

Graph Clustering pada Pengelompokan Tujuan Distribusi Barang Berdasarkan Matriks Adjacency

Yunio Heri Kristianto¹, Daniel Swanjaya, M.Kom²

Teknik Informatika, Fakultas Teknik, Universitas Nusantara PGRI Kediri

E-mail: yuniohk@gmail.com, daniel@unpkediri.ac.id

Abstrak – Dalam melakukan proses distribusi atau menyalurkan barang dari pabrik ke konsumen, distributor dituntut untuk tepat waktu sehingga prosesnya tidak boleh terdapat suatu masalah. Dalam hal ini pengelompokan tujuan distribusi barang adalah salah satu permasalahan yang seringkali muncul dikarenakan hanya dilakukan secara manual atau secara intuisi manusia saja sehingga hasil yang diperoleh kurang maksimal. Untuk mengatasi permasalahan tersebut dilakukan penelitian untuk mengelompokkan tujuan distribusi barang dengan graph clustering berdasarkan jarak terpendek antar tujuan dengan menggunakan algoritma *Hypergraph-Partitioning* dan *Agglomerative Hierarchical Clustering (AHC)*. Hasil yang didapatkan penulis dari penelitian ini adalah kedua metode mampu dalam mengelompokkan tujuan secara optimal, namun dilihat dari hasil nilai ASW didapatkan bahwa *Agglomerative Hierarchical Clustering (AHC)* lebih baik daripada algoritma *Hypergraph-Partitioning* dalam hal graph clustering pengelompokan tujuan distribusi barang berdasarkan jarak terpendek antar tujuan.

Kata Kunci —Pengelompokan Data, *Hypergraph-Partitioning*, Algoritma Hierarchical Clustering

1. PENDAHULUAN

Distributor merupakan perantara yang menyalurkan produk maupun barang dari pabrik pembuat ke konsumen atau pengecer. Dalam proses penyaluran atau pendistribusian barang tersebut, distributor dituntut untuk tepat waktu sehingga dalam pendistribusian barang tidak boleh terdapat masalah. Pengelompokan tujuan distribusi barang adalah salah satu permasalahan yang seringkali didapati dikarenakan hanya dilakukan secara manual dan intuisi manusia saja, sehingga hasil yang di dapat kerap kali tidak mencapai hasil yang memuaskan bahkan hasil yang di dapat serengkali kurang optimal. Tujuan pengiriman yang benar sangat dibutuhkan agar hasil lintasan yang diperoleh menjadi optimal sehingga tidak akan mengganggu proses jalannya proses pendistribusian barang dari lokasi satu menuju lokasi lainnya.

Dalam menghadapi permasalahan ini distributor dituntut untuk mengambil suatu keputusan atau jalan keluar yang tepat agar permasalahan dapat diatasi. Salah satu jalan keluarnya adalah memanfaatkan suatu teknologi atau sistem yang mampu membantu mereka dalam mengambil suatu keputusan yang tepat dan akurat. Saat ini sudah ada banyak system yang mampu membantu mengatasi permasalahan dalam pengelompokan tujuan distribusi barang atau system yang mampu mengelompokkan data tujuan yang dimiliki distributor.

Pada penelitian sebelumnya oleh Eka Wahyuni telah berhasil menghasilkan suatu sistem yang dapat membantu menentukan rute pada pendistribusian paket pos dengan mengelompokkan tujuan pengiriman menjadi beberapa wilayah dengan efektif dan efisien dengan menggunakan *Single Linkage* untuk mengelompokkan data, serta menentukan rute distribusi menggunakan Algoritma *Dijkstra* dan *Branch & Bound* [1].

Pada penelitian Lisna Zahrotun telah mampu mengelompokkan jumlah penumpang bus Trans Jogja berdasarkan jam, halte dan jalur bus dengan algoritma K-Means dan *Agglomerative Hierarchical Clustering (AHC)* [2].

Dari permasalahan dan penelitian di atas, untuk mengatasi permasalahan pengelompokan tujuan distribusi barang maka penulis menggunakan algoritma *Hypergraph-Partitioning* dan *Agglomerative Hierarchical Clustering (AHC)*.

2. METODE PENELITIAN

Prosedure yang digunakan pada penelitian ini adalah teknik/ metode *waterfall*. Alasan menggunakan metode ini dikarenakan metode *waterfall* melakukan pendekatan secara sistematis dan berurutan sehingga sistem yang akan dihasilkan akan mencapai hasil yang maksimum. Tahapan metode *waterfall* seperti gambar 1.

1. Identifikasi Masalah

Pada tahapan ini adalah mengidentifikasi masalah pada penelitian ini dan menentukan solusi untuk permasalahan tersebut yang sesuai dan menghasilkan hasil yang optimal

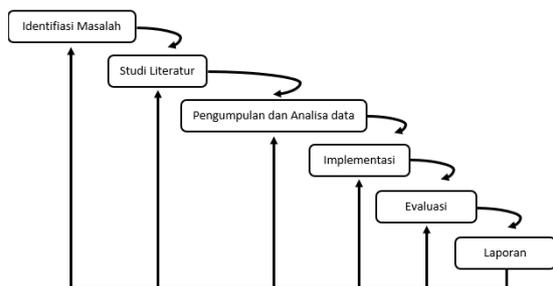
2. Studi Literatur

Studi literatur dilakukan dengan cara mengumpulkan literasi baik berupa jurnal, artikel maupun buku yang terkait dengan tujuan penulis dapat mempelajari dan memahami bahasan tentang Pengelompokan Data, algoritma *Hypergraph-Partitioning* dan *Agglomerative Hierarchical Clustering (AHC)*.

3. Pengumpulan dan Analisa Data

Pada tahapan ini dilakukan pengumpulan data yang diambil dari Google Maps dalam kurun

waktu tertentu, yaitu dari tanggal 6 Desember 2019 – 20 Desember 2019. Kemudian juga dilakukan Analisa terhadap data tersebut.



Gambar 1. Tahapan alur metode Waterfall

4. Implementasi

Pada tahapan implementasi dilakukan dengan cara mengeksekusi data yang sudah dibuat menggunakan algoritma *Hypergraph-Partitioning* dan *Agglomerative Hierarchical Clustering (AHC)* dengan tahapan-tahapan yang ada pada kedua algoritma.

5. Evaluasi

Pada tahapan ini dilakukan evaluasi setelah dilakukan implementasi, apakah ada kesalahan (*error*) yang ada dalam proses penelitian dan mengambil suatu saran yang akan berguna untuk penelitian yang akan datang.

6. Penyusunan Laporan

Tahapan ini dilakukan dengan tujuan setiap kegiatan yang dilakukan dapat dilaporkan dan dapat dilihat dengan mencantumkan hasil dari *Graf Clustering* menggunakan kedua algoritma

2.1 Teori Graf

Graf (*Graph*) adalah pasangan terurut $(V(G), E(G))$ yang terdiri dari himpunan $V(G)$ dari simpul dan himpunan $E(G)$, terpisah dari $V(G)$, tepi, bersama-sama dengan kejadian fungsi ψ_G yang terkait dengan setiap tepi G pasangan tidak teratur dari (tidak harus berbeda) dari simpul G . Dinamakan demikian karena dapat direpresentasikan secara grafis, dan representasi grafis inilah yang membantu kita memahami banyak dari sifatnya. [3]

2.2 Clustering (Pengelompokan Data)

Clustering merupakan suatu proses dalam pengelompokan data menjadi beberapa cluster sehingga pada cluster tersebut memiliki tingkat kesamaan atau kemiripan baik maksimum maupun minimum [4].

Clustering adalah seperangkat metode untuk menyelidiki data dengan tujuan menilai hubungan setiap sampel dalam kumpulan data. Analisis cluster adalah nama yang diberikan untuk seperangkat teknik dengan tujuan menentukan karakteristik structural dari suatu set data dengan membagi menjadi cluster. Jadi clustering adalah teknik populer yang dasar

tujuannya adalah untuk menemukan pengelompokan sampel dalam data [5].

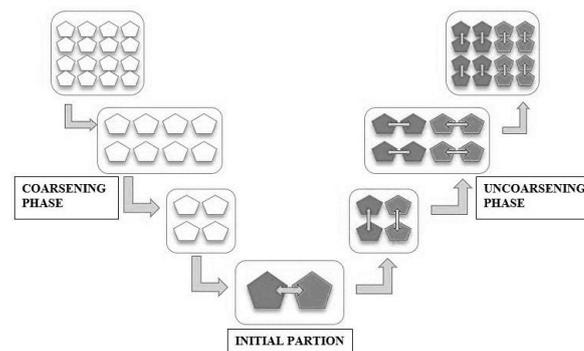
Clustering merupakan alat yang penting dalam data mining. Lebih luas lagi, clustering dapat di definisikan sebagai proses pembagian sekelompok objek ke dalam cluster-cluster [6].

2.3 Hypergraph-Partitioning

Hypergraph adalah suatu bentuk *graph* yang dimana tepi (*edge*) dapat menghubungkan dua atau lebih *vertices* yang disebut juga *hyperedges*. Sedangkan *Hypergraph-Partitioning* adalah suatu proses untuk membagi - bagi *hypergraph* ke dalam sub-sub *hypergraph* atau *graph*.

Proses pembagian $graph G = \{V, E\}$, dimana node atau $V = \{v_1, v_2, \dots, v_n\}$, *edge* yang menghubungkan antar node atau $E = \{e_1=\{v_1, v_2\}, e_2=\{v_1, v_3\}, \dots, e_n\}$, maka himpunan partisinya atau $P = \{S_1, S_2, S_3, \dots, S_n\}$ dimana S merupakan himpunan yang berisi node yang berkelompok pada S , misal $S_1 = \{v_1, v_3, v_5, v_6\}$.

Gambar 2 merupakan tahapan dari *Hypergraph-Partitioning* yang terdiri dari 3 fase, yaitu *Coarsening Phase* yang bertujuan menyederhakan suatu *graph* agar menjadi bentuk partisi awal untuk proses *balancing*, *Initial Partition/Balancing* yang bertujuan menyeimbangkan dua partisi *graph* berdasarkan *edge* yang menghubungkan partisi dari *graph* tak berbobot dan memaksimalkan antar jarak keduanya pada *graph* berbobot, serta *Uncoarsening Phase* yang bertujuan untuk mengoptimalkan kelompok dari hasil proses *coarsening* agar memiliki keseimbangan dengan melakukan proses *balancing* pada setiap pasangan partisi.



Gambar 2. Algoritma Hypergraph-Partitioning [7]

2.4 Agglomerative Hierarchical Clustering (AHC)

Hierarchical clustering adalah metode pengelompokan data yang paling banyak digunakan. Titik awal *hierarchical clustering* adalah kesamaan matriks yang dibentuk oleh komputasi pertama jarak antara semua pasangan titik dalam kumpulan data. Setiap jarak kemudian dikonversi menjadi nilai kesamaan

Pada persamaan 1, S_{ik} (bervariasi dari 0 hingga 1) adalah kesamaan antara sampel I dan k , d_{ik} adalah jarak Euclidean antar sampel I dan k , dan d_{max}

adalah jarak antara dua sampel paling berbeda (misal jarak terbesar) dalam kumpulan data.

Agglomerative Hierarchical clustering (AHC) menempatkan setiap objek dalam kelompoknya sendiri dan secara bertahap menggabungkan kelompok atom ini menjadi kelompok yang lebih besar dan lebih besar sampai semua proyek berada dalam satu kelompok [8].

Langkah-langkah untuk Agglomerative Hierarchical clustering adalah :

1. Mulai langkah n cluster, masing-masing hanya berisi satu objek.
2. Cari matriks ketidaksamaan D untuk pasangan yang paling mirip. Biarkan pasangan yang dipilih menjadi terkait dengan elemen drs sehingga objek r dan s dipilih.
3. Gabungkan objek r dan s ke dalam cluster baru (rs) yang menggunakan beberapa kriteria dan kurangi jumlah cluster dengan 1 dengan menghapus baris dan kolom untuk objek r dan s. hitung perbedaan antara cluster (rs) dan semua cluster yang tersisa, gunakan kriteria, dan tambahkan baris dan kolo ke matriks ketidaksamaan yang baru.
4. Ulangi langkah 2 dan 3, (n-1) kali hingga semua objek membentuk satu cluster. Setiap langkah, mengidentifikasi cluster yang digabung dan nilai ketidaksamaan di mana cluster digabung.

$$S_{ik} = 1 - \frac{d_{ik}}{d_{max}} \dots\dots\dots (1)$$

$$d_{(R)(S)} = \min\{d_{rs} | r \in R \text{ and } s \in S\} \dots\dots\dots (2)$$

$$I(a) = \frac{1}{|\pi_i|} \sum_{x \in \pi_i} d(x, a) \dots\dots\dots (3)$$

$$O_{j(a)} = \frac{1}{|\pi_j|} \sum_{x \in \pi_j} d(x, a) \dots\dots\dots (4)$$

$$s(a) = \frac{O(a)-I(a)}{\max\{O(a), I(a)\}} \dots\dots\dots (5)$$

2.5 Single Linkage

Menerapkan single link method adalah dengan menggabungkan objek dalam kelompok menggunakan minimum perbedaan antar cluster [9]. Membiarkan r mewakili elemen apapun di cluster R, $r \in R$, dan s elemen apapun dalam cluster S, $s \in S$, dari agglomerative cluster, jarak antara R dan S dihitung menggunakan persamaan 2.

2.6 Silhouette Coefficient

Silhouette Coefficient digunakan untuk mengukur suatu kualitas pada pengelompokan data sekaligus mengetahui derajat kepemilikan setiap objek yang berada pada sebuah klaster.

Tahapan Metode Silhouette Coefficient [10] adalah :

1. Hitung jarak rata-rata dari objek a ke objek lain di cluster yang sama menggunakan persamaan 3.
2. Hitung jarak rata-rata dari objek a ke objek lain dalam cluster yang berbeda menggunakan persamaan 4.
3. Hitung nilai silhouette coefficient menggunakan persamaan 5

2.7 Dataset

Data pada penelitian ini berupa data koordinat dari 40 titik swalayan atau toko yang diambil dari google maps dengan kurun waktu tertentu, yaitu dari tanggal 6 Desember 2019 – 20 Desember 2019.

Tabel 1. 20 Data dari 40 Toko/ Swalayan

Id	Nama	Lg	Lt
01	Borobudur	112,013013	-7,817443
02	Arta Swalayan Hayam Wuruk	112,017168	-7,815169
03	Swalayan Surya Dlopo	112,035380	-7,801741
04	Mekar Swalayan PT. Gudang Garam	112,025671	-7,797366
05	Toko Purwodadi Rembang	112,001861	-7,881877
06	Hypermart Kediri	112,015748	-7,812048
07	Golden Swalayan	112,018970	-7,815251
08	Swalayan Surya Kwadungan	112,027449	-7,791210
09	Toko Mekar Sari Kediri	112,010108	-7,829326
10	Sumber Rejeki Jl. Dandang Gendis	112,041007	-7,817101
11	Alfamidi Veteran	111,994139	-7,811099
12	Surya Pagu Swalayan	112,087679	-7,808148
13	RISKITAMA SWALAYAN	111,995903	-7,899378
14	Arta Swalayan Mojoroto	112,002982	-7,803267
15	Namimart Kediri	112,019244	-7,831557
16	Banaran Swalayan	112,036410	-7,831258
17	Trimithra Mart Kediri	111,989780	-7,827930
18	TOP Swalayan Kediri	112,197869	-7,774400
19	Tsamaniya Bendo	112,155683	-7,765044
20	Toko Mina	111,967855	-7,746799

Tabel 2. Data jarak antar toko dalam satuan Kilo Meter (contoh dari 5 Lokasi)

Id	00	01	02	03	04	05
00	~	4,2	3,5	1,8	0,55	12,3
01	4,2	~	1,5	4,4	4,2	7,6
02	4,5	1,5	~	4,3	4	8,3
03	1,8	4,4	4,3	~	1,3	11,7
04	0,55	4,2	4	1,3	~	11,2
05	12,3	7,6	8,3	11,7	11,5	~

Tabel 1 merupakan contoh dari data sampel (contoh dari 20 toko) yang dibutuhkan pada penelitian ini dengan keterangan :

1. Id : sebagai identitas/ pengenalan dari nama toko atau swalayan
2. Nama : nama dari distributor dan beberapa toko
3. Longitude : koordinat atau titik yang digunakan untuk menentukan lokasi yang sering juga disebut garis bujur
4. Latitude : koordinat atau titik yang digunakan untuk menentukan lokasi yang sering juga disebut garis lintang

Kemudian dari keseluruhan data toko yang ada, diolah menjadi data jarak antar lokasi toko dengan jarak a ke b (jarak berangkat) dan jarak b ke a (jarak pulang) disamakan dengan jarak berangkat. Tabel 2 merupakan contoh dari data jarak antar lokasi berdasarkan jarak terdekat.

3. HASIL DAN PEMBAHASAN

Dari 40 data sampel yang telah dibuat, kemudian dilakukan pengelompokan data menggunakan *Algoritma Hypergraph-Partitioning* dan *Agglomerative Hierarchical Clustering* dengan 4 contoh pengelompokan data, yaitu : 18 tujuan (N = 18), 20 tujuan (N = 20), 22 tujuan (N = 22) dan 24 tujuan (N = 24) serta klaster dibagi menjadi 3, 4 dan 5 klaster (tanpa menggunakan data ke 00/ distributor).

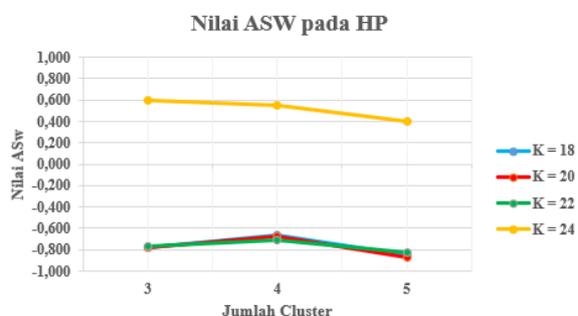
3.1 Pengelompokan data dengan *Hypergraph-Partitioning*

Pada Pengujian pertama data dikelompokkan menggunakan *Hypergraph-Partitioning*, Nilai *Average Silhouette Width* (ASW) dari percobaan banyak cluster dan banyak data didapatkan hasil seperti tabel 3. Nilai ASW negatif menunjukkan banyak data yang salah cluster. Dari Gambar 3 terlihat hasil terbaik didapat pada percobaan dengan banyak data 24 dan banyak cluster 3, dengan nilai ASW 0,596.

Akan tetapi dari hasil pengelompokan menggunakan *Hypergraph-Partitioning* didapat banyak data tiap cluster merata.

Tabel 3. Nilai ASW Pengujian pertama

Jumlah Cluster	Average Silhouette Width			
	N = 18	N = 20	N = 22	N = 24
3	-0,783	-0,775	-0,772	0,596
4	-0,660	-0,675	-0,712	0,555
5	-0,853	-0,868	-0,828	0,397



Gambar 3. Grafik nilai ASW pada *Hypergraph-Partitioning*

3.2 Hasil Pengelompokan data dengan *Agglomerative Hierarchical Clustering* (AHC)

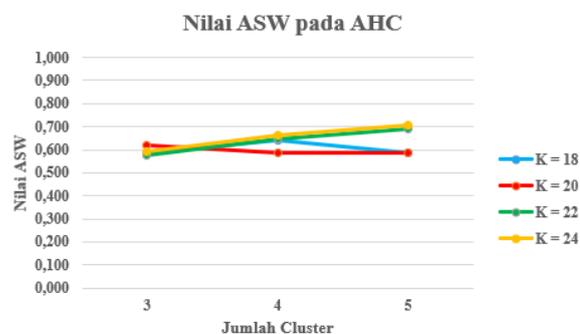
Pada Pengujian kedua data dikelompokkan menggunakan *Agglomerative Hierarchical Clustering* (AHC), Nilai *Average Silhouette Width*

(ASW) dari percobaan banyak cluster dan banyak data didapatkan hasil seperti tabel 4. Nilai ASW yang semakin besar menunjukkan kualitas pengelompokan yang lebih baik. Dari Gambar 4 terlihat hasil terbaik didapat pada percobaan dengan banyak data 24 dan banyak cluster 5, dengan nilai ASW 0,708.

Akan tetapi dari hasil pengelompokan menggunakan *Agglomerative Hierarchical Clustering* (AHC) didapat banyak data tiap cluster tidak merata.

Tabel 4. Nilai Average Silhouette Width pada *Agglomerative Hierarchical Clustering*

Jumlah Cluster	Average Silhouette Width			
	N = 18	N = 20	N = 22	N = 24
3	0,584	0,619	0,573	0,593
4	0,642	0,584	0,645	0,664
5	0,585	0,584	0,691	0,708



Gambar 4. Grafik nilai ASW pada *Agglomerative Hierarchical Clustering*

Dari kedua pengujian pengelompokan data menggunakan metode graph clustering, yang telah dilakukan metode *Agglomerative Hierarchical Clustering* (AHC), memberikan nilai ASW lebih tinggi daripada metode *Hypergraph-Partitioning*, hal ini disebabkan oleh proses pengelompokan yang ada, dimana metode *Hypergraph-Partitioning* pada satu tahap semua data yang berdekatan dikelompokkan, sementara metode AHC satu tahap hanya mengelompokkan sepasang data yang terdekat, sehingga hasil metode AHC yang didapat memiliki nilai ASW yang tinggi. Akan tetapi banyak anggota cluster yang didapat menggunakan metode AHC tidak merata bahkan terdapat cluster yang hanya berisi 1 data saja, sedangkan *Hypergraph-Partitioning* bisa merata dengan baik. Dari segi kecepatan *Hypergraph-Partitioning* memiliki kompleksitas yang lebih sedikit daripada AHC.

4. SIMPULAN

Kesimpulan yang dapat diambil dari penelitian ini adalah sebagai berikut :

1. *Hypergraph-Partitioning* Mampu membantu penulis melakukan pengelompokan data berdasarkan jarak terdekat dengan baik, seimbang dan merata

2. Pengelompokkan data tujuan distribusi barang menggunakan *Agglomerative Hierarchical Clustering* (AHC) juga mendapatkan hasil yang cukup baik serta memiliki hasil nilai ASW yang baik.
3. Dalam penelitian ini graf clustering yang cocok untuk mengelompokkan tujuan distribusi barang adalah *Agglomerative Hierarchical Clustering* (AHC) berdasarkan nilai silhouette yang lebih baik daripada *Hypergraph-Partitioning*

5. SARAN

Berdasarkan hasil dari penelitian ini, saran yang dapat diberikan oleh penulis agar diperoleh hasil yang lebih baik adalah :

1. Memperbaiki data yang ada dikarenakan data pada penelitian ini bersifat sama dari jarak berangkat dan pulang atau nilai a ke $b = b$ ke a
2. Dapat ditambahkan pengujian yang lebih kompleks dengan data yang berbeda yaitu data jarak berdasarkan waktu tercepat

DAFTAR PUSTAKA

- [1] Wahyuni, Eka. 2015. *IMPLEMENTASI HIERARCHIAL CLUSTERING DAN BRANCH AND BOUND PADA SIMULASI PENDISTRIBUSIAN PAKET POS*. *Journal of UN PGRI Kediri*.
- [2] Zahrotun, Lisna. 2015. ANALISIS PENGELOMPOKAN JUMLAH PENUMPANG BUS TRANS JOGJA MENGGUNAKAN METODE CLUSTERING K-MEANS DAN AGGLOMERATIVE HIERARCHICAL CLUSTERING (AHC). *JURNAL INFORMATIKA Vol. 9, No. 1, Jan 2015*
- [3] Bondy, Adrian. & Murty, U.S.R. 2008. *Graph Theory*. New York : Springer
- [4] Irwansyah, Edy., dan Faisal, Muhammad, 2019. *Advanced Clustering:: Teori dan Aplikasi*. Yogyakarta : CV. Budi Utama.
- [5] Lavine, Barry K. 2000. *Clustering and Classification of Analytical Data*. Potsdam
- [6] Basu, Sugato, Davidson, Ian and Wagsta Kiri L. 2009. *Constrained Clustering : Advances in Algorithms, Theory, and Applications*. New York : CRC Press
- [7] Papa, David A., and Igor L. Markov. "Hypergraph Partitioning and clustering." *Approximation algorithms and metaheuristics* (2007): 61-1.
- [8] Jain, Anil K & Dubes, Richard C. 1988. *Algorithms for Clustering Data*. New Jersey: Prentice Hall
- [9] Han, Jiawei & Kamber Micheline. 2006. *Data Mining: Concepts and Techniques (2nd ed)*. San Fransisco: Morgan Kaufman Publishers
- [10] Cogan, Jacob. 2006. *Introducion to Clustering Large and High-Dimensional Data*. New York :. Cambridge University Press

[Halaman ini Sengaja Dikosongkan]