

PENERAPAN DATA MINING PADA ALGORITMA *HIERARCHICAL CLUSTERING* TENTANG PENGELOLAAN MITRA PERJALANAN WISATAWAN BALI BACKPAKER

Putu Nandhika Pratama Artana, Eka Prakarsa Mandyartha, Made Hanindia Prami S,

Program Studi Informatika, Fakultas Ilmu Komputer

Universitas Pembangunan Nasional “Veteran” Jawa Timur

Jl. Rungkut Madya No.1, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur

19081010143@student.upnjatim.ac.id

ABSTRAK

Wisatawan merupakan orang yang dimana akan melakukan sebuah kunjungan atau berkunjung ke suatu daerah atau lokasi dari tempat kediamannya tanpa orang tersebut menetap ditempat yang di datangnya atau hanya untuk sementara waktu yang akan di datangnya. Perkembangan teknologi informasi pada era sekarang, sangat cepat dalam penerapan serta pemanfaatannya sehingga teknologi informasi dapat berkembang secara luas merata diberbagai bidang, salah satunya pada backpacker. Dengan berkembangnya teknologi saat ini sangat memudahkan seorang backpacker untuk berpergian menjelajahi dunia terutamanya Bali. Sehingga permasalahan yang ada di Bali banyak backpacker yang kesusahan untuk mencari sebuah rekomendasi jarak mulai dari rekomendasi hotel, wisata, souvenir, dan tempat ibadah. Penelitian ini bertujuan juga untuk menghasilkan rekomendasi jarak dari backpacker yang berpergian ke pulau Bali yang direkomendasikan pada kolom pencarian nantinya. Penelitian ini menggunakan sebuah Metode Hierarchical Clustering dengan perhitungan Manhattan Distance yang dimana dilakukan untuk menentukan obyek mana saja yang memiliki lokasi berdekatan dari backpacker tersebut sehingga dimana bisa membentuk adanya klaster yang dilakukan berdasarkan jarak paling dekat dengan klaster menuju klaster berikutnya sesuai dengan kelompok obyek dari Hotel, Wisata, Souvenir dan Tempat Ibadah. Sehingga hasil akurasi yang diperoleh sebesar 86,90 % dan nilai Precision sebesar 96,5% serta nilai Recall sebesar 91,2%.

Kata kunci: *Hierarchical Clustering, Backpacker, Teknologi Informasi*

1. PENDAHULUAN

Wisatawan merupakan orang yang mengadakan sebuah perjalanan dari tempat kediamannya tanpa menetap ditempat yang didatanginya, atau hanya untuk sementara waktu yang ditempat tinggal didatanginya. Mereka yang dianggap wisatawan adalah orang yang melakukan sebuah kesenangan. Pertumbuhan pariwisata di Indonesia mulai sangat berkembang terutama pada sebuah destinasi wisata yang sangat menarik. pulau yang ada di Indonesia salah satunya yaitu pada pulau Bali yang dimana Bali sangat dikenal oleh para wisatawan asing dari mancanegara maupun para wisatawan dalam negeri (domestik). Pulau Bali juga memiliki nama sebutan yaitu yang dikenal sebagai Pulau Dewata dikarenakan keindahan sebuah alamnya yang dimilikinya tersebut. Bali juga banyak menyediakan tempat wisata yang bisa di tuju oleh para wisatawan mancanegara maupun wisatawan dalam negeri (domestik) yaitu mulai dari wisata Pura Tanah Lot, Pantai Kuta, Pura Uluwatu, Pura Ulun Danu Beratan Bedugul, Gunung Batur Kintamani, Bali Safari, Pantai Tanjung Benoa, Ubud Monkey Forest, Seminyak, Pantai Pandawa, dan masih banyak lagi tempat wisata yang bisa di kunjungi oleh wisatawan tidak hanya wisata alam saja ada juga tempat ibadah yang bisa di kunjungi oleh wisatawan, ada juga tempat oleh-oleh, maupun hotel yang bisa untuk wisatawan menginap pada saat di Bali nantinya.

Permasalahan yang timbul di pulau Bali dimana banyaknya sebuah wisatawan asing mancanegara

maupun domestik banyak kebingungan terkait tempat wisata, hotel, tempat pembelian souvenir dan tempat ibadah. Sehingga diperlukannya sebuah pengelompokkan berdasarkan sebuah jarak rekomendasi wisatawan dalam berkunjung serta terhadap jalan yang macet. Dimana data pengelompokkan tersebut merupakan sebuah mitra perjalanan wisatawan Bali Backpacker. Sehingga manajemen wisata bisa teroptimalisasi dengan adanya sebuah pengelompokkan tersebut.

Berdasarkan pengelompokkan dalam sebuah mitra perjalanan wisatawan di pulau Bali dalam menggunakan algoritma *Hierarchical clustering* memiliki tujuan utama untuk mengatasi permasalahan yang timbul terkait kebingungan wisatawan terhadap tempat wisata, hotel, souvenir, dan tempat ibadah. Dengan mengelompokkan wisatawan berdasarkan jarak rekomendasi dalam berkunjung serta kondisi jalan yang macet, manajemen wisata dapat ditingkatkan secara signifikan. Pengelompokkan ini memungkinkan wisatawan untuk memiliki panduan yang lebih jelas dan terstruktur mengenai lokasi-lokasi penting yang ingin mereka kunjungi.

Algoritma ini merupakan metode analisis kelompok yang berusaha untuk membangun sebuah hirarki kelompok data. Dimana pada strategi pengelompokannya umumnya ada 2 jenis yaitu yang dimana pada agglomerative (Bottom-Up) dan Divisive (Top-Down) [1]. *Clustering* adalah contoh dari klasifikasi yang tidak mempunyai arah atau

disebut *unsupervised*. *Unsupervised* berarti pengelompokan tidak bergantung pada standard kelas ataupun pelatihan atau training [2].

2. TINJAUAN PUSTAKA

2.1. Pengertian Data Mining

Data mining adalah proses menganalisis sebuah data dan menemukan pola tersembunyi secara otomatis maupun semi otomatis. Pola atau hubungan digunakan sebagai alat untuk menjelaskan sebuah data dan memprediksi data yang baru tersebut. Ada beberapa definisi lain dari data mining tersebut. Dimana data mining adalah sebuah proses menganalisis sebuah data yang sangat besar untuk menemukan sebuah hubungan dan perlu diekstraksi agar menjadi informasi baru dan dapat membantu dalam pengambilan keputusan.

2.2. Clustering

Clustering adalah suatu metode pengelompokan data kedalam beberapa cluster atau kelompok sehingga data yang ada dalam satu cluster mempunyai tingkat kemiripan yang maksimal sedangkan data antar cluster yang berbeda mempunyai tingkat kemiripan yang minimal [2]. Menurut Madhu Yedha *clustering* adalah proses pengorganisasian objek yang berupa data-data kedalam set kelas yang saling berhubungan satu dengan yang lainnya yang disebut *cluster*. *Clustering* adalah contoh dari klasifikasi yang tidak mempunyai arah atau disebut *unsupervised*. *Unsupervised* berarti pengelompokan tidak bergantung pada standard kelas ataupun pelatihan atau training [3].

Clustering pada data yang bersifat kategorial sudah mulai berkembang walaupun masih kalah jauh jika dibandingkan dengan clustering pada data yang bersifat numerik. Data kategorial seara alami tidak bisa diperlakukan sebagaimana data numerik karena beberapa operasi yang ada dalam data numerik yang tidak bisa untuk dilakukan dalam sebuah data kategorial seperti mean dan median. *Clustering* adalah alat utama dalam banyak aplikasi dalam analisa data statistik, data mining, information retrieval, pengolahan citra, dan lain sebagainya. *Clustering* bertujuan untuk mengelompokkan objek-objek yang berupa data kedalam beberapa cluster atau kelompok sehingga objek-objek yang berbentuk data tadi yang berada dalam satu cluster yang sama akan memiliki tingkat kemiripan yang sangat tinggi sedangkan data yang berada di klaster yang berbeda akan memiliki tingkat kemiripan yang bisa dibilang rendah. *Clustering* adalah untuk mempartisi suatu set objek dalam menguraikan cluster homogen, sehingga nantinya objek milik cluster yang sama akan mirip satu dengan yang lainnya dibandingkan objek yang masuk pada kelompok yang berbeda.

2.3. Metode Dalam Analisis Cluster

Pada analisis *cluster* terdapat dua buah metode yang dapat digunakan, yaitu pada metode *Hierarchical* dan *non-Hierarchical*. Pengelompokan pada

Hierarchical Clustering adalah metode yang menganalisis *cluster* dengan cara membangun sebuah hierarki kelompok tersebut. Strategi untuk pengelompokan hierarki pada umumnya bisa dibagi menjadi dua jenis yaitu *agglomerative* (pemusatan) dan *divisive* (penyebaran). Metode *agglomerative* merupakan metode pengelompokan hierarki dengan pendekatan bawah-atas (*bottom up*). Metode *agglomerative* biasanya dipergunakan pada bidang ekonomi dan social masyarakat. Adapun algoritma pada pengelompokan sebuah hierarki dengan menggunakan metode *agglomerative* (pemusatan) sebagai berikut:

1. Single Linkage

Single Linkage ini merupakan prosedur pengelompokan *agglomerative* berdasarkan jarak terkecil antar objek. Algoritma pengelompokan *single linkage* diawali dengan memilih jarak terkecil dalam matriks $D = \{d_{ij}\}$, kemudian menggabungkan sebuah objek yang bersesuaian misalnya U dan V untuk memperoleh *cluster* (UV). Langkah berikutnya adalah mencari jarak antara (UV) dengan *cluster* lainnya, misalnya W sehingga dapat dituliskan sebagai berikut dibawah ini:

$$d_{(UV)W} = \min(d_{UW}, d_{VW}) \quad (1)$$

Jadi dengan d_{UW} merupakan jarak tetangga terdekat dari *cluster* U dan W serta d_{VW} adalah jarak tetangga terdekat dari *cluster* V dan W.

2. Average Linkage

Average Linkage merupakan prosedur pengelompokan *agglomerative* berdasarkan rata-rata antar sebuah objek. Algoritma *average linkage* diawali dengan mendefinisikan matriks $D = \{d_{ij}\}$ untuk memperoleh objek yang paling dekat, sebagai contoh U dan V, kemudian objek ini digabung ke dalam suatu bentuk sebuah cluster (UV) dan selanjutnya jarak antara (UV) dengan cluster lainnya W, sehingga dapat dituliskan sebagai berikut dibawah ini:

$$d_{(UV)W} = \frac{d_{(UV)} + d_{(VW)}}{n_{(UV)}n_W} \quad (2)$$

Dengan $n_{(UV)}$ adalah banyaknya anggota dalam *cluster* (UV) dan n_W adalah banyaknya anggota dalam *cluster* W tersebut.

3. Jarak Manhattan Distance

Jarak *Manhattan Distance* adalah metode pengukuran jarak antara dua titik dalam ruang geometri. *Manhattan distance* yang dimana mengukur sebuah perbedaan absolut antara nilai koordinat pada masing-masing sumbu antara dua titik data. Untuk titik A dengan koordinat (x1, y1) dan titik B dengan koordinat (x2, y2), pada rumus dari sebuah *Manhattan distance* dapat dilihat dibawah ini:

$$\text{Manhattan Distance} = |x_1 - x_2| + |y_1 - y_2| \quad (3)$$

Keterangan:

x_1, x_2 : Koordinat sumbu x dari sebuah titik

y_1, y_2 : Koordinat sumbu y dari sebuah titik

3. METODE PENELITIAN

3.1. Pengumpulan Data

Untuk mendukung penelitian pada kali ini dimana akan dilakukannya sebuah proses dari pengumpulan data. Pada penelitian ini dalam pengumpulan sebuah data dilakukan untuk mendapatkan sumber dataset yang akan digunakan dengan pengumpulan data seperti hotel, wisata, souvenir dan tempat ibadah di daerah bali yang berasal dari website traveling dan google maps. Jumlah data sampel yang dipakai pada penelitian ini sebanyak 336 data dan terdiri dari 6 atribut.

3.2. Pengolahan Data

Pada tahapan dari pengolahan data ini yaitu penelitian dimana harus terdapat sebuah pengolahan data, sehingga dimana data yang akan di olah merupakan sebuah data yang telah diperoleh sebelum dilakukannya pengolahan yang akan di analisis datanya lalu dilakukan dengan pembersihan data pemilihan sebuah atribut sebelum menguji sebuah hasil yang menggunakan algoritma *Hierarchical Clustering* dengan perhitungan *Manhattan Distance* tersebut.

3.3. Menentukan Fungsi Jarak Manhattan Distance

Manhattan distance adalah metode pengukuran jarak antara dua titik dalam ruang geometri. Manhattan distance yang dimana mengukur sebuah perbedaan absolut antara nilai koordinat pada masing-masing sumbu antara dua titik data. Untuk titik A dengan koordinat (x_1, y_1) dan titik B dengan koordinat (x_2, y_2), pada rumus dari sebuah Manhattan distance dapat dilihat dibawah ini:

$$\text{Manhattan Distance} = |x_1 - x_2| + |y_1 - y_2| \quad (4)$$

Keterangan:

x_1, x_2 : Koordinat sumbu x dari sebuah titik

y_1, y_2 : Koordinat sumbu y dari sebuah titik

Dibawah ini merupakan sebuah contoh perhitungan dari rumus *Manhattan Distance* di atas sebagai berikut:

$$\begin{aligned} D_{man} &= |(-8.144374) - (-8.70400)| \\ &\quad + |114.656581 - 115.17990| \\ D_{man} &= -7.273.974 + 103.138.591 \\ D_{man} &= 95.864.617 \end{aligned}$$

Tabel 1. Tabel Rata-Rata Latitude dan Longitude

Mitra	Latitude	Longitude
Amertha Bali Villas	-8.144374	114.656581
Krisna Oleh - Oleh Bali Sunset Road	-8.70400	115.17990
Pura Tanah Lot	-6.153308	106.713137
Prana Dewi Mountain Resort	-8.388295	115.109190
Masjid Besar Al Hidayah	-8.27791	115.16530

3.4. Proses Klusterisasi

Dalam proses klusterisasi dimana melaksanakan sebuah perhitungan fungsi jarak pada Manhattan Distance dengan berdasarkan perhitungan secara diagonal dengan fungsi pythagoras sehingga jarak yang diperoleh secara langsung ditarik garis lurus pada dua titik yang dimana lokasi yang diwakili dengan koordinat dari masing-masing mitra Bali Backpacker tersebut dan setelah itu dihitung. Sedangkan untuk jarak tempuh berdasarkan maps merupakan sebuah perhitungan yang dilakukan berdasarkan panjang jalan yang bisa dilalui. Seperti pada tabel dibawah ini:

Tabel 2. Proses Klusterisasi *Manhattan Distance*

No	Mitra	1	2	3	4	5
1	Amertha Bali Villas	0	95.864.617	5.952.378	-208.688	95.823.468
2	Krisna Oleh - Oleh Bali Sunset Road	-95.864.617	0	-89.912.239	-96.073.305	-41.149
3	Pura Tanah Lot	-5.952.378	89.912.239	0	-6.161.066	89.871.090
4	Prana Dewi Mountain Resort	208.688	96.073.305	6.161.066	0	96.032.156
5	Masjid Besar Al Hidayah	-95.823.468	41.149	-89.871.090	-96.032.156	0

Menghitung sebuah jarak yang dimana antar kelompok (2 dan 4) dengan kelompok lain yang tersisa, yaitu seperti dibawah ini mendapatkan data 1,3 dan 5.

$$\begin{aligned} d_{(24)3} &= \min\{d_{23}, d_{43}\} \\ &= \min\{-89.912.239, 6.161.066\} \\ &= -89.912.239 \\ d_{(24)1} &= \min\{d_{21}, d_{41}\} = \min\{-96.073.305, 0\} \\ &= -96.073.305 \end{aligned}$$

$$\begin{aligned} d_{(24)5} &= \min\{d_{25}, d_{45}\} \\ &= \min\{-41.149, 96.032.156\} = \\ &= -41.149 \end{aligned}$$

Jadi dibawah ini merupakan sebuah tabel yang dimana dari hasil penentuan sebuah kelompok lain yang tersisa pada saat melakukan proses klusterisasi awal yaitu data yang tersisa adalah 1,3 dan 5.

Tabel 3. Proses 4 Klasterisasi

No	Mitra	1	(2,4)	3	5
1	Amertha Bali Villas	0	-96.073.305	5.952.378	95.823.468
(2,4)	Krisna Oleh - Oleh Bali Sunset Road dan Prana Dewi Mountain Resort	-96.073.305	0	-89.912.239	-41.149
3	Pura Tanah Lot	-5.952.378	89.912.239	0	89.871.090
5	Masjid Besar Al Hidayah	-95.823.468	41.149	-89.871.090	0

Jadi untuk $\min(D_{man}) = \min(d_{(24)5}) = -41.149$
 Menghitung sebuah jarak antar kelompok ((24) dan 5)
 dengan kelompok lain yang tersisa, yaitu 1 dan 3.

$$\begin{aligned}
 d_{(245)1} &= \min\{d_{21}, d_{41}, d_{51}\} \\
 &= \min\{95.864.617, -208.688, 95.823.468\} \\
 &= 5.952.370
 \end{aligned}$$

$$\begin{aligned}
 d_{(245)3} &= \min\{d_{23}, d_{43}, d_{53}\} \\
 &= \min\{89.912.239, -6.161.066, 89.871.090\} \\
 &= -6.161.066
 \end{aligned}$$

Tabel 4. Proses 3 Klasterisasi

No	Mitra	(2,4,5)	1	3
(2,4,5)	Krisna Oleh - Oleh Bali Sunset Road, Prana Dewi Mountain Resort dan Masjid Besar Al Hidayah	0	-208.688	-6.161.066
1	Amertha Bali Villas	-208.688	0	5.952.378
3	Pura Tanah Lot	-6.161.066	-5.952.378	0

$\min(D_{man}) = \min(d_{(245)1}) = -208.688$ Menghitung
 sebuah jarak antar kelompok ((245) dan 4) dengan
 kelompok lain yang tersisa, yaitu 3.

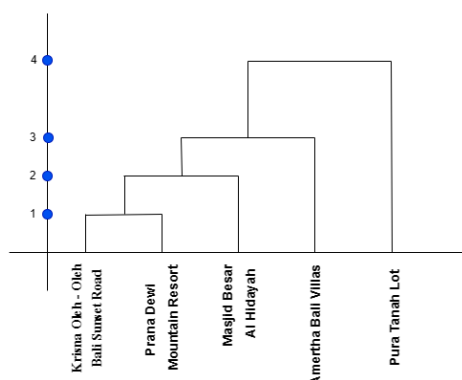
$$d_{(2451)(3)}$$

$$\begin{aligned}
 &= \min\{d_{23}, d_{43}, d_{53}, d_{13}\} \\
 &= \min\{89.912.239, -6.161.066, 89.871.090, -5.952.378\} \\
 &= -6.161.066
 \end{aligned}$$

Tabel 5. Proses Akhir Klasterisasi

No	Mitra	(2,4,5,1)	3
(2,4,5,1)	Krisna Oleh - Oleh Bali Sunset Road, Prana Dewi Mountain Resort, Masjid Besar Al Hidayah dan Amertha Bali Villas	0	-616.066
3	Pura Tanah Lot	-616.066	0

Jadi di atas ini merupakan hasil akhir dari sebuah klasterisasi menggunakan *Hierarchical Clustering* dengan perhitungan dari *Manhattan Distance*. Ada juga untuk melihat hasil dendrogram dari klasterisasi wisatawan bisa dilihat dibawah ini:



Gambar 1. Hasil Dendrogram Perjalanan Wisatawan

4. HASIL DAN PEMBAHASAN

Penelitian ini pada proses penerapan data mining dengan memakai software jupyter notebook dan algoritma *Hierarchical Clustering* dengan perhitungan *Manhattan Distance*. Tujuan penelitian kali ini untuk menghasilkan rekomendasi sebuah jarak terdekat hotel, wisata, souvenir dan tempat ibadah dengan

lokasi yang dicari wisatawan. Pada table 6 menunjukkan deskripsi atribut dari dataset yang di pakai untuk penelitian ini.

Tabel 6. Deskripsi atribut dataset

No	Atribut	Keterangan
1	Nama	Nama dataset
2	Alamat	Mengetahui identitas dataset
3	Latitude	Sistem koordinat geografis untuk menunjukkan posisi suatu titik dipermukaan bumi
4	Longitude	Penentuan navigasi dan penentuan waktu
5	Deskripsi	Memberikan informasi
6	Harga	Menyatakan harga dari dataset

4.1. Filtering Data Latitude dan Longitude

Filtering ini merupakan sebuah hasil dari data latitude dan longitude dari dataset hotel, wisata, souvenir dan tempat ibadah yang dimana sudah di filtering dari 336 data dan gambar dibawah ini adalah potongan tampilan hasil filtering data latitude dan longitude dari jupyter notebook tersebut.

	NAMA	ALAMAT	LATITUDE	LONGITUDE	DIKEMPER
0	Janettha Bai Villas	Jl. Satek Glimanak No.10, Pemaderan, Kec. S...	-1144274	114000001	Janettha Bai Villas adalah sebuah resort yang...
1	Sumbakina Hill Resort	Sumbakina, Genggah, Buleleng Regency, Bali	-8101629	114000000	Sumbakina Hill Resort terletak di Desa Sumb...
2	Tika Bai Homestay	Jl. Satek Glimanak, Pemaderan, Kec. Genggah	-1140334	114031100	Kantor Superior Double atau Twin dengan 1 P...
3	Vila Senadhi	Pemaderan, Genggah, Buleleng Regency, Bali	-1140101	114014032	Menyajikan pemandangan laut dan panorama bu...
4	Biorok Homestay	Pemaderan, Genggah, Pemaderan, Genggah, Bali	-1140335	114044313	Biorok Homestay menyediakan akomodasi yang...
5	NORA QUINOLA Resort and Spa	Jl. Banyuwedang, Pajajaran, Kec. Genggah, Bali	-1130002	114000407	Nora Quinola Resort and Spa memiliki sebuah ...
6	Adi Anas Beach Resort & Spa	Jl. Satek Glimanak, Pemaderan, Kec. Genggah	-1140057	114000034	Adi Anas Beach Resort dan spa menyediakan ...
8	Margo Tree Inn	Jl. Genggah Glimanak, Pemaderan, Kec. Genggah	-1140057	114003140	Margo Tree Inn menawarkan akomodasi dengan ...
9	Habit Paradiis Sari Beach Villa Resort/ Deluxe Bu	Pemaderan, Genggah, Buleleng Regency, Bali	-1140329	114000011	Menyediakan akomodasi 24 jam dan kolam renang ...
10	The Mangrove by Libellulafabrics	Jl. Raya Denpasar - Glimanak No.10, K...	-1140676	114047712	The Mangrove by Libellulafabrics terletak di ...
14	Residence Jungle Resort and Spa Bali	Kuta J. Poppes 1 Gg. Sopa, Kuta, Kuta, Bali	-1141539	115230306	Selain standar Indonesia Care, semua kam...
15	Kupa Kupa Barong Villas & Tree Spa by L'Occitane	Kedondutan P.O. Box 7 Ubud, Kedondutan, Ubud, Bali	-1141700	115240000	Kupa Kupa Barong Villas and Tree Spa by L'Occitane
16	Tiyana Resort Canggang	Jl. Tabak Ajiyung 103, Canggang, Kec. Petang...	-1140104	115224000	Tiyana Resort Canggang adalah Resort dengan ...
20	Yi Anas Bai	Jl. Mengenta Barong Kedondutan, Mengenta, Kec. Pa...	-1150170	115142110	Yi Anas Bai menawarkan akomodasi di Pantai...
21	Manikpa & Rio Carbon Reserve	Jl. Raya Kedondutan, Banjar, Kedondutan, Bali	-1140402	115240002	Manikpa & Rio Carbon Reserve adalah sebuah ...

Gambar 2. Hasil Filtering Data

4.2. Menghitung Jarak Manhattan Distance

Menghitung jarak *Manhattan distance* ini merupakan sebuah hasil dari penampilan perhitungan jarak *Manhattan distance* antara dataset yang dimana pada dataset tersebut akan menghitung jarak dengan *metric='cityblock'* dan mendapatkan sebuah potongan hasil seperti dibawah ini pada jupyter notebook.

```
[54228. 5577. 21178. ... 19467. 7201. 12266.]
```

Gambar 3. Hasil Jarak Manhattan Distance

4.3. Melakukan Hierarchical Clustering Complete Linkage

Complete linkage ini merupakan sebuah hasil dari hierarchical clustering complete linkage yang dimana akan melakukan perhitungan proses complete linkage. Pada gambar dibawah ini merupakan potongan hasil dari complete linkage pada jupyter notebook.

```
[4.40000000e+01 1.05000000e+02 2.68000000e+02 2.00000000e+00]
[6.40000000e+01 8.40000000e+01 4.11000000e+02 2.00000000e+00]
[6.70000000e+01 1.08000000e+02 5.92000000e+02 2.00000000e+00]
[1.06000000e+02 1.07000000e+02 6.12000000e+02 2.00000000e+00]
[5.70000000e+01 6.30000000e+01 6.42000000e+02 2.00000000e+00]
[6.80000000e+01 8.30000000e+01 6.80000000e+02 2.00000000e+00]
[4.00000000e+00 9.00000000e+00 8.04000000e+02 2.00000000e+00]
[3.70000000e+01 6.20000000e+01 1.11700000e+03 2.00000000e+00]
[2.20000000e+01 2.50000000e+01 1.27100000e+03 2.00000000e+00]
[8.00000000e+00 1.16000000e+02 1.69000000e+03 3.00000000e+00]
[8.20000000e+01 9.20000000e+01 2.06100000e+03 2.00000000e+00]
[1.30000000e+01 2.60000000e+01 2.10200000e+03 2.00000000e+00]
[4.70000000e+01 8.80000000e+01 2.16000000e+03 2.00000000e+00]
[7.70000000e+01 1.14000000e+02 2.17700000e+03 3.00000000e+00]
[8.10000000e+01 9.50000000e+01 2.53700000e+03 2.00000000e+00]
[6.90000000e+01 9.60000000e+01 2.66800000e+03 2.00000000e+00]
[1.12000000e+02 1.15000000e+02 2.97500000e+03 4.00000000e+00]
[5.60000000e+01 1.01000000e+02 3.00400000e+03 2.00000000e+00]
[1.50000000e+01 1.02000000e+02 3.08800000e+03 2.00000000e+00]
```

Gambar 4. Hasil Complete Linkage

4.4. Mengubah Matriks Jarak Menjadi Matriks Persegi

Mengubah matriks jarak menjadi matriks persegi ini merupakan sebuah hasil dari mengubah matriks jarak menjadi matriks persegi yang dimana menggunakan variabel *dist_matrix_square* dan memanggil variabel dari perhitungan jarak Manhattan distance antar data tersebut. Seperti pada gambar dibawah ini yang dimana potongan jupyter notebook dari hasil mengubah matriks jarak menjadi matriks persegi.

```
[ [ 0. 54228. 5577. ... 971456. 951989. 964255. ]
[ 54228. 0. 48651. ... 1012374. 992907. 1005173. ]
[ 5577. 48651. 0. ... 972705. 953238. 965504. ]
...
[ 971456. 1012374. 972705. ... 0. 19467. 7201. ]
[ 951989. 992907. 953238. ... 19467. 0. 12266. ]
[ 964255. 1005173. 965504. ... 7201. 12266. 0. ] ]
```

Gambar 5. Hasil Mengubah Matriks Jarak

4.5. Mentransformasi Matriks Persegi Menjadi Dataframe

Mentransformasi matriks ini merupakan sebuah hasil dari mentransformasikan sebuah matriks persegi menjadi sebuah dataframe yang dimana memasukkan perhitungan dari matriks jarak menjadi matriks persegi lalu indexnya merupakan dataset itu sendiri dan kolom juga sama menggunakan dataset. Karena yang dihitung pada mentransformasikan sebuah hasil matriks persegi menjadi sebuah dataframe tersebut seperti dibawah ini yang dimana sesuai dengan jumlah dari dataset yang ada pada penelitian tersebut. Dibawah ini merupakan hasil transformasi matriks tersebut.

```
0 0 0 1 2 3 4 5 6 \
0 0.0 54228.0 5577.0 21178.0 3407.0 95596.0 4136.0
1 54228.0 0.0 48651.0 67952.0 53099.0 54678.0 57998.0
2 5577.0 48651.0 0.0 22427.0 4448.0 94347.0 9347.0
3 21178.0 67952.0 22427.0 0.0 24585.0 116774.0 17042.0
4 3407.0 53099.0 4448.0 24585.0 0.0 92189.0 7543.0
... ..
105 915411.0 956329.0 916660.0 894233.0 918818.0 1011007.0 911275.0
106 971598.0 1012516.0 972847.0 950420.0 975905.0 1067194.0 967462.0
107 971456.0 1012374.0 972705.0 950278.0 974863.0 1067052.0 967320.0
108 951989.0 992907.0 953238.0 930811.0 955396.0 1047585.0 947853.0
109 964255.0 1005173.0 965504.0 943077.0 967662.0 1059851.0 960119.0

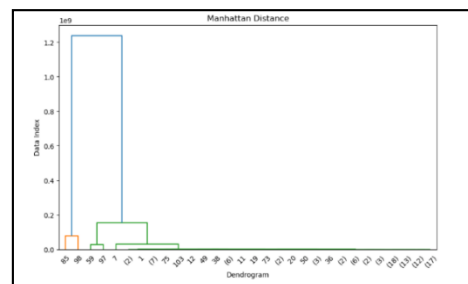
0 7 8 9 ... 100 101 102 \
0 30003403.0 3649.0 2815.0 ... 975037.0 964940.0 926051.0
1 30004321.0 52213.0 53903.0 ... 1015955.0 1005858.0 966969.0
2 30004652.0 3562.0 5252.0 ... 976286.0 966189.0 927300.0
3 29982225.0 24827.0 23993.0 ... 953859.0 943762.0 904873.0
4 30006810.0 886.0 804.0 ... 978444.0 968347.0 929458.0
... ..
105 29727242.0 919060.0 918226.0 ... 59626.0 49529.0 10640.0
106 29736525.0 975247.0 974413.0 ... 12705.0 12446.0 45547.0
107 29735913.0 975105.0 974271.0 ... 12093.0 13058.0 45405.0
108 29725706.0 955638.0 954804.0 ... 23048.0 23265.0 25938.0
109 29733454.0 967904.0 967070.0 ... 10782.0 15517.0 38204.0

0 103 104 105 106 107 108 109
0 1027826.0 949278.0 915411.0 971598.0 971456.0 951989.0 964255.0
1 1068744.0 990196.0 956329.0 1012516.0 1012374.0 992907.0 1005173.0
2 1029075.0 950527.0 916660.0 972847.0 972705.0 953238.0 965504.0
3 1006648.0 928100.0 894233.0 950420.0 950278.0 930811.0 943077.0
4 1031233.0 952685.0 918818.0 975005.0 974863.0 955396.0 967662.0
... ..
105 550353.0 33867.0 0.0 56187.0 56045.0 36578.0 48844.0
106 559636.0 24354.0 56187.0 0.0 612.0 19609.0 7343.0
107 559024.0 24966.0 56045.0 612.0 0.0 19467.0 7201.0
```

Gambar 6. Hasil Mentransformasi Matriks Persegi Menjadi Dataframe

4.6. Menampilkan Hasil Dendrogram

Dendrogram ini merupakan sebuah hasil dari klusterisasi Manhattan distance yang dimana sebuah hasil dari diagram dendrogram tersebut akan ditampilkan dalam bentuk dendrogram yang sudah di klusterisasi dari dataset tersebut. Seperti hasil pada gambar dibawah ini adalah hasil dendrogramnya.



Gambar 7. Hasil Dendrogram

Pada pengujian sistem dan perhitungan manual ini, dimana dilakukannya sebuah perhitungan kecocokan dari sebuah akurasi sistem serta kesesuaian dari perhitungan manual dengan menggunakan *Hierarchical Clustering Manhattan Distance*. Dalam pengujian sistem dan perhitungan manual ini dimana menggunakan data uji seperti data Hotel sebanyak 110 data, Wisata sebanyak 96 data, Souvenir sebanyak 20 data dan Tempat Ibadah sebanyak 110 data. Sehingga tabel dibawah ini merupakan pengujian akurasi rekomendasi jarak terdekat sistem.

Tabel 7. *Confusion Matrix* Rekomendasi Jarak Terdekat Sistem

Rekomendasi	P01	P02	P03	P04	Jumlah
P01	81	29	0	0	110
P02	0	96	0	0	96
P03	0	0	20	0	20
P04	15	0	0	95	110
Jumlah	96	125	20	95	336

Perhitungan sebuah akurasi rekomendasi jarak terdekat sistem dilakukan dengan menjumlah nilai dari *True Positive* (TP), kemudian membagi nilai dari total *True Positive* (TP) tersebut dengan jumlah data yang di uji kemudian dilakukan perkalian yaitu dikali dengan 100%, seperti pada perhitungan dibawah ini.

$$\begin{aligned} \text{True Positive (TP)} &= 81 + 96 + 20 + 95 = 292 \\ \text{Jumlah Data Uji} &= 336 \end{aligned}$$

$$\text{Accuracy} = \frac{292}{336} \times 100\% = 86.90\%$$

Dari perhitungan akurasi di atas, didapatkan hasil akurasi sebesar 80.89%, dengan jumlah data uji sebanyak 336 buah data dengan data uji yang sesuai yaitu sebanyak 292 data dan data uji yang tidak sesuai sebanyak 44 data uji. Kemudian dilakukannya sebuah perhitungan *precision*, sebagai berikut.

Tabel 8. Perhitungan *Precision*

	P01	P02	P03	P04
<i>True Positive</i> (TP)	81	96	20	95
<i>False Positive</i> (FP)	0	0	0	15
<i>Precision</i>	81 / (81 + 0) = 1	96 / (96 + 0) = 1	20 / (20 + 0) = 1	95 / (95 + 15) = 0,86

$$\text{Precision} = \frac{1 + 1 + 1 + 0,86}{4} \times 100\% = 96.5\%$$

Pada hasil perhitungan *precision* diatas, diperoleh sebuah presentase dari nilai *precision* sebesar 96,5%. Terdapat sebuah beberapa kesalahan pada data pengujian, diantaranya pada dataset Hotel diperoleh nilai *False Positive* (FP) sebesar 29, lalu

pada dataset Ibadah terdapat nilai *False Positive* (FP) sebesar 15. Selanjutnya dilakukannya perhitungan *recall*, sebagai berikut.

Tabel 9. Perhitungan *Recall*

	P01	P02	P03	P04
<i>True Positive</i> (TP)	81	96	20	75
<i>False Negatif</i> (FN)	44	0	0	0
<i>Recall</i>	81 / (81 + 44) = 0,64	96 / (96 + 0) = 1	20 / (20 + 0) = 1	75 / (75 + 0) = 1

$$\text{Recall} = \frac{0,64 + 1 + 1 + 1}{4} \times 100\% = 91,2\%$$

Pada hasil perhitungan *Recall* diatas, diperoleh sebuah presentase dari nilai *Recall* sebesar 91,2%. Terdapat beberapa kesalahan pada data pengujian diantaranya pada data Hotel yang dimana terdapat memperoleh nilai *False Negative* (FN) sebesar 44 data.

5. KESIMPULAN DAN SARAN

Berdasarkan sebuah penelitian yang sudah dilakukan dan berhasil menyelesaikan masalah dalam penerapan algoritma *hierarchical clustering* untuk menentukan atau rekomendasi jarak terdekat pada backpacker yang ada di pulau Bali. Berikut ini adalah beberapa kesimpulan yang bisa di ambil dari hasil penelitian ini seperti proses rekomendasi pada backpacker yang ada di Bali bisa dibantu dengan implementasi sebuah algoritma *Hierarchical Clustering* dengan menerapkan sebuah algoritma dari perhitungan Manhattan Distance kedalam sebuah system website rekomendasi jarak terdekat oleh pada backpacker yang ada di pulau Bali. Sehingga dari proses pengujian sistem dengan menggunakan 4 dataset yaitu Hotel, Wisata, Souvenir dan Ibadah memiliki banyak data sebesar 336 data uji, yang dimana diperoleh sebuah hasil akurasi sebesar 86,90% dan nilai *precision* sebesar 96,5% serta nilai *recall* sebesar 91,2%. Algoritma *Hierarchical Clustering* juga memiliki sebuah kelemahan tersendiri yang dimana cenderung sensitive jika terdapat adanya sebuah Outliers dalam sebuah data pada dataset hotel, wisata, souvenir, dan ibadah, yang dimana akan dapat mempengaruhi sebuah pembentukan klaster dan menghasilkan klaster-klaster yang tidak relevan dan tidak representative. Adapun saran dari peneliti berikutnya Dapat menggunakan sebuah metode atau algoritma selain *Hierarchical clustering* atau bisa dikombinasikan dengan algoritma atau metode lainnya, sehingga dapat bisa membandingkan sebuah rekomendasi jarak dengan algoritma selain *hierarchical clustering*. Dapat menambahkan sebuah data mulai hotel, wisata, souvenir, maupun ibadah yang lebih banyak, supaya rekomendasi jarak terdekat bisa lebih efektif, dan yang lebih penting bisa diimplementasikan kedalam bentuk aplikasi mobile.

DAFTAR PUSTAKA

- [1] Febuariyanti, H. & Santoso, D. "HIERARCHICAL AGGLOMERATIVE CLUSTERING UNTUK PENGELOMPOKAN SKRIPSI MAHASISWA," *Prosiding SINTAK*, 2017.
- [2] Gede, I., Komang, I., Tri, N., & Hardiantara, K. "Implementasi Metode Analytical Hierarchy Process Dan Interpolasi Linier Dalam Penentuan Lokasi Wisata Di Kabupaten Karangasem," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, pp. 866-878, 2021.
- [3] Tan, P. "Introduction to Data Mining," *Pearson Education*, 2006.
- [4] Budi, S., Handayani, D., & Saefurrohman, "MODEL ANALISIS PRE-DEFINED SINGLE LINKAGE CLUSTERING PADA SISTEM REKOMENDASI OBYEK WISATA DI KOTA SEMARANG," *Jurnal DINAMIKA*, 2019.
- [5] Aditya, W., Pandu, A., & Adinugroho, S., "Clustering Dokumen Skripsi Dengan Menggunakan Hierarchical Agglomerative," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, pp. 6227-6234, 2018.
- [6] Beckstead, J., "Using Hierarchical ClusterAnalysis in Nursing Research," *Western Journal of Nursing Research*, pp. 307-319, 2002.
- [7] Gupitha, R., "PENENTUAN STRATEGI MARKETING SEKOLAH MENENGAH KEJURUAN TERPADU LAMPUNG SUBANG MENGGUNAKAN METODE K-MEANS CLUSTERING," *Jurnal Global*, vol. 2, pp. 17-24, 2018.
- [8] Iqbal, M., Dian, R., & Yusnita, I., "CLUSTERING CALON PENUMPANG KERETA CEPAT JAKARTA - BANDUNG CLUSTERING ON POTENTIAL PASSENGERS OF JAKARTA - BANDUNG HIGH SPEED RAILWAY," *Jurnal Manajemen Transportasi & Logistik*, 2017.
- [9] Kohn, H., & Hubert, L., "Hierarchical Cluster Analysis," *Statistics Reference Online*, pp. 1-13, 2015.
- [10] Maulana, R., Adi, D., Nugraha, N., & Rahmasai, A., "Implementasi Algoritma Hierarchical Clustering Untuk Klasterisasi Data Pelanggan Mall," *Gunung Djati Conference Series*, 2021.
- [11] Mukharil, A., Dhamayanti, D & Lutfi, R., "PENERAPAN METODE HIERARCHICAL AGGLOMERATIVE CLUSTERING UNTUK SEGMENTASI PELANGGAN POTENSIAL DI JEGER JERSEY INDONESIA," *KOMPUTA*, 2017.
- [12] Salsabila, D., & Yunus, H., "Analisis Kondisi Pemberdayaan Gender di Indonesia Tahun 2020 dengan Agglomerative Hierarchical Clustering dan Biplot (Analysis of Indonesia's Gender Empowerment Conditions in 2020 with Agglomerative Hierarchical Clustering and Biplot)," *Seminar Nasional Official Statistics*, pp. 1-10, 2021.
- [13] Tri, R., Wahyusriningsih, S., & Arista, N., "Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu," *Jambura Journal of Mathematics*, 2019.
- [14] Widodo, E., Ermayanti, P., Nur, L., & Tri, A., "Pengelompokan Provinsi di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Analisis Hierarchical Agglomerative Clustering (Indonesian Province Grouping Based on Poverty Level Using Hierarchical Agglomerative Clustering Analysis)," *Seminar Nasional Official Statistics*, 2021.
- [15] Yim, O., & Ramdeen, K., "Hierarchical Cluster Analysis: Comparison of Three Linkage Measures and Application to Psychological Data," 2015.