

# MODEL ANALISIS *PRE-DEFINED SINGLE LINKAGE CLUSTERING* PADA SISTEM REKOMENDASI OBYEK WISATA DI KOTA SEMARANG

Dwi Budi Santoso<sup>1</sup>, Dewi Handayani U.N<sup>2</sup>, Saefurrohman<sup>3</sup>

<sup>1,2,3</sup>Faculty Information Technology, Stikubank University

email: <sup>1</sup>db@edu.unisbank.ac.id, <sup>2</sup>dewi\_h@edu.unisbank.ac.id, <sup>3</sup>saefurr@edu.unisbank.ac.id

## Abstrak

Perjalanan wisata selalu berkaitan dengan bagaimana seorang wisatawan mengunjungi obyek wisata berdasar kebutuhan informasi tentang rute yang harus dilalui ke obyek wisata yang diinginkan dan transportasi apa saja yang bisa digunakan berdasar waktu yang dimiliki. Petunjuk arah juga sangat diperlukan untuk kemudahan akses perjalanan.

Model analisis *pre-defined single-linkage* dilakukan untuk menentukan obyek mana saja yang memiliki lokasi berdekatan bisa membentuk satu klaster dan kunjungan dilakukan berdasar jarak paling dekat dengan klaster menuju klaster berikutnya sesuai dengan kelompok jenis obyek wisata. Perhitungan jarak antar obyek dengan pengguna menggunakan metode *Euclidean Distance* untuk mencari jarak berdasar ketetanggaan terdekat.

Hasil analisis penentuan obyek wisata di kota Semarang dengan metode *pre-defined single-linkage* divisualisasikan menggunakan dendrogram yang bisa menggambarkan kelompok obyek dalam satu klaster berdasar ketetanggaan terdekat. Dari dendrogram bisa dilihat jumlah klaster yang terbentuk dari 14 obyek wisata.

**Kata Kunci:** *analisa klastering, pre-defined single linkage, metode nearest neighbour, rute obyek wisata, dendrogram*

## 1. PENDAHULUAN

Perjalanan wisata selalu berkaitan dengan bagaimana seorang wisatawan mengunjungi obyek wisata dengan berdasar kebutuhan informasi tentang rute yang harus dilalui untuk melakukan kunjungan ke obyek wisata yang diinginkan. Petunjuk arah juga sangat diperlukan untuk kemudahan akses perjalanan. Kebutuhan informasi tentang seberapa banyak obyek wisata yang bisa dikunjungi dan bisa dijangkau sesuai dengan waktu yang dimiliki.

Pada saat seorang wisatawan ingin mengunjungi sebuah obyek wisata, hal pertama yang dilakukan adalah membuka internet untuk mendapatkan informasi yang dibutuhkan. *Website* selalu memberikan informasi yang sama pada wisatawan yang berbeda, tanpa memperhatikan kebutuhan personal masing-masing wisatawan. Data yang diberikan kepada wisatawan bisa sesuai dengan permintaan, kebutuhan dan minat, diperlukan satu sistem rekomendasi yang bisa memberikan beberapa alternatif rekomendasi yang membuat situs pariwisata menjadi lebih cerdas [1].

Kebutuhan untuk menentukan rute yang bisa dilalui berdasar posisi pengguna berada dengan jarak terdekat dari klaster obyek wisata yang terbentuk dengan mengelompokkan berdasarkan kesamaan jenis obyek wisata.

*Clustering* merupakan teknologi yang digunakan untuk menemukan pola-pola tersembunyi. Obyek-obyek data dikelompokkan dalam kelompok kemiripan yang disebut *cluster*. Dalam satu *cluster* obyek-obyek data didalamnya memiliki kemiripan satu sama lain dan antar *cluster* memiliki perbedaan [3]. Untuk mengelompokkan obyek-obyek wisata yang berdekatan membentuk satu klaster menggunakan analisa klaster dengan metode *Hierarchical Agglomerative Cluster* (HAC).

Analisa kluster yang merupakan salah satu teknik data mining bertujuan mengidentifikasi sekelompok obyek yang mempunyai kemiripan karakteristik tertentu dapat dipisahkan dengan kelompok obyek lainnya, sehingga obyek berada pada kelompok yang berbeda.

Metode *nearest neighbour* merupakan salah satu metode yang digunakan pada teknologi clustering. Pengelompokan dimulai dengan mengelompokkan dua obyek yang memiliki jarak yang terdekat menjadi satu *cluster* [3]. Proses pengelompokan dilanjutkan ke obyek-obyek dengan jarak obyek yang semakin jauh, sampai akhirnya semua obyek terhubung menjadi sebuah rute. Analisis jarak terdekat yang bisa dikunjungi berdasarkan posisi wisatawan berada dibanding dengan posisi terdekat pada kluster. Wisatawan akan mendapatkan rekomendasi informasi urutan obyek yang dapat dikunjungi sesuai dengan kluster peruntukan obyek wisata. Pengelompokan juga dapat diatur seberapa jauh perjalanan yang akan ditempuh dan memakan waktu berapa lama.

Pemodelan kembali rute optimal yang dihasilkan dari urutan waktu kunjungan menggunakan *single linked list* dengan titik awal dari obyek yang dituju dengan metode penentuan titik awal dimulai dari user berada.

Obyek wisata yang dijadikan acuan dalam perhitungan untuk analisis jarak paling dekat obyek wisata yang bisa dijangkau oleh pengguna terdiri dari beberapa obyek wisata sesuai dengan pembagian peruntukan yang bisa mewakili kluster yang diidentifikasi. Data yang digunakan sebagai sampel pemodelan diambil dari beberapa obyek wisata yang biasa dikunjungi oleh wisatawan ke kota Semarang sejumlah 14 obyek wisata dengan posisi (berdasar koordinat latitude dan longitude) yang dirasa bisa mewakili kunjungan yaitu:

Tabel 1. Obyek Wisata Di Semarang Dengan Posisi Latitude dan Longitude

No	Obyek Wisata	Koordinat	
		Latitude	Longitude
1	Lawang Sewu	-6.9844579	110.4077796
2	Stasiun Tawang	-6.973959	110.4094061
3	Stasiun Besar Poncol	-6.9787517	110.408679
4	Klenteng Sam Po kong	-6.9888498	110.3996613
5	Masjid Agung Jawa Tengah (Baiturrahman)	-6.986011	110.4295715
6	Masjud Agung Jawa Tengah (MAJT)	-7.0086145	110.4121065
7	Masjid Besar Kauman	-6.9778208	110.4120934
8	Museum Ronggowarsito	-6.9859542	110.3838605
9	Museum Mandala Bakti	-6.9847668	110.4087782
10	Candi Tugu	-6.9787612	110.3292208
11	Gereja Blenduk	-6.9743371,	110.4093101
12	Kawasan Kota Lama	-6.9684447	110.4259028
13	Pagoda Buddhagaya	-7.0862474,	110.409091
14	Gedung Marba	-6.9696509	110.4225325

Untuk mengetahui jarak antar obyek dengan posisi terdekat dengan pengguna menggunakan metode Euclidian Distance dan berdasar layanan google map.

Melakukan predefine single linked list metode *nearest neighbour* berdasar posisi wisatawan berada dengan kluster obyek wisata terdekat di kota Semarang.

Penelitian tentang analisa klustering dengan metode Single Linked list sudah banyak dilakukan sebelumnya. Sebagai salah satu metode IR, *clustering* melakukan pengelompokan data berdasarkan *cluster*/kelas dan merupakan teknik untuk mengorganisasikan data yang tidak terstruktur tersebut menjadi suatu struktur data yang mempunyai nilai informasi tertentu. Tujuan dari algoritma ini adalah untuk mengelompokkan dokumen yang memiliki kesamaan

(*similarity*) ke dalam satu *cluster* tertentu. Bisa dikatakan bahwa dokumen dalam satu *cluster* akan memiliki relevansi satu sama lain, tetapi tidak memiliki relevansi dengan dokumen di dalam *cluster* lain. Dalam proses yang baik, data yang dihasilkan dalam satu *cluster* akan memiliki tingkat kesamaan yang tinggi, dan akan memiliki nilai kesamaan yang rendah dengan *cluster* yang lainnya.

*Clustering* melakukan pengelompokan data tanpa berdasar pada kelas data tertentu yang sudah ditetapkan dari awal. Proses ini sangat berbeda dengan proses pada *classification* yang pada awal proses harus memberikan kelas-kelas data. Sehingga *clustering* sering disebut dengan pengelompokan data yang tidak terstruktur.

Dalam algoritma *clustering*, biasanya diperlukan fungsi jarak yang digunakan untuk mengukur jarak (kemiripan suatu data dengan data yang lain). Beberapa fungsi jarak yang dapat digunakan dalam proses *clustering* adalah:

- a. *Euclidean Distance*, merupakan perhitungan jarak dengan menggunakan 2 keadaan.
- b. *Manhattan Distance*, disebut juga dengan *taxicab-norm*, yang menghitung jarak dengan menggunakan 1 keadaan.
- c. *Mahalanobis Distance*, merupakan fungsi jarak yang digunakan untuk menghitung data dalam skala dan korelasi yang berbeda.

Analisis cluster termasuk dalam analisis multivariat yang mewakili seluruh hubungan interdependensi, tidak ada perbedaan variabel bebas dan variabel tak bebas (*independent and dependent variables*) dalam analisis ini. “Analisis cluster adalah teknik yang digunakan untuk mengidentifikasi objek atau individu yang serupa dengan memperhatikan beberapa kriteria” [4]. Analisis cluster merupakan bagian penting dari teknologi data mining untuk menemukan pola distribusi data yang menarik dalam data potensial dalam kondisi tanpa pengalaman sebelumnya [9]. Analisis cluster merupakan teknik multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya.

Analisis cluster mengklasifikasi objek sehingga setiap objek yang paling dekat kesamaanya dengan objek lain berada dalam cluster yang sama. Kelompok-kelompok yang terbentuk memiliki homogenitas internal yang tinggi dan heterogenitas eksternal yang tinggi. Berbeda dengan teknik multivariat lainnya, analisis ini tidak mengestimasi set variabel secara empiris sebaliknya menggunakan set variabel yang ditentukan oleh peneliti itu sendiri. Fokus dari analisis cluster adalah membandingkan objek berdasarkan set variabel, hal inilah yang menyebabkan para ahli mendefinisikan set variabel sebagai tahap kritis dalam analisis cluster. Set variabel cluster adalah suatu set variabel yang mempresentasikan karakteristik yang dipakai objek-objek.

Analisis cluster dilakukan untuk tujuan:

- a. Menggali data/eksplorasi data,
- b. Mereduksi data menjadi kelompok data baru dengan jumlah lebih kecil atau dinyatakan dengan pengkelasan (klasifikasi) data,
- c. Menggeneralisasi suatu populasi untuk memperoleh suatu hipotesis dan,
- d. Menduga karakteristik data-data.

Algoritma klaster harus mampu diimplementasikan pada berbagai macam bentuk data seperti data nominal, ordinal maupun gabungannya dimana tujuan utama analisis cluster adalah mengelompokkan objek-objek berdasarkan kesamaan karakteristik diantara objek-objek tersebut [5].

*Clustering* adalah studi formal metode dan algoritma untuk partisi atau pengelompokan. Analisis *clustering* tidak menggunakan pelabelan kategori sebelumnya. *Clustering* bersifat *unsupervised learning* atau tidak mempunyai tahap pelatihan data, berbeda dengan klasifikasi [6].

Pada saat seorang wisatawan ingin mengunjungi sebuah obyek wisata, hal pertama yang dilakukan adalah membuka internet untuk mendapatkan informasi yang dibutuhkan. Untuk menyampaikan informasi, *website* selalu memberikan informasi yang sama pada wisatawan yang berbeda, tanpa memperhatikan kebutuhan personal masing-masing wisatawan. Supaya sektor pariwisata lebih efektif dan efisien, data yang diberikan kepada wisatawan seharusnya sesuai permintaan, kebutuhan dan minat. Penerapan sistem rekomendasi akan membuat situs pariwisata lebih cerdas [1].

Metode *nearest neighbour* menggunakan jarak tetangga terdekat untuk mengukur ketidakmiripan antara dua kelompok, termasuk dalam kategori *agglomerative (bottom up) clustering*, yang akan menghasilkan *dendrogram(tree)* dari tingkat yang paling bawah, menggabungkan dua *cluster* yang paling mirip pada tiap tingkat(*level*) untuk menuju ke tingkat yang lebih tinggi. Proses berulang sampai semua kelompok data digabungkan menjadi satu *cluster* [2].

### Jarak Euclidean

Jarak *Euclidean* adalah besarnya jarak suatu garis lurus yang menghubungkan antar objek. Misalkan ada dua objek yaitu A dengan koordinat  $(x_1, y_1)$  dan B dengan koordinat  $(x_2, y_2)$  maka jarak antar kedua objek tersebut dapat diukur dengan rumus kurat jarak atau ketidaksamaan antar objek ke- $i$  dengan objek ke- $j$ , disimbolkan dengan  $d_{ij}$  dan  $k=1, \dots, p$ . Nilai  $d_{ij}$  diperoleh melalui perhitungan jarak kuadrat *Euclidean* sebagai berikut:

$$d_{ij} = \sqrt{\sum_{k=1}^F (x_{ik} - x_{jk})^2}$$

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Keterangan:

$d_{ij}$  =Jarak Kuadrat *Euclidean* antar objek ke- $i$  dengan objek ke- $j$

P =Jumlah variabel cluster

$X_{ik}$  =Nilai atau data dari objek ke- $i$  pada variabel ke- $k$

$X_{jk}$  =Nilai atau data dari objek ke- $j$  pada variabel ke- $k$  [6].

## 2. METODE PENELITIAN

### a. Hierarchical Agglomerative Clustering (HAC)

Hierarchical Clustering adalah metode analisis kelompok yang berusaha untuk membangun sebuah hirarki kelompok data. Strategi pengelompokannya umumnya ada 2 jenis yaitu *Agglomerative (Bottom-Up)* dan *Devisive (Top-Down)*. (Pada bagian ini akan dibatasi hanya menggunakan konsep *Agglomerative*).

HAC mengasumsikan satu fungsi untuk menentukan kesamaan/kemiripan dua kejadian. Salah satunya juga bisa mengasumsikan satu fungsi kesamaan dan sebaliknya dalam beberapa operasi ( seperti jarak minimal  $\rightarrow$  maksimal kemiripan) dalam algoritma untuk membangun kesamaan. Contoh Algoritma HAC adalah sebagai berikut:

### b. Fungsi Jarak Euclidean Distance

*Euclidean distance* adalah sebuah metode yang digunakan untuk mengukur jarak (*distance*). *Euclidean distance* sebenarnya merupakan generalisasi dari *teorema pythagoras*.

Berikut ini adalah contoh perhitungan dengan menggunakan *Euclidean Distance*. Jika terdapat dua buah titik pada sebuah bidang dua dimensi ( $R^2$ ).  $u = (x_1, y_1)$  dan  $v = (x_2, y_2)$  maka untuk mengukur jarak dari kedua belah titik tersebut dapat digunakan persamaan pythagoras.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

$x_1, x_2$  = koordinat sumbu x dari sebuah titik

$y_1, y_2$  = koordinat sumbu y dari sebuah titik

Jarak tersebut menyebabkan sebuah metrik pada  $R^2$ , yang disebut sebagai **Euclidean** metric pada  $R^2$ . Bila terdapat dua buah vektor dengan n dimensi,  $a = (a_1, a_2, \dots, a_n)$  dan  $b = (b_1, b_2, \dots, b_n)$  maka formula pythagoras dapat digeneralisasi ke dalam n dimensi ( $R^2$ ) menjadi:

$$d = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2}$$

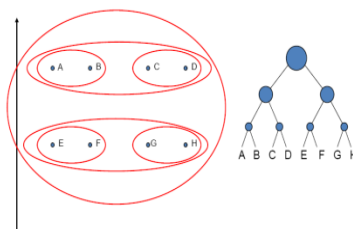
Perhatikan kemiripan dari dua buah formula diatas. Formula *Euclidean Distance* juga merupakan metrik pada  $R^n$ , dikenal sebagai *Euclidean Space*:

Berikut ini adalah bentuk umum dari *Euclidean Distance*:

$$d = (\sum_{n=1}^N (x_n - y_n)^p)^{1/p}$$

N = jumlah dimensi vektor

$x, y$  = vektor



Gambar 1. Contoh algoritma HAC

Algoritma *Agglomerative Hierarchical Clustering* :

- a. Matrik Jarak antar data.
- b. langkah c dan d hingga hanya satu kelompok yang tersisa.
- c. Gabungkan dua kelompok terdekat berdasarkan parameter kedekatan yang ditentukan.
- d. Matrik Jarak antar data untuk merepresentasikan kedekatan diantara kelompok baru dan kelompok yang masih tersisa.
- e. Selesai.

### 3. HASIL DAN PEMBAHASAN

#### a. Perhitungan Fungsi Jarak

Untuk menentukan jarak paling dekat dengan ketetanggan terdekat antar obyek untuk menghasilkan klaster dihitung berdasarkan posisi kordinat obyek wisata berdasar (latitude, longitude) yang diberikan di tabel 2. Ada 14 obyek wisata yang dijasikan sampel data untuk dilakukan analisis menggunakan metode single linkage HAC.

Tabel 2. Perhitungan Jarak User ke Obyek

User	-6.988327	110.4148582	Jarak user ke obyek	
			Euclidean Distance	Google Map
Lawang Sewu	-6.9840506	110.4077796	0.9	1.1
Stasiun Tawang	-6.9644338	110.4257718	2.9	3.3
Stasiun Besar Poncol	-6.9729977	110.4125928	1.7	2.5
Klenteng Sam Po kong	-6.9960734	110.3967853	2.2	3.3
Masjid Agung Jawa Tengah (MAJT)	-6.9834409	110.4429559	3.2	4.4
Masjud (Baiturrahman)	-6.9891405	110.4202544	0.6	
Masjid Besar Kauman	-6.9724511	110.420478	1.9	2.2
Museum Ronggowarsito	-6.9859489	110.3816718	3.7	4.1

Museum Mandala Bakti	-6.9847668	110.4087782	0.8	1.1
Candi Tugu	-6.9787612	110.329221	9.6	10.5
Gereja Blenduk	-6.9743371	110.4093101	1.7	3.3
Kawasan Kota Lama	-6.9684447	110.4259028	2.5	3.2
Pagoda Buddhagaya	-7.0862474	110.409091	10.9	17.5
Gedung Marba	-6.9696509	110.4225325	2.2	3.2

Sebagai perbandingan untuk perhitungan jarak berdasar posisi user berada (sebagai acuan dipakai posisi pengguna berada di UNISBANK dengan posisi latitude, longitude = -6.9883268, 110.4148582).

Tabel 3. Koordinat Posisi Obyek Wisata

No	Obyek Wisata	Koordinat	
		Latitude	Longitude
1	Lawang Sewu	-6.9844579	110.4077796
2	Stasiun Tawang	-6.973959	110.4094061
3	Stasiun Besar Poncol	-6.9787517	110.408679
4	Klenteng Sam Po kong	-6.9888498	110.3996613
5	Masjid Agung Jawa Tengah (Baiturrahman)	-6.986011	110.4295715
6	Masjud Agung Jawa Tengah (MAJT)	-7.0086145	110.4121065
7	Masjid Besar Kauman	-6.9778208	110.4120934
8	Museum Ronggowarsito	-6.9859542	110.3838605
9	Museum Mandala Bakti	-6.9847668	110.4087782
10	Candi Tugu	-6.9787612	110.3292208
11	Gereja Blenduk	-6.9743371,	110.4093101
12	Kawasan Kota Lama	-6.9684447	110.4259028
13	Pagoda Buddhagaya	-7.0862474,	110.409091
14	Gedung Marba	-6.9696509	110.4225325

Perhitungan jarak dimulai dari posis user berada ( $d_{user} = -6.9883268, 110.4148582$ )

**b. Perbandingan Jarak Berdasar *Euclidean Distance* Dengan Google Map**

Perbandingan jarak dilakukan untuk melihat realitas jarak tempuh dengan perhitungan fungsi jarak menggunakan *Euclidean Distance* dengan jarak yang dihasilkan dari Google Map yang diberikan dibawah ini

Perhitungan fungsi jarak *Euclidian Distance* dengan berdasar perhitungan secara diagonal dengan fungsi *Phytagoras* sehingga jarak yang diperoleh secara langsung ditarik garis lurus pada dua titik lokasi (diwakili dengan koordinat) dan dihitung. Sedangkan untuk jarak tempuh berdasar google map merupakan perhitungan yang dilakukan berdasarkan panjang jalan yang bisa dilalui.

1) Klasterisasi Obyek Wisata

Klasterisasi obyek wisata yang akan dikunjungi berdasar jarak terdekat pertama yang bisa dijangkau oleh pemakai ditentukan sebagai titik awal perhitungan untuk menentukan dalam klaster mana saja yang dihasilkan sesuai dengan batasan jarak tempuh yang diharapkan oleh wisatawan.

2) Perhitungan Jarak Antar Obyek

Proses clustering dengan metode *single linkage* merupakan penggabungan dua objek atau lebih yang mempunyai jarak terdekat. Proses pengklasteran dihitung dengan matrik jarak adalah sebagai berikut:

(a) Matrik 14 x 14 dibuat yang mewakili 14 obyek wisata. Langkah pertama mencari jarak antara dua obyek untuk semua kemungkinan obyek yang ada. Tahap berikutnya dari

setengah matrik (diagonal) dicari jarak yang paling pendek. Jarak dari Lawang Sewu dan Museum Mandala Bakti dengan jarak terdekat 0.1 km, kedua objek tersebut menjadi satu cluster dan membentuk cluster baru. Baris dan kolom yang berhubungan Lawang Sewu dan Museum Mandala Bakti dihilangkan kemudian dijadikan satu menjadi Lawang Sewu/Museum Mandala Bakti, sehingga tersisa 13 cluster yang membentuk matriks yang baru. Kemudian dilakukan perbaikan matriks jarak menggunakan metode *single linkage* dengan persamaan:

$$h(B_r, B_s) = \min\{d(x_i, x_j); x_i \text{ anggota } B_r, x_j \text{ anggota } B_s\}$$

Hasil perhitungan semua jarak dari ke 14 titik lokasi adalah sebagai berikut:

Tabel 4. Proses klusterisasi Lawang Sewu-Museum Mandala Bakti

No	Wisata Religi	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Masjid Baiturrahman	0.0	1.5	1.9	2.6	1.4	2.0	4.3	2.4	10.2	2.8	2.2	2.0	10.9	2.7
2	Lawang Sewu	1.5	0.0	1.8	0.1	0.1	1.5	0.3	2.5	0.8	3.1	2.3	1.7	16.1	1.9
3	Masjid Kauman Semarang	1.9	1.8	0.0	1.7	1.9	0.3	2.1	0.6	1.0	1.3	0.4	0.1	17.9	3.7
4	Masjid Agung MAJT	2.6	0.1	1.7	0.0	0.2	1.4	0.4	2.4	0.7	3.0	2.2	1.6	16.2	2.0
5	Museum Mandala Bakti	1.4	0.1	1.9	0.2	0.0	1.6	0.2	2.6	0.9	3.2	2.4	1.9	16.0	1.8
6	Gereja Bledug( Kristen)	2.0	1.5	0.3	1.4	1.6	0.0	1.8	0.9	0.7	1.6	0.7	0.2	17.6	3.4
7	Museum Ronggowarsito	4.3	0.3	2.1	0.4	0.2	1.8	0.0	2.8	1.1	3.4	2.6	2.0	15.8	1.6
8	Kawasan Kota Lama	2.4	2.5	0.6	2.4	2.6	0.9	2.8	0.0	1.6	0.6	0.2	0.7	18.5	4.3
9	Candi tugu	10.2	0.8	1.0	0.7	0.9	0.7	1.1	1.6	0.0	2.3	1.4	0.9	16.9	2.7
10	Stasiun Tawang Stasiun Poncol	2.8	3.1	1.3	3.0	3.2	1.6	3.4	0.6	2.3	0.0	0.8	1.3	19.2	5.0
11	Gedung Marba	2.2	2.3	0.4	2.2	2.4	0.7	2.6	0.2	1.4	0.8	0.0	0.5	18.4	4.2
12	Stasiun Poncol	2.0	1.7	0.1	1.6	1.9	0.2	2.0	0.7	0.9	1.3	0.5	0.0	17.8	3.6
13	Pagoda Buddha Gaya	10.9	16.1	17.9	16.2	16.0	17.6	15.8	18.5	16.9	19.2	18.4	17.8	0.0	14.2
14	Klenteng Sam Po Kong	2.7	1.9	3.7	2.0	1.8	3.4	1.6	4.3	2.7	5.0	4.2	3.6	10.0	0.0

Di sini yang mengalami perubahan adalah jarak yang melibatkan cluster baru (cluster yang anggotanya Lawang Sewu (lokasi ke-2) dan Museum Mandala Bakti (lokasi ke-5)) sehingga diperoleh sebagai berikut:

$$\begin{aligned}
 d(2/5)1 &= \min \{d_{21}, d_{51}\} \\
 &= \min \{1.5, 1.4\} \\
 &= 1.4 \\
 d(2/5)3 &= \min \{d_{23}, d_{53}\} \\
 &= \min \{1.8, 1.9\} \\
 &= 1.8 \\
 d(2/5)4 &= \min \{d_{24}, d_{54}\} \\
 &= \min \{0.1, 0.2\} \\
 &= 0.1 \\
 d(2/5)6 &= \min \{d_{26}, d_{56}\} \\
 &= \min \{1.5, 1.6\} \\
 &= 1.5 \\
 d(2/5)7 &= \min \{d_{27}, d_{57}\} \\
 &= \min \{0.3, 0.2\} \\
 &= 0.2 \\
 d(2/5)8 &= \min \{d_{28}, d_{58}\} \\
 &= \min \{2.5, 2.6\}
 \end{aligned}$$

$$\begin{aligned}
 &= 2.5 \\
 d(2/5)9 &= \min \{d29, d59\} \\
 &= \min \{0.8, 0.9\} \\
 &= 0.8 \\
 d(2/5)10 &= \min \{d210, d510\} \\
 &= \min \{3.1, 3.2\} \\
 &= 3.1 \\
 d(2/5)11 &= \min \{d211, d511\} \\
 &= \min \{2.3, 2.4\} \\
 &= 2.3 \\
 d(2/5)12 &= \min \{d212, d512\} \\
 &= \min \{1.7, 1.9\} \\
 &= 1.7 \\
 d(2/5)13 &= \min \{d213, d513\} \\
 &= \min \{16.1, 16.0\} \\
 &= 16.1 \\
 d(2/5)14 &= \min \{d214, d514\} \\
 &= \min \{1.9, 1.8\} \\
 &= 1.8
 \end{aligned}$$

Dari hasil perbandingan dengan mencari nilai minimal dihasilkan matrik baru yang merupakan representasi dari kluster (lawang sewu-Museum Mandala). Dari matrik di cari nilai minimal baru yang digambarkan dengan tabel 5.

Tabel 5. Proses klusterisasi Masjid Kauman-Gedung Marba

No	2/5	1	3	4	6	7	8	9	10	11	12	13	14
2/5	0	1.4	1.9	3.8	1.1	2.9	2.6	8.8	2.9	2.3	1.3	11.3	1.8
1	1.4	0.0	1.9	2.6	2.0	4.3	2.4	10.2	2.8	2.2	2.0	10.9	2.7
3	1.9	1.9	0.0	2.8	1.3	4.6	0.8	10.2	1.1	0.4	0.9	12.7	3.7
4	3.8	2.6	2.8	0.0	3.9	6.8	2.5	12.7	2.9	2.7	3.6	12.0	5.3
6	1.1	2.0	1.3	3.9	0.0	3.3	2.0	8.9	2.1	1.6	0.4	12.5	2.8
7	2.9	4.3	4.6	6.8	3.3	0.0	5.3	5.9	5.5	4.9	3.7	11.6	2.0
8	2.6	2.4	0.8	2.5	2.0	5.3	0.0	10.8	0.4	0.4	1.6	13.2	4.5
9	8.8	10.2	10.2	12.7	8.9	5.9	10.8	0.0	10.9	10.4	9.3	14.9	7.8
10	2.9	2.8	1.1	2.9	2.1	5.5	0.4	10.9	0.0	0.7	1.7	13.7	4.8
11	2.3	2.2	0.4	2.7	1.6	4.9	0.4	10.4	0.7	0.0	1.2	13.1	4.1
12	1.3	2.0	0.9	3.6	0.4	3.7	1.6	9.3	1.7	1.2	0.0	12.6	3.1
13	11.3	10.9	12.7	12.0	12.5	11.6	13.2	14.9	13.7	13.1	12.6	0.0	10.1
14	1.8	2.7	3.7	5.3	2.8	2.0	4.5	7.8	4.8	4.1	3.1	10.1	0.0

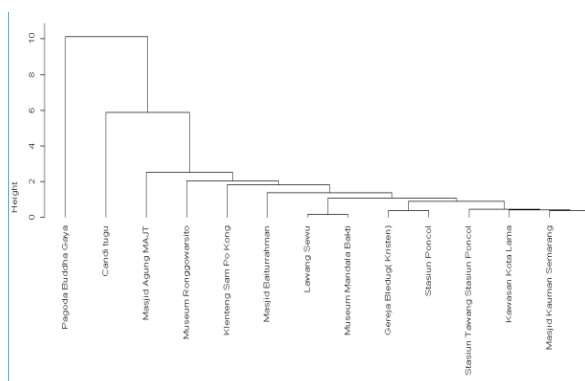
- (b) Dengan demikian matrik jarak menjadi seperti dalam tabel 5 Dari matrik jarak pada tabel 4.5 kembali mencari jarak terdekat. Jarak terdekat yang diperoleh adalah jarak antara Masjid Kauman (Lokasi ke-3) dengan Gedung Marba (Lokasi ke-11) yaitu 0.4 km, sehingga Masjid Kauman dan Gedung Marba menjadi 1 kluster. Sekarang tersisa 10 obyek/kluster. Kemudian dilakukan kembali perbaikan matrik jarak menggunakan persamaan *singlelinkage*.
- (c) Proses perhitungan dilakukan sama seperti diatas sampai menghasilkan matrik dengan dua obyek.



Hasil perhitungan akhir membentuk satu kluster saja dengan keanggotaan {Candi Tugu-Pagoda Buddha Gaya} diberikan di tabel 6 dengan jarak 14.9 km. Untuk memperjelas proses penggabungan satu demi satu dapat digambarkan dalam bentuk Dendrogram yang dilustrasikan pada gambar 2 berikut ini.

Tabel 6. Proses Akhir Kluster

No	9	13
9	0.0	14.9
13	14.9	0.0



Gambar 2. Dendrogram Wisata Kota Semarang

Gambar 2 menunjukkan dendrogram semua kluster yang terbentuk dari proses perhitungan. Dari dendrogram bisa dilihat jumlah cluster yang terbentuk dari 14 obyek wisata. Setiap kelompok yang terbentuk berdasar ketetanggaan terdekat sehingga dihasilkan kluster dengan keanggotaan sebagai berikut:

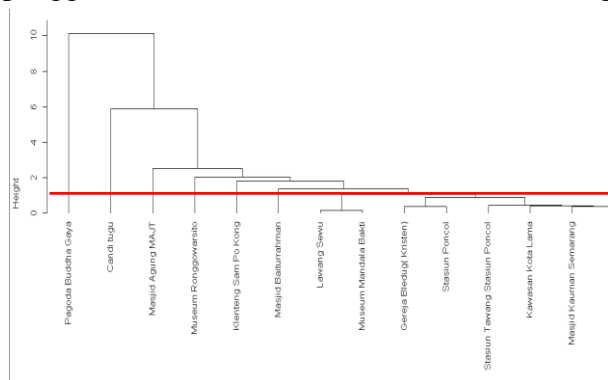
- Kluster pertama dengan keanggotaan {Masjid Kauman Semarang, Gedung Marba} yang berada dalam satu kluster,
- Kluster kedua dengan jarak yang berdekatan membentuk kluster baru dengan keanggotaan {Masjid Kauman Semarang, Gedung Marba} dengan {Kawasan Kota Lama, Stasiun Tawang},
- Kluster ketiga yang terbentuk {gereja Bleduk, Stasiun Poncol}
- Kluster keempat (Lawang Sewu- Museum Mandala Bakti)
- Kluster keenam gabungan dari keanggotaan {gereja Bleduk , Stasiun Poncol} dengan {Masjid Kauman Semarang, Gedung Marba, Kawasan Kota Lama, Stasiun Tawang},
- Kluster ketujuh dengan keanggotaan {{masjid Baiturrahman}{Lawang Sewu, Museum Mandala Bakti}}{Gereja Bleduk, Stasiun Poncol}}{Kawasan Kota Lama, Masjid Kauman, Gedung Marba}}
- Kluster kedelapan dengan keanggotaan {Klenteng Sam Pokong}{{{masjid Baiturrahman} {Lawang Sewu, Museum Mandala Bakti}}{Gereja Bleduk, Stasiun Poncol}}{Kawasan Kota Lama, Masjid Kauman, Gedung Marba}}
- Kluster kesembilan dengan keanggotaan {{museum ronggowarsito}}{Klenteng Sam Pokong} {masjid Baiturrahman} {Lawang Sewu, Museum Mandala Bakti} {Gereja Bleduk, Stasiun Poncol} {Kawasan Kota Lama, Masjid Kauman, Gedung Marba}}
- Kluster kesepuluh {{Masjid Agung MAJT}}{museum ronggowarsito}{{Klenteng Sam Pokong} {masjid Baiturrahman} {Lawang Sewu, Museum Mandala Bakti} {Gereja Bleduk, Stasiun Poncol} {Kawasan Kota Lama, Masjid Kauman, Gedung Marba}}
- Kluster kesebelas {{candi Tuggu}}{Masjid Agung MAJT}{{museum ronggowarsito} {Klenteng Sam Pokong} {masjid Baiturrahman} {Lawang Sewu, Museum Mandala

Bakti}{Gereja Blenduk, Stasiun Poncol}{Kawasan Kota Lama, Masjid Kauman, Gedung Marba}}

- k. Klaster keduabelas {{Pagoda Buddha Gaya}{ candi Tuggu}}{Masjid Agung MAJT}{museum ronggowarsito}{Klenteng Sam Pokong}{masjid Baiturrahman}{Lawang Sewu, Museum Mandala Bakti}{Gereja Blenduk, Stasiun Poncol}{Kawasan Kota Lama, Masjid Kauman, Gedung Marba}}

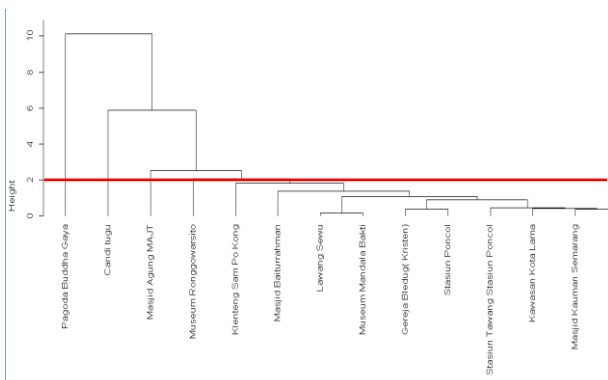
Berdasarkan Klaster yang ada bisa digunakan untuk mencari jarak paling dekat yang bisa dilakukan oleh pengguna (mengunjungi obyek wisata yang berada dalam satu klaster) dengan memberikan batasan/kontrain untuk melihat obyek mana saja yang bisa dikunjungi dengan satu satuan jarak yang dimiliki.

Misal pengguna memberikan batasan waktu yang diinginkan untuk waktu kunjungan dengan jarak 1 km, obyek mana saja yang bisa dilakukan kunjungan berdasar jarak paling dengan dengan posisi pengguna berada. Hasil analisis bisa dilihat di gambar 3.



Gambar 3. Dendrogram Wisata Dengan Batasan 1 km

Saat diberikan batasan 1km, maka terlihat bahwa terbentuk 8 cluster yakni cluster {Gedung Marba, Masjid Kauman, Kawasan Kota Lama, Gereja Blenduk, Stasiun Poncol}, {Museum Mandala Bakti, Lawang Sewu}, {Masjid Baiturrahman}, {Klenteng Sam Po Kong}, {Museum Ronggowarsito}, {Masjid Agung}, {Candi Tugu} dan {Pagoda Budha Gaya}. Pengguna dapat memilih cluster yang paling dekat dengan lokasinya, dan cluster yang berdekatan dengan cluster yang dikunjungi. Pengguna juga dapat menaikkan batasan untuk melihat perubahan jumlah cluster, misalnya batasan dinaikkan menjadi 2 km seperti pada gambar 4.



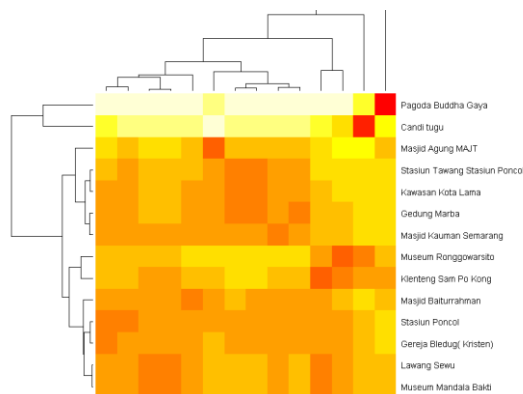
Gambar 4. Dendrogram Wisata Dengan Batasan 2 km

Saat diberikan batasan 2km, maka terlihat bahwa terbentuk 5 cluster yakni cluster {Gedung Marba, Masjid Kauman, Kawasan Kota Lama, Gereja Blenduk, Stasiun Poncol, Museum Mandala Bakti, Lawang Sewu, Masjid Baiturrahman, Klenteng Sam Po Kong}, {Museum Ronggowarsito}, {Masjid Agung}, {Candi Tugu} dan {Pagoda Budha Gaya}.

Dengan menaikkan batasan, maka terlihat bahwa jumlah cluster menjadi berkurang, tetapi jumlah obyek dalam satu cluster menjadi bertambah. Sehingga obyek yang bisa dikunjungi oleh pengguna juga menjadi lebih banyak.

Untuk memvisualisasikan kluster yang terbentuk bisa menggunakan HeatMap-Dendrogram. Heatmap sangat membantu dalam memvisualisasikan data yang sangat besar dalam clustering, cluster yang berdekatan akan divisualkan dengan tingkat kontras warna yang rendah. Semakin jauh jarak cluster maka semakin tinggi tingkat kontras warna. Contohnya adalah obyek wisata Mandala Bakti yang berwarna kuning berjarak sangat jauh dengan Pagoda Budha Gaya yang berwarna merah.

Visualisasi warna yang tingkat kontrasannya berbeda (otliner) sama sekali dengan kluster yang lain, bisa dikatakan bahwa kluster tersebut berada di posisi Terluar dari kelompok kluster yang ada. Di gambar 5 di berikan contoh kluster dengan keanggotaan {Pagoda Budha Gaya, Candi Tugu}.



Gambar 5. Heatmap-Dendrogram

## 5. KESIMPULAN

- Model ini dapat memberikan saran bagi pengguna kumpulan obyek yang harus dikunjungi berdasarkan ketersediaan waktu dengan batasan jarak maksimum antar cluster
- Jumlah cluster fleksibel berdasarkan kebutuhan pengguna
- Heatmap memudahkan pengguna dalam melihat cluster obyek wisata yang akan dikunjungi
- Semakin tinggi nilai batasan, jumlah kluster yang dihasilkan akan semakin sedikit, dan sebaliknya

## 6. SARAN

- Dibutuhkan algoritma lain dengan obyek data yang sama sehingga dapat dihasilkan cluster yang lebih baik
- Dapat dikembangkan menjadi sistem rekomendasi berdasarkan jarak sesungguhnya, bukan berdasarkan perhitungan euclidean distance

## DAFTAR PUSTAKA

- [1] Agarwal, J., Sharma, N., Kumar, P., Parshav, Vishesh, Srivastava, A., Goudar, R.H., 2013, "Intelligent search in E-Tourism services using Recommendation System: Perfect guide for tourist," *Intelligent Systems and Control (ISCO), 2013 7th International Conference on* , vol. 410, no. 415, pp. 4-5 Jan

- [2] Gerharz, M., de Waal, C., Martini, P., James, P., 2003, "A cooperative nearest neighbours topology control algorithm for wireless ad hoc networks," *Computer Communications and Networks*, 2003. ICCCN 2003. Proceedings. The 12th International Conference on , vol. 412, no. 417, pp. 20-22
- [3] Liu, B., 2009, *Web Data Mining : Exploring Hyperlinks, Contents, and Usage Data*, Springer, Heidelberg, Germany
- [4] Kuncoro, M. 2003. *Metode Riset untuk Bisnis dan Ekonomi* . Jakarta: Penerbit Erlangga
- [5] Han, J., Kamber, M., Pei, J.: *Data Mining Concept and Techniques*, 3rd ed. Morgan Kaufmann-Elsevier, Amsterdam (2012)
- [6] Jain. A.K 2009. *Data Clustering: 50 Years Beyond K-Means*. Pattern Recognition Letters, 2009.
- [7] Johnson and Wichern. 2007., *Applied Multivariate Statistical Analysis (6th Edition)*, Prentice Hall.
- [8] Shafiee, S., Shafiee, H., 2014, "Studying the effective factors on domestic tourists trust in offering e-services in e-tourism: Case study Iran Country," *e-Commerce in Developing Countries: With Focus on e-Trust (ECDC)*, 2014 8th International Conference on , vol. 1, no. 9, pp. 24-25
- [9] Yang, W., Miao, R., 2014, "The Application of Cluster Analysis in the Classification of China's Modern Service Industries",
- [10] Zhang, M., Chen, Y., Zhang, X., Lai, J., 2009 "Study on the Recommendation Technology for Tourism Information Service," *Computational Intelligence and Design*, 2009. ISCID '09. Second International Symposium on , vol.1, no. 415, pp. 12-14