saling memiliki kemiripan, namun berbeda jauh dengan objek yang ada pada *cluster* lain. Terdapat dua metode analisis yang dikenal dalam *clustering*, yaitu metode hirarki *clustering* dan metode parsial *clustering*. Metode hirarki *clustering* merupakan proses pengelompokkan data pada suatu bagan yang bersifat hirarki, dimana terdapat penggabungan dua grup yang terdekat pada setiap iterasinya. Sedangkan metode parsial, data dikelompokkan kedalam sejumlah *cluster* tanpa

adanya struktur hirarki antara satu dengan lainnya (Agusta, 2007; Irwansyah dan Faisal, 2015).

Penelitian ini akan melakukan *clustering* menggunakan beberapa algoritma yaitu algoritma k-means, dan k-median pada data pendidikan Propinsi Jawa Tengah. Data yang akan digunakan merupakan dataset kabupaten/kota pada Propinsi Jawa Tengah berdasarkan penduduk usia sekolah, guru, dan ruang kelas. Data tersebut nantinya akan di*clustering*kan berdasarkan algoritma masing-masing. Hasil dari *clustering* menggunakan beberapa algoritma ini nantinya dapat dijadikan sebagai acuan oleh pihak pemerintah sebagai pertimbangan pengambilan keputusan terkait penyaluran bantuan pendidikan di Propinsi Jawa Tengah.

2. Tinjauan Pustaka

Penelitian terdahulu yang dijadikan sebagai acuan dalam penelitian ini berjudul Analisis klaster k-means dan k-median Pada Data Indikator Kemiskinan. Pada penelitian ini digunakan analisis diskriminan sebagai metode pembanding dalam mengetahui ketepatan *cluster* antara metode k-means dan k-median menggunakan data indikator kemiskinan kabupaten di Indonesia tahun 2009. hasil dari penelitian ini dijelaskan bahwa metode k-means lebih unggul berdasarkan nilai klasifikasi (Febriyana, 2011).

Penelitian lain dengan judul Analisa Perbandingan Metode Hierarchical Clustering, K-means dan Gabungan Keduanya dalam Cluster Data (Studi kasus : Problem Kerja Praktek Jurusan Teknik Industri ITS). Pada penelitian ini dilakukan beberapa pengujian performa *cluster*, salah satunya yaitu *Cluster Variance* terhadap tiga metode *clustering* yang berbeda yaitu metode hierarki *clustering*, k-means *clustering* dan kombinasi algoritma *hierarchical clustering* dengan K-means. Berdasarkan hasil uji permforma yang dilakukan didapatkan hasil kombinasi algoritma *hierarchical clustering* dengan k-means menghasilkan pengelompokan data yang lebih baik dibandingkan algortima k-means (Alfina dkk, 2012).

Berdasarkan penelitihan terdahulu yang dijadikan sebagai acuan seperti yang dipaparkan diatas, pada penelitian-penelitian yang sudah dilakukan sebelumnya perbandingan algoritma *clustering* dilakukan dengan menguji performa dan validitas pada *cluster*. Namun pada penelitian ini, perbandingan pada algoritma yang digunakan yaitu algoritma *clustering* k-means dan k-median dilakukan dengan pengujian terhadap nilai centroid untuk melihat seberapa besar pengaruh nilai yang diberikan terhadap *cluster* yang dibentuk.

Clustering merupakan proses partisi satu set objek data kedalam himpunan bagian yang disebut *cluster*. Objek dalam sebuah *cluster* memiliki kemiripan karakteristik antara satu dengan yang lain dan akan berbeda dengan *cluster* yang lain. Kemiripan karakteristik dalam suatu *cluster* diukur secara numerik menggunakan pengukuran kesamaan dengan membandingkan jarak antara objek. Dimana semakin kecil jarak antara objek, maka semakin tinggi kemiripan karakteristik objek tersebut (Agusta, 2007; Irwansyah dan Faisal, 2015). Terdapat beberapa pengukuran jarak yang dapat digunakan dalam *clustering*, salah satunya *Euclidean Distance.*

Euclidean Distance merupakan perbandingan jarak dua buah objek dengan mengetahui nilai dari masing-masing atribut pada kedua objek tersebut. Pengukuran jarak *euclidean* didefinisikan sebagai berikut:

$$d_{ij} = \left(\sum_{j=1}^{p} \left(\overline{X}_i - \overline{X}_j \right)^2 \right)^{\frac{1}{2}} \quad (1)$$

Dimana (d_{ij}) adalah jarak antara objek *i* ke objek *j*, $\overline{X}_i$ adalah nilai tengah gerombol ke-*i*, $\overline{X}_j$ adalah nilai tengah gerombol ke-*j*, dan *p* adalah banyaknya peubah yang diamati (Lathifaturrahmah, 2010).

K-means merupakan teknik pengelompokkan non-hirarki yang sering digunakan dalam membagi data kedalam sebuah *cluster*/kelompok. Langkah-langkah *clustering* data menggunakan algoritma k-means secara umum didefinisikan sebagai berikut (Agusta, 2007; Aggarwal dan Reddy, 2013):

1. Tentukan k sebagai *cluster* yang ingin dibentuk.
2. bangkitkan k *centroid* (titik pusat) awal secara random.
3. Hitung jarak setiap data ke masing-masing pusat *cluster* dengan menggunakan *euclidean distance.*

4. kelompokkan setiap data berdasarkan jarak terdekat antara data dengan pusat *cluster.*
5. tentukan posisi pusat *cluster* baru dengan cara menghitung nilai rata-rata dari data yang ada pada pusat *cluster* yang sama.
6. Kembali ke langkah 3, apabila masih terdapat data yang berpindah *cluster.*

K-Median merupakan salah satu teknik pengelompokkan data, dimana setiap proses atau tahapan yang dilakukan sama seperti teknik pengelompokkan pada k-means. Jika pada proses k-means pengelompokkan dihitung berdasarkan nilai rata-ratanya, pada proses k-median pengelompokkan dihitung berdasarkan nilai median. Algoritma *clustering* k-median memilih k sebagai pusat *cluster* dengan tujuan untuk meminimalkan jumlah ukuran jarak setiap titik dari pusat *cluster* terdekat. Misalkan terdapat ${}_nX_p$ gugus data yang mempunyai *n* objek dan *p* peubah. Jarak antara objek ke-*i*, x_i dan objek ke-*j*, x_j, dinotasikan dengan $d(i,j)$. Dalam pemilihan suatu objek yang representatif dalam suatu *cluster* (median awal), y_i didefinisikan sebagai variabel biner 0 dan 1, dimana $y = 1$ jika dan hanya jika objek *i* (i= 1,2,...,n) dipilih sebagai median awal. Penempatan setiap objek ke-*j* ke salah satu median awal dituliskan sebagai z_{ij}, dimana z_{ij} didefinisikan sebagai variabel biner 0 dan 1.z_{ij} bernilai 1 jika dan hanya jika objek *j* ditempatkan ke *cluster*, dimana objek *i* sebagai median awal. Model optimasi k-median didefinisikan sebagai berikut (Aggarwal dan Reddy, 2013; Flowrensia, 2010) :

$$\text{minimize } \sum_{j=1}^{n}\sum_{j}^{n} d(i,j)z_{ij} \quad (2)$$

dimana :

$$\sum_{i=1}^{n} z_{ij} = 1 \qquad j = 1, 2,...,n \quad (3)$$

$$z_{ij} \le y_i, \qquad i, j = 1, 2,..,n \quad (4)$$

$$\sum_{i=1}^{n} y_i = k, \qquad k = \text{Jumlah gerombol} \quad (5)$$

$$y_i, z_{ij} \in \{0,1\}, \quad i, j = 1, 2,...,n \quad (6)$$

Persamaan (3) menyatakan bahwa setiap objek *j* harus ditempatkan ke hanya satu median awal. Persamaan (3) dan (6) berimplikasi bahwa untuk suatu *j*, z_{ij} akan bernilai 1 atau 0. Persamaan (5) menyatakan bahwa hanya ada *k* objek yang akan dipilih sebagai median. Langkah-langkah dasar algoritma k-median didefinisikan sebagai berikut;

1. Tentukan jumlah *cluster* yang ingin dibentuk.
2. Alokasi data kedalam *cluster* secara random.
3. Hitung jarak setiap data ke masing-masing pusat *cluster* dengan menggunakan *euclidean distance.*
4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan pusat *cluster.*
5. tentukan posisi pusat *cluster* baru (C_{kj}) dengan cara menghitung nilai median dari data yang ada pada pusat *cluster* yang sama.
6. Kembali ke langkah 3 apabila masih terdapat data yang berpindah *cluster.*

3. Metode Penelitian

3.1 Data

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa data baseline Propinsi Jawa Tengah tahun 2008. Digunakan tiga variabel yang digunakan untuk menguji pengaruh nilai centroid terhadap banyak iterasi dan banyak angota pada setiap *cluster*, sehingga dapat dijadikan sebagai referensi untuk menentukan kab/kota di Propinsi Jawa Tengah yang dapat diberikan alokasi dana pendidikan.

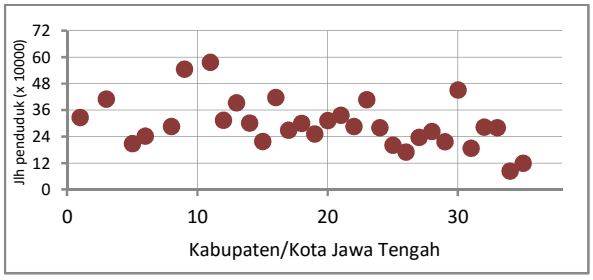
Propinsi Jawa Tengah terdiri dari 35 Kabupaten/Kota (Kab/kota), untuk memudahkan dialam visualisasi diagram (*Scatter Plot*) maka diberikan indeks sebagai subtitusi pengganti kab/kota pada Tabel 1.

Tabel 1, Centroid Berdasarkan Nilai Kuartil pada Setiap Cluster dengan Metode K-Means

Index	Kabupaten / Kota	Index	Kabupaten / Kota
1	Kab. Banjarnegara	19	Kab. Pemalang
2	Kab. Banyumas	20	Kab. Purbalingga
3	Kab. Batang	21	Kab. Purworejo
4	Kab. Blora	22	Kab. Rembang
5	Kab. Boyolali	23	Kab. Semarang
6	Kab. Brebes	24	Kab. Sragen
7	Kab. Cilacap	25	Kab. Sukoharjo
8	Kab. Demak	26	Kab. Tegal
9	Kab. Grobogan	27	Kab. Temanggung
10	Kab. Jepara	28	Kab. Wonogiri
11	Kab. Karanganyar	29	Kab. Wonosobo
12	Kab. Kebumen	30	Kota Magelang
13	Kab. Kendal	31	Kota Pekalongan
14	Kab. Klaten	32	Kota Salatiga

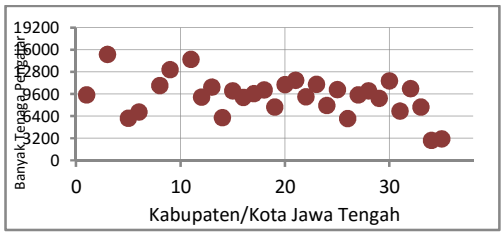
15	Kab. Kudus	33	Kota Semarang
16	Kab. Magelang	34	Kota Surakarta
17	Kab. Pati	35	Kota Tegal
18	Kab. Pekalongan		

Data Penduduk Propinsi Jawa Tengah pada tahun 2008 diberikan pada Gambar 1. Diagram Scatter digunakan sehingga dapat dilihat variasi data dan penyebarannya.



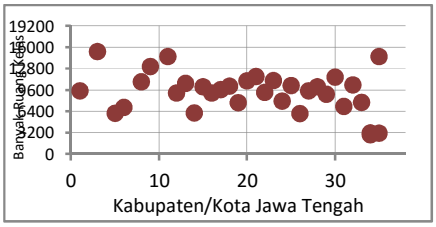
Gambar 1. Data Penduduk Jawa Tengah 2008 Berdasarkan Kabupaten/Kota

Banyak tenaga pengajar diperlukan untuk mengatahui kebutuhan kab/kota terhadap banyak penduduk usia sekolah. Data banyak pengajar diberikan di Gambar 2.



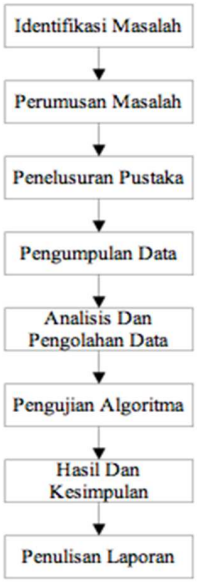
Gambar 2. Data BanyakTenaga Pengajar Jawa Tengah 2008 Berdasarkan Kabupaten/Kota

Ruang kelas menjadi variabel yang ketiga dalam penentuan alokasi dana pendidikan, data diberikan pada Gambar 3.



Gambar 3. Data Banyak Ruang Kelas Jawa Tengah 2008 Berdasarkan Kabupaten/Kota

3.2 Tahapan Penelitian



Gambar 4. Tahapan Penelitian

Terdapat beberapa tahapan yang akan dilakukan dalam penelitian ini. Seperti yang dapat dilihat pada Gambar 4. Tahapan pertama yang dilakukan dalam penelitian ini yaitu identifikasi malasalah. Identifikasi masalah dilakukan untuk melihat tujuan dan sasaran yang ingin dicapai dari penelitian. Tahapan kedua yaitu perumusan masalah. Pada tahapan ini masalah yang telah ditentukan sebelumnya perlu dijawab atau dijelaskan solusi pemecahan masalahnya. Tahap selanjutnya yaitu penelusuran pustaka. Pada tahapan ini informasi serta sumber pustaka yang berkaitan dengan penelitian dikumpulkan untuk memperkuat landasan teori dalam penelitian ini, atau bisa juga digunakan untuk menghindari duplikasi penelitian. Tahap selanjutnya adalah tahap pengumpulan data. Pada tahap ini data yang digunakan pada penelitian ini yaitu data sekunder berupa data baseline Propinsi Jawa Tengah. Tahapan berikutnya yaitu analisis dan pengolahan data. Dalam tahapan ini data yang telah dikumpulkan dianalisis dan kemudian diolah menggunakan algoritma *clustering.* Tahap selanjutnya yaitu pengujian algoritma. Pada tahap ini, algoritma *clustering* k-means dan k-median

yang ada akan diuji dengan membangkitkan nilai centroid dengan menggunakan nilai kuartil dan nilai-nilai random secara berturut-turut dengan tujuan untuk menguji pengaruh nilai centroid terhadap penentuan anggota didalam *cluster*. Tahapan selanjutnya adalah hasil dan kesimpulan. Pada tahapan ini ditarik kesimpulan dari setiap hasil tahapan yang telah dilakukan sebelumnya. Selanjutnya merupakan tahapan terakhir dalam penelitian ini yaitu penulisan laporan. Pada tahapan ini seluruh proses dan hasil dari penelitian dijabarkan ke dalam bentuk tulisan berupa laporan penelitian (Hasibuan, 2007).

4. Hasil dan Pembahasan

Penelitian ini menguji metode *clustering* yang dapat digunakan sebagai rekomendasi dalam menentukan keputusan alokasi bantuan dana pendidikan di Propinsi Jawa Tengah. Metode yang akan diujikan adalah algoritma *clustering* k-means dan k-median, dengan menggunakan data penduduk, jumlah pengajar, dan ruang kelas.

Pengujian dilakukan pada kedua algoritma menggunakan nilai *centroid* untuk melihat pengaruhnya pada banyak iterasi yang diperlukan. Secara teoritis, nilai *centroid* pada metode *cluster* ditentukan secara acak dengan interval pada nilai minimum dan nilai maksimum. Asumsi yang dibangun adalah penentuan nilai *centroid* terbaik akan berpengaruh pada proses iterasi optimum.

Setiap data set akan mempunyai nilai sari data yang digunakan sebagai sumber informasi. Nilai kuartil dipilih sebagai sari data, karena secara teori dapat membagi data menjadi empat bagian yang sama. Secara kebutuhan *cluster* dibagi menjadi tiga, dan masing-masing *cluster* memerlukan nilai centroid. Dalam kuartil terdapat tiga nilai yaitu Q_1, Q_2, dan Q_3. Dimana secara statistik nilai kuartil diperoleh karena data telah diurutkan, dan setiap nilai kuartil adalah nilai tengah dari pembagian dari data utama.

Pengurutan data dari sebuah data set, yang digunakan sebagai nilai *centroid* awal. Nilai kuartil dari setiap data jumlah penduduk, ruang kelas, dan jumlah guru secara berturut-turut menjadi nilai *centroid* pada *cluster* 1, *cluster* 2, dan *cluster* 3 diberikan pada Tabel 2.

Tabel 2, Centroid Berdasarkan Nilai Kuartil pada Setiap Cluster dengan Metode K-Means

Nilai Kuartil			
	Centroid 1	**Centroid 2**	**Centroid 3**
cluster 1	212,842	3,615	7,444
cluster 2	281,350	6,720	9,494
cluster 3	332,343	10,124	10,740

Tabel 3, Centroid Berdasarkan Nilai Random 1 pada Setiap Cluster dengan Metode K-Means

Nilai Random 1			
	Centroid 1	**Centroid 2**	**Centroid 3**
cluster 1	189,900	13,783	10,543
cluster 2	570,670	16,234	6,980
cluster 3	421,490	9,780	3,636

Tabel 4, Centroid Berdasarkan Nilai Random 2 pada Setiap Cluster dengan Metode K-Means

Nilai Random 2			
	Centroid 1	**Centroid 2**	**Centroid 3**
***cluster* 1**	397,000	10,392	13,651
***cluster* 2**	151,433	16,263	13,039
***cluster* 3**	190,358	3,012	3,416

Tabel 5, Centroid Berdasarkan Nilai Random 3 pada Setiap Cluster dengan Metode K-Means

Nilai Random 3			
	Centroid 1	**Centroid 2**	**Centroid 3**
***cluster* 1**	379,782	3,733	14,742
***cluster* 2**	156,280	3,526	14,871
***cluster* 3**	174,661	14,570	9,947

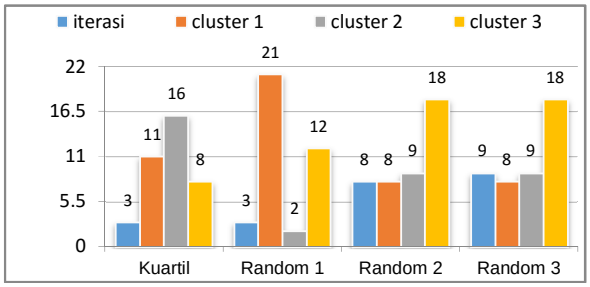
Nilai awal *centroid* awal juga dibangkitkan secara random yang diberikan pada Tabel 3, Tabel 4, dan Tabel 5. Nilai kuartil dan nilai random

digunakan untuk memperoleh iterasi yang optimum pada sebuah algoritma *clustering* dan juga melihat kemampuan algoritma berdasarkan inputan dengan jumlah hasil anggota pada sebuah *cluster*.

Penentuan optimal suatu *clustering* dilihat dari banyak iterasi yang diperlukan, dimana iterasi ke-(n−1) dan iterasi ke-n mempunyai hasil *clustering* yang sama. Pengujian untuk kedua algoritma dilakukan dengan input nilai centroid awal.

4.1 Pengujian *Clustering* k-means

Penggunaan metode *cluster* k-means dengan perhitungan jarak *euclidean distance* seperti pada persamaan (1). Hasil yang diperoleh berdasarkan nilai centroid awal pada Tabel 2, Tabel 3, Tabel 4, dan Tabel 5 diberikan berupa histogram pada Gambar 5.



*Gambar 5.*Banyak iterasi dan Kab/Kota Jawa Tengah pada tiap *cluster*

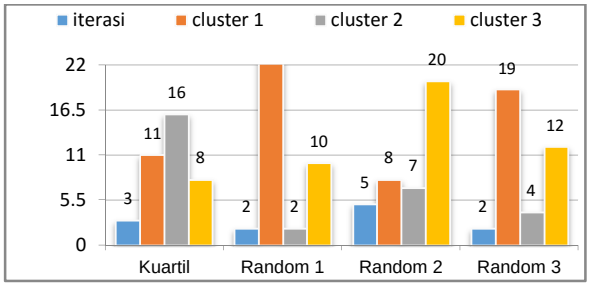
Setiap kelompok terdapat empat histogram, yang terdiri dari banyak iterasi yang diperlukan adalah histogram pertama, histogram kedua adalah banyak anggota pada *cluster* 1, hal yang sama secara berturut-turut untuk setiap histogram berikutnya menujukkan banyak anggota pada *cluster* 2 dan *cluster* 3.

Sebagai contoh, pada inputan centroid dengan nilai kuartil, diperlukan tiga iterasi dan memperoleh sebelas kab/kota pada berada pada *cluster* 1, enam kab/kota di *cluster* 2, dan 8 kab/kota di *cluster* 3. Bersasarkan hasil ini, pengambil keputusan menggunakan *cluster* pertama sebagai sebagai acuan maka dapat diputuskan sebelas kab/kota yang memerlukan bantuan dana pendidikan di Propinsi Jawa Tengah.

4.2 Pengujian *Clustering* k-median

Pengujian pada algoritma k-median analog dengan dilakukan pada algoritma k-means. Berdasarkan nilai centroid awal yang dipilih pada Tabel 2, Tabel 3, Tabel 4, dan Tabel 5. Hasil secara lengkap disajikan dalam histogram yang ditunjukkan pada Gambar 6.

Clustering k-median mempunyai banyak iterasi yang lebih sedikit dibandingkan dengan k-means. Hal ini terlihat pada nilai centroid dengan random 1 dan random 3 hanya memerlukan dua iterasi sudah memperoleh hasil yang optimum.

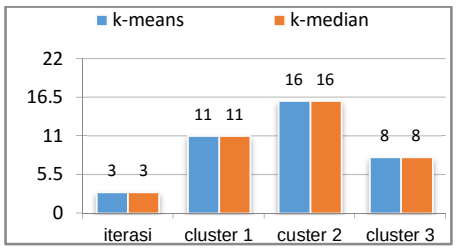


Gambar 6. Pengujian *Clustering* K-median

4.3 Pengujian Centroid

Pengujain terhadap nilai centroid awal dilakukan dengan membandingkan secara langsung antara kedua metode. Nilai yang dikaji adalah banyak iterasi, dan banyak anggota pada setiap *cluster*. Penentuan banyak iterasi menjadi salah satu kajian penting dari penelitian ini, karena dari banyak proses akan mengurangi efesiensi sebuah algoritma untuk menemukan sebuah solusi.

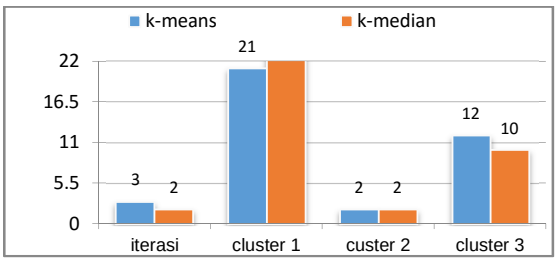
Banyak anggota pada sebuah *cluster* juga menajdi kajian penting yang kedua. Kajian ini dipandang penting karena dalam kasus alokasi bantuan dana pendidikan, terlalu banyak atau terlalu sedikit anggota pada sebuah *cluster* tentu bukan sebuah solusi yang tepat untuk ditawarkan pada pemerintah Jawa Tengah.



Gambar 7. Pengujian Berdasarkan Nilai Kuartil

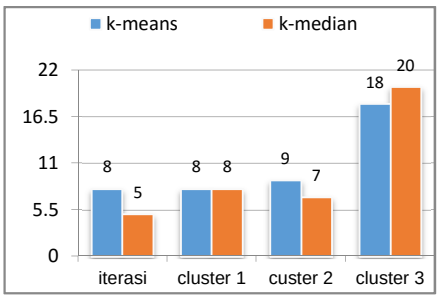
Perbandingan berdasarkan nilai kuartil yang diberikan pada Gambar 7, kedua algoritma mempunyai hasil yang seimbang untuk proses iterasi atau banyak elemen pada setiap *cluster*. Apabila yang dikaji banyak terhadap penentuan alokasi dana pendidikan, maka akan terdapat sebelas kabupaten/kota di Propinsi Jawa Tengah yang membutuhkan bantuan pendidikan. Pengujian menggunakan nilai kuartil memberikan hasil yang sama menujukkan bahwa setiap algoritma mempunyai kekuatan yang seimbang.

Apabila dikaji lebih dalam, salah satu faktor yang membuat kedua algoritma memberikan hasil yang sama yaitu nilai kuartil karena merupakan nilai tengah yang membagi data terurut menjadi empat bagian yang sama. Algoritma k-means dan k-median secara rumus hampir sama, perbedaannya hanya pada penggunaan nilai mean atau nilai median. Secara teori, nilai mean dan median dari sebuah data selalu mempunyai hasil yang tidak terlalu jauh jaraknya. Tetapi argumen ini menjadi tidak tepat, karena pada inputan dengan nilai random 1, random 2, dan random 3 memberikan hasil yang berbeda seperti yang ditunjukkan pada Gambar 8, Gambar 9, dan Gambar 10. Hasil inputan dengan nilai random 1, algoritma k-median memerlukan iterasi yang lebih sedikit. Sedangkan untuk banyak anggota pada setiap *cluster* memperoleh hasil yang bervariasi, dimana k-median lebih banyak pada *cluster* 1 sedangkan sebaliknya k-means lebih banyak pada *cluster* 3, sedangkan pada *cluster* 2 keduanya mempunyai hasil yang sama.



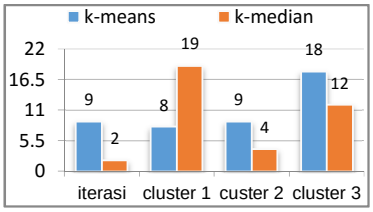
Gambar 8. Pengujian Berdasarkan Nilai Random 1

Gambar 9 merupakan perbandingan dengan nilai random 2 sebagai centroid. Pada Kebutuhan iterasi untuk mencapai hasil optimum metode k-means memerlukan 23% lebih banyak dari k-median. Sedangkan untuk banyak elemen pada kedua metode pada *cluster* 1 menghasilkan nilai yang sama. Banyak elemen pada *cluster* 2, k-median lebih banyak dua kabupaten atau sekitar 13% dari jumlah algoritma k-means. Sedangkan pada *cluster* 3 algoritma k-median lebih banyak anggota dibandingkan k-means.



Gambar 9. Pengujian Berdasarkan Nilai Random 2

Hasil yang diperoleh dengan nilai random 3, memperoleh hasil yang sangat bervariasi. Iterasi yang diperlukan k-median jauh lebih baik dibandingkan dengan k-means yang memerlukan sembilan iterasi. Hasil selengkapnya ditunjukkan pada Gambar 10.



Gambar 10. Pengujian Berdasarkan Nilai Random 3

4.4 Analisis Hasil Pengujian

Analisis terhadap nilai *cluster* merupakan interpretasi terhadap nilai centroid yang menjadi inputan pada kedua algoritma *clustering*, sehingga memberikan pengaruh pada banyak iterasi yang diperlukan dan banyak elemen dari sebuah *cluster*.

Nilai centroid pertama diberikan pada Gambar 7. Pengujian berdasarkan nilai kuartil memberikan hasil yang unik, yaitu pada kedua algoritma memberikan hasil yang sama. Apabila dikaji lebih dalam, salah satu penyebab penggunaan nilai kuartil dari setiap variabel pada masing-masing *cluster*. Misalnya nilai kuartil pertama pada variabel banyak penduduk adalah 212,842, sedangkan untuk banyak ruang kelas adalah 3,615 dan variabel banyak tenaga pengajar adalah 7,444 semuanya digunakan sebagai centroid pada *cluster 1*. Hal yang sama untuk *cluster* 2 dan *cluster* 3 digunakan nilai kuartil 2 dan kuartil 3 dari masing-masing variabel.

Nilai kuartil merupakan nilai tengah yang membagi data terurut menjadi empat bagian yang sama. Algoritma k-means dan k-median secara rumus hampir sama, perbedaannya hanya pada penggunaan nilai mean atau nilai median sebagai centroid baru pada ieterasi ke-2 dan seterusnya. Secara teori, nilai mean dan median dari sebuah data selalu mempunyai hasil yang tidak terlalu jauh jaraknya. Sehingga penggunaan nilai centroid yang baru tidak akan membrikan hasil yang berbeda.

Argumen ini terkait penggunaan nilai input yang sama pada kedua algoritma *clustering* akan memberikan hasil yang sama karena secara statistik nilai median dan mean selalu berdekatan menjadi tidak tepat, karena pada inputan dengan nilai random 1, random 2, dan random 3 memberikan hasil yang berbeda seperti yang ditunjukkan pada Gambar 8, Gambar 9, dan Gambar 10. Sehingga simpulan sementara untuk kondisi inputan yang sama pada kedua metode belum tentu memberikan hasil yang sama, dapat juga mengeluarkan hasil yang berbeda sangat signifikan seperti pada Gambar 10.

Nilai centroid awal yang diinputkan akan memberikan pengaruh yang sangat besar terhadap penyelesaian metode *clustering*. Keunikan hasil yang diperoleh pada Gambar 7, Gambar 8, Gambar 9, dan Gambar 10 menujukkan perbedaan kondisi awal dimana perubahan hasil dan bentuk pola yang dapat digeneralisasi sebagai sebuah hubungan antar input dan output. Kondisi yang dapat diperhatikan adalah penggunaan nilai kuartil dan nilai acak. Nilai kuartil adalah nilai yang diambil dari data yang sudah terurut, dan kuartil juga merupakan nilai tengah dari setiap bagian data. Pengambilan nilai kuartil i yang seragam pada *cluster* i (untuk i = 1, 2, dan 3) juga membarikan pengaruh. Misalnya kuartil 1 untuk *cluster* 1 pada setiap variabel, membuat jarak yang diperoleh dengan rumus *euclidean* tidak berbeda jauh karena centroid sudah diposisikan di tengah data.

Hasilnya tentu akan berbeda untuk nilai acak. Nilai yang dibangkitkan secara random, membuat keduanya tidak memberikan hasil yang konstan. Pengambilan nilai yang acak akan mengakibatkan jarak nilai centroid terhadap data yang diukur dapat menjadi lebih jauh atau bahkan lebih dekat. Kondisi nilai acak kadang tidak memberikan pengaruh karena secara kebetulan terpilih bilangan yang secara jarak dan posisi tepat berada ditengah dari data yang ada. Hal ini yang membuat centroid awal dengan kuartil akan memberikan hasil yang lebih terukur, dibandingkan dengan pengambilan secara acak.

4.5 Penentuan Algortima Terbaik

Berdasarkan nilai centroid awal pada metode k-means dengan menggunakan nilai kuartil atau nilai acak akan memberikan pengaruh terhadap banyak proses iterasi dan banyak anggota pada setiap *cluster*.

Tabel 6, Nilai Rata-rata Banyak Iterasi dan Banyak Elemen dalam Centroid Berdasarkan Algotima

Algoritma	Iterasi	*Cluster* 1	*Cluster* 2	*Cluster* 3
k-means	5,75	12	9	14
k-median	3,75	11,75	13,5	9,75

Terdapat dua hal penting yang diperhatikan untuk menentukan algoritma terbaik. Pertama

pengaruh nilai centroid terhadap banyak iterasi yang diperlukan. Hal kedua, nilai centroid terhadap banyak elemen pada setiap *cluster*. Hasil dari kedua algortima secara rata-rata diberikan pada Tabel 6. Pada table 6 dijelaskan bahwa algortima k-median lebih unggul secara berdasarkan banyak iterasi yang diperlukan untuk mendapatkan hasil optimum. Rata-rata anggota pada *cluster* 1, k-median juga memperoleh hasil yang lebih sedikit. Walaupun pada *cluster* 2 algoritma k-means unggul dari banyak elemen/anggota pada setiap *cluster*.

Banyak elemen atau anggota pada setiap *cluster* berdasarkan algoritma tetap diperhatikan, karena sangat memberikan pengaruh yang cukup signifikan dalam pengambilan keputusan. Pada kasus pemberian bantuan dana pendidikan pada setiap kab/kota di propinsi Jawa Tengah, menjadi tidak mungkin bagi pemerintah Jateng untuk memberikan bantuan pada kab/kota yang berjumlah terlalu banyak atau terlalu sedikit karena akan menjadi tidak efisien.

Penelitian ini menggunakan tiga *cluster* untuk menentukan daerah kab/kota untuk alokasi dana pendidikan. Bila di bagi berdasarkan rating untuk masing-masing *cluster* maka *cluster* 1 menjadi sangat penting, *cluster* 2 dengan rating penting, dan *cluster* 3 sebagai rating kurang penting. Berdasarakan hal tersebut maka metode k-median lebih baik atau lebih efisien dalam menentukan keputusan dibandingkan dengan k-means.

4.6 Penentuan Daerah Alokasi Dana Pendidikan

Penentuan daerah sebagai alokasi dana pendidikan dilakukan dengan mengambil hasil pada *cluster* 1 yang dikategorikan sangat penting. Penelitian ini menggunakan empat inputan pada dua algoritma, sehingga akan menghasilkan delapan hasil *cluster* 1. Sehingga dilihat secara irisan dengan memperhatikan daerah kab/kota yang berada pada setiap *cluster* 1. Terdapat enam kab/kota yaitu Kabupaten Batang, Kab. Karanganyar, Kab. Purworejo, dan Kab. Temanggung menjadi daerah yang sangat membutuhkan bantuan pendidikan.

5. Kesimpulan

Simpulan yang dapat diambil dari penelitian ini adalah

1. Penentuan nilai centroid awal pada algoritma k-means dan k-median akan mempengaruhi banyak proses iterasi dan banyak elemen pada setiap *cluster*.
2. Penggunaan nilai kuartil sebagai centroid awal dari data pada masing-masing variabel memberikan hasil yang sama untuk kedua algoritma yang berbeda.
3. Berdasarkan banyak iterasi dan perolehan banyak anggota pada setiap *cluster* dan konsistensi dan efesiensi terhadap nilai centroid maka algortima k-median lebih baik dari algoritma k-means.
4. Kabupaten/Kota di Propinsi Jawa Tengah yang membutuhkan bantuan dana pendidikan adalah Kabupaten Batang, Kab. Karanganyar, Kab. Purworejo, dan Kab. Temanggung.

Daftar Pustaka

Aggarwal, Charu C., & Reddy, Chandan., (Ed.), (2013), Data *Cluster*ing: Algorithms and Applications, Boca Raton : CRC Press

Agusta, Y., (2007). K-Means-Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem dan Informatika*, Vol.3, 47-60

Alfina, T., Santosa, B., Barakbah, A.R. (2012). Analisa Perbandingan Metode Hierarchical Clustering, K-means dan Gabungan Keduanya dalam Cluster Data (Studi kasus : Problem Kerja Praktek Jurusan Teknik Industri ITS), *Jurnal Teknik ITS*, Vol.3, 521-525.

Annisa, (2014). Potret Pendidikan di Indonesia. Seputar Malang. 19 Agustus 2014.

Febriyana. (2011). *Analisis Klaster K-Means dan K-Median Pada Data Indikator Kemiskinan*. Jakarta: Universitas Islam Negeri Syarif Hidayatullah.

Flowrensia,Y., (2010). Perbandingan Penggerombolan K-Means dan K-Medoid Pada Data Yang Mengandung Pencilan (Skripsi S1, Universitas Pertanian Bogor), dari IPB Repository : http://repository.ipb.ac.id/

Hasibuan, Zainal A, (2007), *Metode Penelitian Pada Bidang Ilmu Komputer Dan Teknologi Informasi: Konsep,*

Teknik Dan Aplikasi, Jakarta : Fakultas Ilmu Komputer Universitas Indonesia.

Irwansyah, E., Faisal, M., (2015). Advance *Cluster*ing: Teori dan Aplikasi. Jakarta: Bina Nusantara University.

Lathifaturrahmah., (2010). Perbandingan Hasil Penggerombolan Metode *k-means*, *Fuzzy k-means*, dan *Two Step Cluster*, Bogor: Institut Pertanian Bogor.

Raharjo, A., Akbar, M., (2016). Kabupaten Semarang Kekurangan Guru. *Republika*, 2 Agustus 2016.

Biodata Penulis

Yuniaz Hezron Salulolo, menempuh pendidikan di Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana.

Andeka Rocky Tanaamah, memperoleh gelar S1 di Universitas Kristen Satya Wacana. Memperoleh gelar S2 di Universitas Gadjah Mada. Saat ini menjadi pengajar di Universitas Kristen Satya Wacana.

Alz Danny Wowor, memperoleh gelar S1 di Universitas Kristen Satya Wacana. Memperoleh gelar S2 di Universitas Kristen Satya Wacana. Saat ini menjadi pengajar di universitas Kristen Satya Wacana.

BERITA ACARA PELAKSANAAN HASIL SEMINAR SESI PARALEL KNASTIK 2016

Judul : Analisis dan Perbandingan Algoritma Clustering dalam Penentuan Alokasi Bantuan Dana Pendidikan Propinsi Jawa Tengah

Pemakalah : Yuniaz Hezron Salulolo, Andeka Rocky Tanaamah, Alz Danny Wowor

Moderator : Gloria Virginia, S.Kom., MAI, Ph.D

Notulis : Yoas

Peserta : 11 orang di ruang : D.3.2

Tanya Jawab :

Penyaji: Yuniaz Herzon Salulolo (UKSW)

Pertanyaan (Dari sdr. Henry):

1. Untuk iterasi kenapa nilainya 3, 2, 5, 2 pada 4x percobaan?
2. Kenapa nilai random ada yg banyak sekali ada sedikit sekali?
3. Untuk perbandingan data seperti itu apakah dapat menghasilkan data yg terbaik?
4. Kmeans ada 12, 9, 14, pada 3 cluster maksudnya apa?
5. Berdasar data mana yg lebih baik kmeans atau kmdedian?
6. Daerah mana yg menjadi prioritas/yg mendapat bantuan?

Jawaban:

1. Karena memang proses klasterisasi yang dibuat menghasilkan nilai iterasi sebanyak itu.
2. Nilai k ditentukan sebanyak 3, nilai centroid dibangkitkan secara random. Proses K-means memang dibangkitkan menggunakan nilai random, tetapi penulis mencoba membangkitkan menggunakan nilai kuartil.
3. Untuk nilai kuartil, banyak nya iterasi dan banyaknya klaster lebih konsisten dan sama, dibandingkan yang menggunakan nilai random.
4. Klaster ditentukan sebanyak 3 klaster, Klaster 1 = Sangat memerlukan bantuan, Klaster 2 = Cukup memerlukan bantuan pendidikan, Klaster 3 = Tidak terlalu memerlukan bantuan pendidikan.
5. Semakin sedikit jumlah dalam klaster semakin baik. K-median lebih baik Karena jumlah iterasi klaster pertama lebih sedikit.
6. Kabupaten Batang, Purworejo, Temanggung, Karanganyar.

Masukan Seminar :

Tidak ada penjelasan tentang data & cara evaluasi yang digunakan, sehingga sulit untuk mengikuti penjelasan hasil penelitian.

Yogyakarta, 19 November 2016

Moderator Kelas

Gloria Virginia, S.Kom., MAI, Ph.D.

Penyaji Makalah

Yuniaz Hezron Salulolo