

ANALISIS PENGELOMPOKAN JUMLAH PENUMPANG BUS TRANS JOGJA MENGGUNAKAN METODE *CLUSTERING K-MEANS* DAN *AGGLOMERATIVE HIERARCHICAL CLUSTERING (AHC)*

Lisna Zahrotun

Program Studi Teknik Informatika Universitas Ahmad Dahlan

Jl. Prof. Dr. Soepomo, Janturan, Yogyakarta 55164

Telp: (0274) 563515 ext. 3208

E-mail : lisna.zahrotun@tif.ac.id

Abstrak

Teknik kluster merupakan teknik pengelompokan yang sudah di kenal, dimana dalam teknik ini bertujuan untuk mengelompokkan data ke dalam kluster sehingga setiap kluster berisi data yang semirip mungkin. Dengan memanfaatkan data dari dinas perhubungan tentang jumlah penumpang Trans Jogja, maka dalam penelitian ini dilakukan pengelompokan jumlah penumpang berdasarkan jalur bus dan shelter dengan menggunakan metode clustering k-means, dan Agglomerative Hierarchical Clustering (AHC). Penelitian ini menghasilkan 3 cluster untuk metode K-Means dengan jumlah data untuk cluster 0 ada 15 data, cluster 1 ada 44 data dan cluster 2 ada 54 data. Dan metode AHC3 cluster dengan jumlah data untuk cluster 0 ada 2 data, cluster 1 ada 57 data dan cluster 2 ada 54 data. Sehingga dari hasil pengelompokan tersebut dapat terlihat metode K-Means memiliki hasil yang lebih baik dari pada metode AHC

Kata kunci : shelter, clustering k-means, agglomerative hierarchical clustering .

1. PENDAHULUAN

Pertumbuhan penduduk di suatu negara akan berbanding lurus dengan kebutuhan sarana transportasi. Trans Jogja adalah sebuah sistem transportasi bus cepat, murah, dan ber-AC di seputar Kota Yogyakarta yang dicanangkan Dinas Perhubungan, Komunikasi, dan Informatika (Dishubkominfo) DIY. Berbeda dengan bus kota kebanyakan, bus Trans Jogja membutuhkan halte (*shelter*) khusus sebagai tempat pertemuan antara calon penumpang dengan bus.

Seiring perkembangan waktu, jumlah penumpang bus Patas Trans Jogja mengalami kenaikan dari waktu ke waktu. Namun, kenaikan tersebut belum ditopang dengan peningkatan jumlah shelter, sehingga banyak calon penumpang yang tidak terangkut di jalur bus Trans Jogja. Sedangkan disisi lain pada shelter-shelter tertentu hanya terdapat sedikit calon penumpang pada jam-jam tertentu. Selain itu banyak warga yang mengeluhkan sejumlah tempat yang masuk jalur, tetapi kurang ditemui adanya shelter. Sehingga banyak diantaranya yang ingin menggunakan jasa bus Trans Jogja, tetapi tidak terpenuhi karena tidak adanya shelter. (<http://transjogja.com>)

Teknik kluster merupakan teknik pengelompokan yang sudah di kenal, dimana dalam teknik ini bertujuan untuk mengelompokkan data ke dalam kluster sehingga setiap kluster berisi data yang semirip mungkin. Menurut Santosa (2007) ada dua pendekatan dalam clustering, yaitu partisioning dan hirarki.

Pada algoritma *hierarchical clustering* menurut Tang, dkk (2006) terdapat beberapa keunggulan yaitu tidak perlu menentukan jumlah kluster yang diinginkan karena proses dapat langsung dihentikan pada saat jumlah kluster sesuai dengan yang diinginkan. Namun algoritma ini juga memiliki kelemahan bergantung pada pemilihan teknik *intercluster similarity* yang lebih dikenal dengan istilah *linkage*. Beberapa kelemahan dari linkage tersebut adalah sensitif terhadap adanya *outlier*, kesulitan menangani variasi bentuk dan ukuran, dan memisahkan cluster yang besar.

Sedangkan Berkhin menyebutkan beberapa kelemahan algoritma *K-means* adalah: (1) sangat bergantung pada pemilihan nilai awal centroid, (2) tidak jelas berapa banyak cluster *k* yang terbaik, (3) hanya bekerja pada atribut numerik.

Algoritma *K-means* ini juga memiliki kelebihan yaitu dinilai cukup efisien, yang ditunjukkan dengan kompleksitasnya $O(kn)$, dengan catatan n adalah banyaknya obyek data, k adalah jumlah cluster yang dibentuk, dan t banyaknya iterasi. Biasanya, nilai k dan t jauh lebih kecil daripada nilai n . Selain itu, dalam iterasinya, algoritma ini akan berhenti dalam kondisi optimum lokal (Tang, dkk 2005).

Dengan mengetahui kelebihan dan kelemahan *Agglomerative Hierarchical Clustering* (AHC) dan juga permasalahan pada Dinas Perhubungan tentang jumlah penumpasan bus Trans Jogja maka dilakukan penelitian lanjutan dari penelitian sebelumnya yang telah dilakukan oleh penulis yaitu mengelompokkan jumlah penumpang bus dengan metode *clustering K-Means* ke dalam penelitian ini dengan judul “Analisis Pengelompokan Jumlah Penumpang Bus Trans Jogja Menggunakan Metode *Clustering K-Means* dan *Agglomerative Hierarchical Clustering* (AHC)”.

2. TINJAUAN PUSTAKA

Irdiansyah (2007) dalam jurnalnya didapatkan informasi dari data penjualan yang besar yaitu sebagai gambaran untuk pengambilan keputusan bagi perusahaan. Pengelompokan dalam penelitian ini juga digunakan oleh perusahaan untuk membuat laporan mengenai karakteristik umum dari grup-grup konsumen yang berbeda. Clustering yang digunakan adalah *Agglomerative Hierarchical Clustering Algorithm* (AHC algorithm).

Pralita (2010) dalam penelitiannya ini penerapan JST digunakan untuk prediksi jumlah penumpang bus di Kabupaten Kebumen dengan menggunakan metode Backpropagation berdasarkan pada data bulan Januari 2004 sampai Juli 2009. Dari hasil pelatihan dan pengujian data yang telah dilakukan, dihasilkan nilai prediksi untuk Agustus 2009 adalah sebesar 21857 orang.

Zahrotun (2013) dalam penelitiannya melakukan pengelompokkan menggunakan metode clustering *k-means*. Dimana dalam penelitian ini masih perlu dikembangkan karena belum dapat menghasilkan data untuk setiap clusternya.

3. DASAR TEORI

Analisis kelompok (*cluster analisis*) adalah pekerjaan mengelompokkan data (objek) yang didasarkan hanya pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan diantaranya (Tan, 2006 dalam buku Prasetyo, 2012). Ada beberapa cara yang dapat digunakan untuk mengukur jarak data ke pusat kelompok, di antaranya Euclidean (Bezdek, 1981 dalam buku Prasetyo, 2012) menggunakan formula sebagai berikut:

$$D(x_2 - x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^p |x_{2j} - x_{1j}|^2} \quad (1)$$

Algoritma Clustering K-Means

K-Means merupakan algoritma clustering yang berulang-ulang. Algoritma K-Means dimulai dengan pemilihan secara acak K, K disini merupakan banyaknya cluster yang ingin dibentuk. Kemudian tetapkan nilai-nilai K secara random, untuk sementara nilai tersebut menjadi pusat dari cluster atau biasa disebut dengan centroid, mean atau “means”. Hitung jarak setiap data yang ada terhadap masing-masing centroid menggunakan rumus Euclidian hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Klasifikasikan setiap data berdasarkan kedekatannya dengan centroid. Lakukan langkah tersebut hingga nilai centroid tidak berubah (stabil) (Pralita, 2010).

Pada tahapan iterasi, nilai centroid cluster ditentukan dengan cara menghitung rata-rata dari data yang terletak pada cluster yang sama menggunakan rumus berikut:

$$v_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} x_{kj} \quad (2)$$

dimana:

i : indeks dari cluster

j : indeks dari variable

k : indeks dari data

v_{ij} : nilai centroid cluster ke-i untuk variabel ke-j

x_{kj} : nilai data ke-k yang ada di dalam cluster tersebut untuk variabel ke-j

N_i : Jumlah data yang menjadi anggota cluster ke-i

Algoritma Clustering AHC

Pengelompokan hierarki aglomeratif merupakan metode pengelompokan hierarki dengan pendekatan bawah atas (bottom up). Proses pengelompokan dimulai dari masing-masing data sebagai satu buah kelompok, kemudian secara rekursif mencari kelompok terdekat sebagai pasangan untuk bergabung sebagai satu kelompok yang besar. Proses tersebut diulang terus sehingga tampak bergerak ke atas membentuk jenjang (hierarki). Cara ini membutuhkan parameter kedekatan kelompok (cluster proximity). (Prasetyo, 2012)

Single Linked (Jarak Terdekat)

Pada metode single linked (jarak terdekat), kedekatan di antara dua kelompok ditentukan dari jarak terdekat (terkecil) diantara pasangan diantara dua data dari dua kelompok yang berbeda (satu dikelompok dari kelompok pertama dan satu lagi dari kelompok yang lain, atau disebut juga kemiripan termaksimal. Formulasi tautan tunggalnya adalah (Prasetyo, 2012):

$$D_{UV} = \min\{d_{UV}\}, d_{UV} \in D \quad (3)$$

4. METODOLOGI

1.1 Subjek Penelitian

Subjek penelitian yang akan dibahas adalah *data mining* untuk mengelompokkan jumlah penumpang Trans Jogja dalam satu hari

1.2 Metode Pengumpulan Data

1.2.1 Studi Pustaka

Merupakan pengumpulan data yang dilakukan dengan mencari, membaca dan mengumpulkan dokumen-dokumen sebagai referensi seperti buku, artikel dan literature-literatur yang berhubungan dengan data mining menggunakan metode clustering k-means, serta browsing di internet.

1.2.2 Observasi

Metode observasi ini digunakan untuk mempelajari dan mengetahui secara langsung objek yang diteliti. Objek yang diteliti itu adalah jumlah penumpang Trans Jogja pada shelter musper untuk tiga jalur bus.

5. HASIL DAN PEMBAHASAN

Penggunaan Algoritma K-Means

Dari hasil pengolahan menggunakan metode K-Means dapat diperoleh hasil bahwa jumlah cluster 0 adalah 15 data, cluster 1 adalah 44 data, dan cluster 2 adalah 54 data. Dengan nilai titik pusat terdapat pada Gambar 4.2

Attribute	Full Data (113)	Cluster#		
		0 (15)	1 (44)	2 (54)
jamkedatangan	3.4248	4	3.3636	3.3148
jalurbus	19.115	24.2667	29.8864	8.9074
jumlahpenumpang	12.3363	39.0667	10.1818	6.6667

Gambar 4.2 Titik Pusat Cluster K-Means

Dan untuk data yang termasuk dalam cluster 0 ditampilkan dalam Tabel 4.3, cluster 1 ditampilkan dalam Tabel 4.4 dan cluster 2 ditampilkan dalam Tabel 4.5. Dalam cluster 0 ini merupakan data dimana jumlah penumpang dikategorikan banyak, cluster 1 jumlah penumpang sedang dan cluster 2 jumlah penumpang sedikit. Sebaran dari data terhadap ketiga cluster menghasilkan jumlah data pada masing-masing cluster sebagai berikut: **cluster 0 = 15 data, cluster 1 = 44 data, dan cluster 2 = 54 data**. Dari hasil clustering terlihat bahwa densitas/kerapatan data pada masing-masing cluster terlihat seimbang, ini memperlihatkan bahwa data memiliki similitas yang baik.

Tabel 4.3 Data pada cluster 0 metode K-Means

Id bus	jam	Jalur bus	Halte/ shelter	jumlah penumpang
49	jp 15-17	3B	Terminal Concat	74
53	jp 15-17	3B	Sarjito Barat	21
55	jp 15-17	1B	Santika	46
63	jp 15-17	1A	Malioboro 3	53
65	jp 15-17	2A	Malioboro 3	33
69	jp 15-17	1A	Janti 3	21
71	jp 15-17	2A	Termconcat	37
73	jp 15-17	2B	Termconcat	39
77	jp 15-17	2B	Papmi Selatan	32
79	jp 15-17	3A	MT Haryono	22
81	jp 15-17	2A	SD Pujokusuman	55
83	jp 15-17	3A	SD Pujokusuman	39
75	jp 15-17	3A	Papmi Selatan	34
97	jp 15-17	2A	Kehutanan	44
99	jp 15-17	3A	Kehutanan	36

Tabel 4.4 Data pada cluster metode 1 K-Means

Id bus	Jam	Jalur bus	Halte/ shelter	jumlah penumpang
58	jp 13-14	3B	Papmi Utara	11
59	jp 15-17	3B	Papmi Utara	11
60	jp 13-14	3B	Tejokusuman	9
61	jp 15-17	3B	Tejokusuman	13
62	jp 13-14	1A	Malioboro 3	11
64	jp 13-14	2A	Malioboro 3	6
66	jp 13-14	3A	Malioboro 3	19
67	jp 15-17	3A	Malioboro 3	14
68	jp 13-14	1A	Janti 2	19
70	jp 13-14	2A	Termconcat	13
72	jp 13-14	2B	Termconcat	16
74	jp 13-14	3A	Papmi Selatan	22
75	jp 15-17	3A	Papmi Selatan	34
76	jp 13-14	2B	Papmi Selatan	8
78	jp 13-14	3A	MT Haryono	6
80	jp 13-14	2A	SD Pujokusuman	10
82	jp 13-14	3A	SD Pujokusuman	6
86	jp 13-14	3A	Tegal turi Utara	2
88	jp 13-14	3A	wirosaban utara	3
89	jp 15-17	3A	wirosaban utara	12
90	jp 13-14	3A	MBS Utara	5
91	jp 15-17	3A	MBS Utara	8
92	jp 13-14	1A	JEC 1	12
92	jp 13-14	1A	JEC 1	12
93	jp 15-17	1A	JEC 1	18
94	jp 13-14	3A	JEC 1	11
95	jp 15-17	3A	JEC 1	16
96	jp 13-14	2A	Kehutanan	26
97	jp 15-17	2A	Kehutanan	44
98	jp 13-14	3A	Kehutanan	27
99	jp 15-17	3A	Kehutanan	36
100	jp 13-14	3A	Janti Utara	7
101	jp 15-17	3A	Janti Utara	24
102	jp 13-14	3A	Disnaker	4
103	jp 15-17	3A	Disnaker	11
104	jp 13-14	3A	Instiper Selatan	11
105	jp 15-17	3A	Instiper Selatan	16
106	jp 13-14	3A	UPN	3
107	jp 15-17	3A	UPN	7
108	jp 13-14	2A	Manggung	2
109	jp 15-17	2A	Manggung	6
110	jp 13-14	3A	Manggung	3
111	jp 15-17	3A	Manggung	11
112	jp 13-14	3A	Sarjito Timur	3
113	jp 15-17	3A	Sarjito Timur	6

Tabel 4.5 Data pada cluster 2 metode K-Means

Id bus	Jam	Jalur bus	Halte/ shelter	jumlah penumpang
1	jp 6-8	1B	Tamrin Senopati	7
2	jp 13-14	1B	Tamrin Senopati	0
3	jp 15-17	1B	Tamrin Senopati	0
4	jp 17-19	1B	Tamrin Senopati	0
5	jp 6-9	2B	Tamrin Senopati	17
6	jp 13-14	2B	Musper	15
7	jp 15-17	2B	Musper	20
8	jp 13-14	3B	Musper	8
9	jp 15-17	3B	Musper	33
10	jp 13-14	3B	Lowanu barat	1
11	jp 15-17	3B	Lowanu barat	6
12	jp 13-14	3B	Wirosaban selatan	5
13	jp 15-17	3B	Wirosaban selatan	2
14	jp 13-14	3B	SMA 7	3
15	jp 15-17	3B	SMA 8	25
16	jp 13-14	3A	Ngabean	8
17	jp 15-17	3A	Ngabean	12
18	jp 13-14	3B	Ngabean	6
19	jp 15-17	3B	Ngabean	18
20	jp 13-14	2B	Ngabean	1
21	jp 15-17	2B	Ngabean	6
22	jp 13-14	1A	Jayakarta	6
23	jp 15-17	1A	Jayakarta	9
24	jp 13-14	1B	Jayakarta	5
25	jp 15-17	1B	Jayakarta	0
26	jp 13-14	3B	Jayakarta	0
27	jp 15-17	3B	Jayakarta	4
28	jp 13-14	1B	HarjoLukito	2
29	jp 15-17	1B	HarjoLukito	9
30	jp 13-14	3B	HarjoLukito	0
31	jp 15-17	3B	HarjoLukito	3
32	jp 13-14	3B	Banguntapan	9
33	jp 15-17	3B	Banguntapan	11
34	jp 13-14	2B	Banguntapan	6
35	jp 15-17	2B	Banguntapan	12
36	jp 13-14	1B	JEC Concat	0
37	jp 15-17	1B	JEC Concat	3
38	jp 13-14	3B	JEC Concat	1
39	jp 15-17	3B	JEC Concat	3
40	jp 13-14	3B	MBS Selatan	0
41	jp 15-17	3B	MBS Selatan	2
42	jp 13-14	3B	Instiper	5
43	jp 15-17	3B	Instiper	9
44	jp 13-14	3B	Stikers	3
45	jp 15-17	3B	Stikers	7
46	jp 13-14	3B	JIH	0
47	jp 15-17	3B	JIH	0
48	jp 13-14	3B	Terminal Concat	17
50	jp 13-14	2B	Kentungan	3
51	jp 15-17	2B	Kentungan	12
52	jp 13-14	3B	Sarjito Barat	8
54	jp 13-14	1B	Santika	6
56	jp 13-14	3B	Santika	2
57	jp 15-17	3B	Santika	10

Penggunaan Algoritma Clustering AHC

Dari hasil pengolahan menggunakan metode AHC dapat diperoleh hasil bahwa jumlah cluster 0 adalah 32 data, cluster 1 adalah 57 data, dan cluster 2 adalah 54 data. Dan untuk data yang termasuk dalam cluster 0 ditampilkan dalam Tabel 4.6, cluster 1 ditampilkan dalam Tabel 4.7 dan cluster 2 ditampilkan dalam Tabel 4.8.

Sebaran dari data terhadap ketiga cluster menghasilkan jumlah data pada masing-masing cluster sebagai berikut: **cluster 0 = 2 data**, **cluster 1 = 57 data**, dan **cluster 2 = 54 data**. Dari hasil clustering terlihat bahwa densitas/kerapatan data pada masing-masing cluster terlihat seimbang, ini memperlihatkan bahwa data memiliki similitas yang kurang baik.

Tabel 4.6 Data pada cluster 0 metode AHC

Id bus	Jam	Jalurbus	Halte/ shelter	Jumlah penumpang
1	jp 6-8	1B	Tamrin Senopati	7
5	jp 6-9	2B	Tamrin Senopati	17

Tabel 4.7 Data pada cluster 1 metode AHC

Id bus	Jam	Jalur bus	Halte/ shelter	Jumlah penumpang
2	jp 13-14	1B	Tamrin Senopati	0
3	jp 15-17	1B	Tamrin Senopati	0
6	jp 13-14	2B	Musper	15
7	jp 15-17	2B	Musper	20
8	jp 13-14	3B	Musper	8
10	jp 13-14	3B	Lowanu barat	1
12	jp 13-14	3B	Wirosaban sltan	5
14	jp 13-14	3B	SMA 7	3
16	jp 13-14	3A	Ngabean	8
18	jp 13-14	3B	Ngabean	6
20	jp 13-14	2B	Ngabean	1
22	jp 13-14	1A	Jayakarta	6
24	jp 13-14	1B	Jayakarta	5
26	jp 13-14	3B	Jayakarta	0
28	jp 13-14	1B	HarjoLukito	2
30	jp 13-14	3B	HarjoLukito	0
32	jp 13-14	3B	Banguntapan	9
34	jp 13-14	2B	Banguntapan	6
36	jp 13-14	1B	JEC Concat	0
38	jp 13-14	3B	JEC Concat	1
40	jp 13-14	3B	MBS Selatan	0
42	jp 13-14	3B	Instiper	5
44	jp 13-14	3B	Stikers	3
46	jp 13-14	3B	JIH	0
48	jp 13-14	3B	Terminal Concat	17
50	jp 13-14	2B	Kentungan	3
52	jp 13-14	3B	Sarjito Barat	8
54	jp 13-14	1B	Santika	6
56	jp 13-14	3B	Santika	2

Id bus	Jam	Jalur bus	Halte/ shelter	Jumlah penumpang
58	jp 13-14	3B	Papmi Utara	11
60	jp 13-14	3B	Tejokusuman	9
62	jp 13-14	1A	Malioboro 3	11
64	jp 13-14	2A	Malioboro 3	6
66	jp 13-14	3A	Malioboro 3	19
68	jp 13-14	1A	Janti 2	19
70	jp 13-14	2A	Termconcat	13
72	jp 13-14	2B	Termconcat	16
74	jp 13-14	3A	Papmi Selatan	22
76	jp 13-14	2B	Papmi Selatan	8
78	jp 13-14	3A	MT Haryono	6
80	jp 13-14	2A	SD Pujokusuman	10
82	jp 13-14	3A	SD Pujokusuman	6
84	jp 13-14	3A	Lowanu Timur	0
86	jp 13-14	3A	Tegal turi Utara	2
88	jp 13-14	3A	wirosaban utara	3
90	jp 13-14	3A	MBS Utara	5
92	jp 13-14	1A	JEC 1	12
94	jp 13-14	3A	JEC 1	11
96	jp 13-14	2A	Kehutanan	26
98	jp 13-14	3A	Kehutanan	27
100	jp 13-14	3A	Janti Utara	7
102	jp 13-14	3A	Disnaker	4
104	jp 13-14	3A	Instiper Selatan	11
106	jp 13-14	3A	UPN	3
108	jp 13-14	2A	Manggung	2
110	jp 13-14	3A	Manggung	3
112	jp 13-14	3A	Sarjito Timur	3

Tabel 4.8 Data pada cluster 1 metode AHC

Id bus	jam	Jalur bus	Halte/ shelter	jumlah penumpang
4	jp 17-19	1B	Tamrin Senopati	0
5	jp 6-9	2B	Tamrin Senopati	17
9	jp 15-17	3B	Musper	33
11	jp 15-17	3B	Lowanu barat	6
13	jp 15-17	3B	Wirosaban selatan	2
15	jp 15-17	3B	SMA 8	25
17	jp 15-17	3A	Ngabean	12
19	jp 15-17	3B	Ngabean	18
21	jp 15-17	2B	Ngabean	6
23	jp 15-17	1A	Jayakarta	9
25	jp 15-17	1B	Jayakarta	0
27	jp 15-17	3B	Jayakarta	4
29	jp 15-17	1B	HarjoLukito	9
31	jp 15-17	3B	HarjoLukito	3
33	jp 15-17	3B	Banguntapan	11
35	jp 15-17	2B	Banguntapan	12

Id bus	jam	Jalur bus	Halte/ shelter	jumlah penumpang
59	jp 15-17	3B	Papmi Utara	11
61	jp 15-17	3B	Tejokusuman	13
63	jp 15-17	1A	Malioboro 3	53
65	jp 15-17	2A	Malioboro 3	33
67	jp 15-17	3A	Malioboro 3	14
69	jp 15-17	1A	Janti 3	21
71	jp 15-17	2A	Termconcat	37
73	jp 15-17	2B	Termconcat	39
75	jp 15-17	3A	Papmi Selatan	34
77	jp 15-17	2B	Papmi Selatan	32
79	jp 15-17	3A	MT Haryono	22
81	jp 15-17	2A	SD Pujokusuman	55
83	jp 15-17	3A	SD Pujokusuman	39
85	jp 15-17	3A	Lowanu Timur	0
87	jp 15-17	3A	Tegal turi Utara	0
89	jp 15-17	3A	wirosaban utara	12

37	jp 15-17	1B	JEC Concat	3
39	jp 15-17	3B	JEC Concat	3
41	jp 15-17	3B	MBS Selatan	2
43	jp 15-17	3B	Instiper	9
45	jp 15-17	3B	Stikers	7
47	jp 15-17	3B	JIH	0
49	jp 15-17	3B	Terminal Concat	74
51	jp 15-17	2B	Kentungan	12
53	jp 15-17	3B	Sarjito Barat	21
55	jp 15-17	1B	Santika	46
57	jp 15-17	3B	Santika	10
91	jp 15-17	3A	MBS Utara	8
93	jp 15-17	1A	JEC 1	18
95	jp 15-17	3A	JEC 1	16
97	jp 15-17	2A	Kehutanan	44
99	jp 15-17	3A	Kehutanan	36
101	jp 15-17	3A	Janti Utara	24
103	jp 15-17	3A	Disnaker	11
105	jp 15-17	3A	Instiper Selatan	16
107	jp 15-17	3A	UPN	7
109	jp 15-17	2A	Manggung	6
111	jp 15-17	3A	Manggung	11
113	jp 15-17	3A	Sarjito Timur	6

Dari hasil pengelompokkan menggunakan metode *clustering K-Means* dan *clustering AHC* maka dalam penelitian ini terlihat hasil dari pengelompokkan *clustering K-Means* memiliki kemiripan yang lebih seimbang untuk setiap kelompoknya dibandingkan dengan hasil dari pengelompokkan *clustering AHC*. Sehingga dapat disimpulkan meskipun pada proses *clustering K-Means* harus menentukan centroid atau titik terlebih dahulu namun dalam kasus ini dapat menghasilkan data pengelompokkan yang lebih baik. Sedangkan untuk hasil pengelompokkan menggunakan metode AHC meskipun lebih efisien karena tidak menggunakan titik pusat pada awal pengelompokkan dan hanya menghitung tingkat kemiripan memiliki hasil yang kurang baik.

6. KESIMPULAN DAN SARAN

Setelah melakukan data mining dengan algoritma *clustering K-Means* dan *clustering AHC* pada jumlah penumpang bus Trans Jogja, maka dapat ditarik kesimpulan sebagai berikut :

1. Hasil dari data mining yang dapat digunakan untuk menampilkan informasi pengelompokkan jumlah penumpang berdasarkan jam, halte/shelter dan jalur bus.
2. Hasil dari proses *clustering K-Means* ini didapatkan 3 cluster dengan jumlah data untuk cluster 0 ada 15 data, cluster 1 ada 44 data dan cluster 2 ada 54 data.
3. Hasil dari proses *clustering K-Means* ini didapatkan 3 cluster dengan jumlah data untuk cluster 0 ada 2 data, cluster 1 ada 57 data dan cluster 2 ada 54 data.
4. Dalaam penelitian ini dihasilkan metode K-Means memiliki hasil yang lebih baik dari pada metode AHC.

Aplikasi *data mining* yang dihasilkan pada penelitian ini masih perlu dikembangkan lebih lanjut dalam hal data yang ditampilkan tidak hanya berupa data namun juga dapat digambarkan dalam bentuk grafik dan dendrogram yang lebih jelas.

DAFTAR PUSTAKA

- Berkhin, Pavel. *Survey on clustering data mining techniques*, http://www.ee.ucr.edu/~barth/EE242/clustering_survey.pdf di akses pada Desember 2013
- Irdiansyah 2007, “*Penerapan Data Mining Pada Penjualan Minuman di PT.Pepsi Cola Indobeverages Menggunakan Metode Clustering*”
- Pralita, F. 2010 ,“*Penerapan Jaringan Syaraf tiruan Untuk Memprediksi Jumlah Penumpang Bus Dengan Metode BackPropagation*”, Skripsi, Fakultas MIPA, Universitas Islam Indonesia.
- Prasetyo, Eko. 2012, “*Data Mining Konsep dan Aplikasi Menggunakan Matlab*”, Andi offset, Yogyakarta.
- Santosa, Budi, 2007, “*Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*”, Graha Ilmu, Yogyakarta.
- Tan, Pang-Ning,; Steinbach,Michael; Kumar ,Vipin, 2006 *Data Mining Cluster Analysis : Basic Concepts and Algorithms*. <http://www-users.cs.umn.edu/~kumar/dmbook/index.php> di akses pada Desember 2013
- Tang, ZhaoHui; MacLennan, Jamie. 2005. *Data Mining with SQL Server 2005*. Indiana Polis : Wiley Publishing
- Zahrotun, L, Perancangan Data Mining Pengelompokan Jumlah Penumpang Trans Jogja menggunakan metode clustering K-Means, Seminar Teknik Informatika, Universitas Ahmad Dahlan. 2013, Yogyakarta. 134-142
- <http://transjogja.com/gunakan-trans-jogja/>, tanggal akses 12 Desember 2013