

S5

by Prada Prada

Submission date: 14-Nov-2023 01:36PM (UTC+0700)

Submission ID: 2227680909

File name: Paper.doc (277K)

Word count: 2487

Character count: 16133

OPTIMALISASI SEGMENTASI PELANGGAN MENGGUNAKAN HIERARCHICAL CLUSTERING

Musthofa Galih Pradana^{1*}, Rifka Dwi Amalia², Kharisma Wati Gusti³

10

^{1,3}Program Studi Informatika, Universitas Pembangunan Nasional Veteran Jakarta

²Program Studi Sistem Informasi, Universitas Pembangunan Nasional Veteran Jakarta

Cilandak, Jakarta Selatan

musthofagalihpradana@upnvj.ac.id¹, rifkadwiamalia@upnvj.ac.id²,

kharismawati@upnvj.ac.id³

Abstract – The position and role of customers in business continuity for a company is very vital, this means that companies must be able to carry out the right policy and decision processes in optimizing all the resources they have to be able to formulate good strategies and policies, including for companies providing shopping center services such as mall. The basis for a good policy or decision can basically be based on a fact or real data that represents the real situation. The process of formulating and determining decisions can also have direct implications for the sustainability and running of a company. One identification that companies can do is by segmenting customers. The segmentation process can be used with various kinds of references and also scientific domains, one of which is in terms of data mining or the process of extracting data into something more valuable. One of the approach techniques that can be applied for segmentation strategies is the clustering process, this is suitable because the unsupervised data grouping process can create and extract new information from a set of data. One algorithm that can be applied is hierarchical clustering, this algorithm connects rows or samples with the same configuration to form a tree structure. The results of this research are that the process of implementing the hierarchical clustering algorithm can provide a new point of view from existing data and has the opportunity to be used as a company segmentation strategy. The clusters formed from this segmentation were 5 clusters with the validation results for the use of the elbow method having the same results.

Keywords - Segmentation, Customer, Clustering, Hierarchical Clustering

Abstrak – Posisi dan peranan pelanggan dalam keberlangsungan usaha bagi sebuah perusahaan sangatlah vital, hal ini menjadikan perusahaan harus bisa melakukan proses kebijakan dan keputusan yang tepat dalam mengoptimalkan segala sumber yang dimiliki untuk dapat merumuskan strategi dan kebijakan yang baik, tak terkecuali bagi perusahaan penyedia layanan pusat perbelanjaan seperti mall. Landasan kebijakan atau keputusan yang baik, pada dasarnya dapat didasarkan pada sebuah fakta ataupun data nyata yang merepresentasikan keadaan sesungguhnya. Proses perumusan dan penentuan keputusan juga dapat berimplikasi secara langsung terhadap keberlangsungan dan jalannya sebuah perusahaan. Salah satu identifikasi yang bisa dilakukan oleh perusahaan diantaranya adalah dengan melakukan segmentasi pelanggan. Proses segmentasi dapat digunakan dengan berbagai macam acuan dan juga ranah keilmuan, salah satunya dalam hal data mining atau proses ekstrasi data menjadi sesuatu yang lebih bernilai. Teknik pendekatan yang dapat diterapkan untuk strategi segmentasi salah satunya dengan proses klustering, hal ini cocok dikarenakan proses pengelompokan data yang secara tidak terbimbing dapat menjadikan dan mengekstrak informasi baru dari sekumpulan data. Salah satu algoritma yang dapat diterapkan adalah hierarchical clustering, algoritma ini menghubungkan baris atau sampel dengan konfigurasi yang sama untuk membentuk struktur pohon. Adapun hasil dari penelitian ini adalah proses penerapan algoritma hierarchical clustering dapat memberikan sudut pandang baru dari data yang ada dan berpeluang dijadikan strategi segmentasi perusahaan. Cluster yang terbentuk dari segmentasi ini sebanyak 5 cluster dengan hasil validasi terhadap penggunaan metode elbow memiliki hasil yang sama.

Kata Kunci - Segmentasi, Pelanggan, Klustering, Pengelompokan Hierarki

I. PENDAHULUAN

Salah satu faktor penting dalam indikator keberhasilan atau kesuksesan sebuah usaha adalah keberadaan pelanggan. Hal ini selaras dengan fakta-fakta yang menyebutkan bahwa posisi pelanggan memiliki pengaruh dalam perkembangan suatu bisnis. Pelanggan merupakan salah satu aspek dasar dari keberhasilan serangkaian strategi bisnis yang diterapkan perusahaan [1]. Pelanggan merupakan salah

satu faktor yang dapat digunakan sebagai parameter ukur dari keberhasilan dan kualitas layanan dari perusahaan [2]. Contoh nyata dalam kasus ini adalah pertumbuhan pelanggan yang terus naik dari tahun ke tahun perusahaan Netflix juga diiringi oleh pertumbuhan profit perusahaan [3] [4].

Berdasarkan beberapa hal tersebut, maka posisi dan peran pelanggan sangat penting dalam suatu bisnis. Salah satu hal konkret yang dapat dilakukan dalam mempertahankan dan meningkatkan kepuasan

dari pelanggan adalah dengan menerapkan strategi bisnis, seperti segmentasi pelanggan. Segmentasi dapat dibagi berdasarkan kebutuhan dan juga tujuan dari strategi [5]. Penerapan strategi segmentasi dapat dilakukan dengan beberapa teknik dan pendekatan teknologi yang bermacam-macam, salah satunya dengan pendekatan data mining yang memiliki banyak cara ataupun teknik di dalamnya untuk mempermudah pengelolaan informasi dan juga data yang erat kaitannya dengan Customer Relationship Management [6].

Pengelolaan data pengguna hamper pasti dilakukan oleh sembilan perusahaan, tak terkecuali bagi pusat perbelanjaan. Pusat perbelanjaan saat ini menjadi salah satu tempat yang menjadi tempat favorit warga Indonesia menghabiskan waktu bersama keluarga, hal ini selaras dengan data dari KataData yang menyatakan rata-rata nasional pengunjung mal atau pusat perbelanjaan nyaris mencapai 100% pada awal 2023 dengan durasi kunjungan hampir mencapai 2 jam [7]. Banyaknya kunjungan ke mall juga mengakibatkan data pengunjung mall juga menjadi banyak dan bervariasi, sehingga hal ini memungkinkan untuk dilakukan pengelolaan dari data, salah satunya dengan melakukan segmentasi.

Proses segmentasi dapat diterapkan dengan menggunakan bantuan teknik data mining salah satunya dengan menerapkan klustering. Klustering adalah proses pembagian populasi yang memiliki karakter dan sifat yang serupa ke beberapa kelompok kecil untuk dikelompokkan [8]. Hal ini sangat cocok untuk proses segmentasi pelanggan. Beberapa penelitian terdahulu dari penerapan klustering untuk mengukur kepuasan pelanggan dihasilkan dari penelitian ini menyatakan dari total 1303 data, yang masuk kategori kluster 0 atau sebesar 65% atau dinyatakan puas dan 697 data masuk kategori kluster 1 atau sebesar 35% atau tidak puas [9]. Penelitian perbandingan dua algoritma yakni K-Means dan Logistic Regression dalam kasus data Universitas menyatakan bahwa penerapan clustering model K-Means memiliki akurasi sebesar 22% yang lebih rendah dibandingkan dengan model Logistic Regression model sebesar 91% [10]. Penerapan klustering hierarchical dengan kombinasi FL+HC memungkinkan persentase klien yang lebih besar untuk mencapai akurasi target dibandingkan dengan FL standar [11]. Dalam penelitian identifikasi metode identifikasi ulang orang (re-ID) hasilnya menunjukkan bahwa hierarchical clustering mencapai 56,4% mAP di Market-1501 dan 50,7% mAP di DukeMTMC-reID [12].

Pengelompokan perilaku pengguna kartu pintar penting dalam bidang analisis permintaan angkutan umum dengan menerapkan CCD dan DTW pada model hierarchical clustering yang menunjukkan bahwa kinerja CCD lebih baik daripada DTW dalam mengklasifikasikan rangkaian waktu [13]. Penelitian yang melakukan pengelompokan data DNA yang

terkena dampak kanker menunjukkan hasil perbandingan dari hierarchical clustering, K-Means dan K-Medoids membuktikan bahwa algoritma hierarchical clustering menunjukkan waktu eksekusi dan efisiensi yang lebih rendah peningkatan akurasi dibandingkan yang lainnya [14].

Klustering

Klustering dikenal sebagai metode mengidentifikasi kelompok umum dalam kumpulan data. Entitas di setiap grup lebih mirip dengan entitas di grup tersebut dibandingkan dengan entitas di grup lain. Sejak tahun 1970-an, segmentasi cluster sangat sering digunakan dalam berbagai penelitian terkait data, khususnya di bidang pemasaran [15].

Hierarchical Clustering

Hierarchical Clustering atau pengelompokan hierarki adalah metode yang umum digunakan untuk mengidentifikasi kelompok data yang berkaitan erat. Pengelompokan hierarki adalah metode yang berhasil menghubungkan baris atau sampel dengan konfigurasi yang sama untuk membentuk struktur pohon. Perbedaan dengan algoritma klustering lain seperti K-Means yang cenderung merupakan metode clustering non-hierarki (datar) yang mengharuskan analis memberikan nomor cluster (K) dan mengalokasikan baris dan sampel ke dalam cluster dengan tepat [16].

Adapun langkah dalam hierarchical clustering adalah sebagai berikut : Diberikan set suatu himpunan yang terdiri dari N item untuk di clusterkan, dan NxN jarak/similarity matriks, proses dasar dari hierarchical clustering adalah : [16]

- Menempatkan setiap item ke dalam cluster masing-masing. Sehingga jika terdapat N item, juga terdapat N cluster yang terdiri dari 1 item.
- Cari pasangan cluster terdekat/termirip dan gabungkan ke dalam 1 cluster.
- Hitung jarak/similarity antara cluster baru dan setiap cluster yang lama.
- Ulangi langkah 2 dan 3 hingga setiap item diclusterkan ke dalam single cluster yang memiliki ukuran N

3

II. METODE PENELITIAN

Penelitian ini menggunakan alur dan metode dengan detail sebagai berikut :

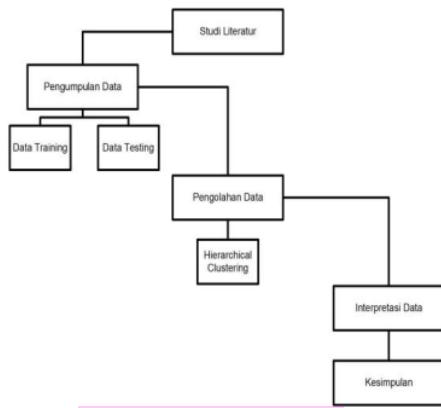
1. Studi Literatur

Tahapan ini melakukan studi pustaka yang mencari dari penelitian relevan terdahulu sebelum melakukan tahapan-tahapan penelitian berikutnya.

2. Pengumpulan Data

Tahapan pengumpulan data menggunakan dataset pengguna mall yang dalam

- pengumpulannya menggunakan kuisioner online.
3. Pengolahan Data
Proses olah data diambil dari data yang sudah terkumpul untuk dilakukan proses pengelompokan hierarki yang bertujuan untuk mendapatkan hasil dari pemodelan pengelompokan hierarki.
 4. Interpretasi Data
Hasil pengolahan data selanjutnya dilakukan proses pemahaman dan pembacaan data dari proses pengelompokan yang sudah dilakukan untuk mendapatkan informasi baru dari proses pengolahan data.
 5. Kesimpulan
Penarikan kesimpulan dari penelitian ini dicermati dari hasil akhir yang didapatkan berdasarkan data latih dan data uji.
Adapun gambaran dari metode penelitian ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

17 III. HASIL DAN PEMBAHASAN

Data penelitian yang digunakan adalah data responden pengunjung mall di Indonesia dengan melakukan pengisian kebutuhan data yang diperlukan untuk penelitian ini, adapun jenis dan deskripsi data pada penelitian ini ditunjukkan pada Tabel 1.

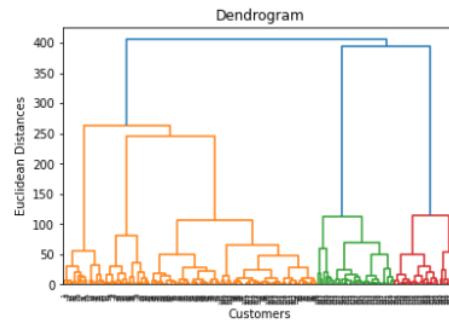
Tabel 1. Atribut Data

No	Jenis Kelamin	Usia	Total Pendapatan	Skor Pengeluaran
1	Laki-Laki	19	Rp. 2.300.000	67
2	Laki-Laki	21	Rp. 2.600.000	70
3	Perempuan	20	Rp. 2.900.000	64
4	Perempuan	53	Rp.	74

5	Perempuan	35	4.700.000	87
			
100	Laki-Laki	48	Rp. 9.300.000	80

Dari 1000 data yang terkumpul, dibutuhkan data pendapatan para responden dan juga angka pengeluaran yang sudah diberikan rating skor. Semakin tinggi rating skor pengeluaran maka pengeluaran responden tersebut juga semakin banyak. Adapun rentang dari rating skor dimulai dari 0 sampai dengan 100.

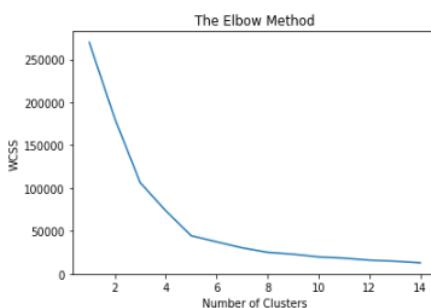
Tahapan berikutnya adalah dengan memproses data yang sudah terkumpul dalam proses pengelompokan hierarki. Proses ini akan membuat data yang ada membentuk kluster sesuai dengan data yang dimiliki. Disinilah perbedaan dari proses klasifikasi dan klustering, dimana kelas labelnya dan jumlah kelompoknya belum diketahui atau disebut *unsupervised learning*. Pada tahapan ini proses pengelompokan hierarki atau *hierarchical clustering* menunjukkan hasil sebagaimana pada Gambar 2.



Gambar 2. Dendrogram Hasil Pengelompokan

Hasil dari dendrogram diatas dapat dilakukan interpretasi hasil. Panjang garis vertikal pada dendrogram menunjukkan seberapa mirip data/objek satu dengan data/objek lainnya. Semakin pendek artinya semakin dekat kemiripannya, semakin panjang berarti semakin jauh kemiripannya (tidak mirip). Jika banyak cluster dari dendrogram dapat dilakukan dengan mencari garis vertikal terpanjang di mana dia tidak berpotongan dengan garis horizontal. Jika ditarik garis horizontal pada jarak cluster yang terpanjang (garis vertikal terpanjang) dan dilihat perpotongannya dengan garis vertikal, maka diperoleh bahwa jumlah cluster yang terbentuk adalah 5 cluster atau 5 kelompok.

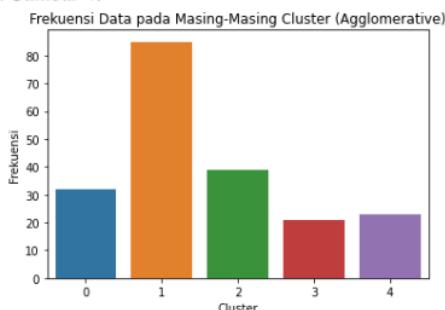
Untuk mengecek optimalisasi hasil dari jumlah cluster yang terbentuk dalam hierarchical clustering, dilakukan proses pengecekan jumlah kluster optimum dengan menggunakan elbow. Hasil dari penerapan metode elbow ditunjukkan pada Gambar 3.



Gambar 3. Hasil Elbow

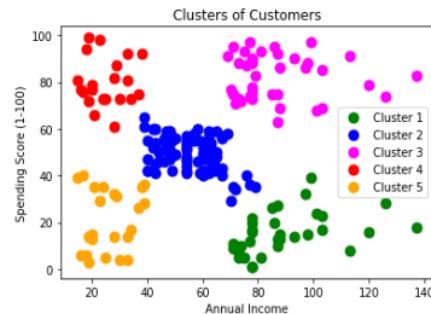
Hasil elbow menunjukkan grafik nilai wcss yang semakin menurun seiring dengan angka jumlah kluster yang semakin banyak, namun setelah K=5 grafik terlihat membentuk seperti sebuah siku. Penurunan nilai wcss setelah K=5 sudah tidak terlalu signifikan. Dengan demikian, dapat diketahui bahwa jumlah kluster yang tepat untuk grafik di atas adalah 5. Hasil ini berarti selaras dengan jumlah kluster yang terbentuk pada diagram dendogram.

Setelah proses validasi hasil kluster dari dendogram dengan hasil elbow, berikutnya dapat dilakukan visualisasi dari masing-masing cluster yang terbentuk. Adapun kelima kluster yang terbentuk masing-masing memiliki frekuensi yang ditunjukkan pada Gambar 4.



Gambar 4. Frekuensi Kluster

Berdasarkan frekuensi kluster, maka dapat dibentuk dan divisualisasikan dalam bentuk grafik pada Gambar 5.



Gambar 5. Hasil Cluster dengan Income

Dengan melihat hasil scatter plot, kita dapat mengetahui perbedaan perilaku dari customer pada masing-masing cluster sebagai berikut:

- Cluster 1 memiliki Annual Income TINGGI, Spending Score RENDAH.
- Cluster 2 memiliki Annual Income SEDANG, Spending Score SEDANG
- Cluster 3 memiliki Annual Income TINGGI, Spending Score TINGGI
- Cluster 4 memiliki Annual Income RENDAH, Spending Score TINGGI
- Cluster 5 memiliki Annual Income RENDAH, Spending Score RENDAH

Dari hasil cluster ini, informasi strategi tertentu dapat diterapkan berdasarkan data. Misalnya berikan 1 kluster 1 karena mereka memiliki pendapatan tinggi namun skor pengeluarannya masih rendah. Hal ini untuk memastikan pelanggan di cluster 1 dapat berbelanja/membeli lebih banyak. Pelanggan pada kluster 3 dapat dikatakan pelanggan setia karena selain mempunyai pendapatan yang tinggi juga mempunyai tingkat pembelanjaan/pembelian yang tinggi. Strategi yang dapat diterapkan pada kluster 3 adalah dengan menawarkan reward atau insentif khusus untuk mempertahankannya. Beberapa strategi dimungkinkan dibangun berlandaskan hasil dari proses pengelompokan atau klasterisasi. Sehingga penerapan klustering ini secara langsung dapat bermaafat dan memberikan impact terhadap pengambilan keputusan yang lebih baik karena didasarkan pada data nyata yang dapat dipertanggungjawabkan validitas datanya.

IV. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan dapat ditarik kesimpulan sebagai berikut :

1. Penerapan klustering ini dapat membantu proses pengelompokan data customer yang dapat bermanfaat bagi proses pengambilan keputusan berdasarkan data nyata.
2. Hasil penerapan metode elbow dengan tujuan mendapatkan nilai kluster maksimal memiliki

hasil yang sama dengan hasil kluster yang terbentuk dari dendogram hierarchical clustering.

Adapun saran yang didapat pada penelitian ini adalah pengoptimalan hasil solusi dari setiap cluster yang terbentuk dapat diterapkan solusi yang lebih konkret sesuai dengan kebutuhan dari segmentasi pelanggan mall.

DAFTAR PUSTAKA

- [1] Z. Ritonga, *Buku Ajar Manajemen Strategi (Teori Dan Aplikasi)*. Deepublish, 2020.
- [2] F. Halim *et al.*, *FullBookManajemenPemasaranJasa*. 2021.
- [3] GoodStats, “Menilik Perkembangan Netflix dalam 10 Tahun Terakhir,” 2023. <https://data.goodstats.id/statistic/agneszfanya-yonatan/menilik-perkembangan-netflix-dalam-10-tahun-terakhir-I8MRp>.
- [4] KataData, “Ditinggal 200 Ribu Pelanggan, Pendapatan Netflix Masih Tumbuh.” <https://databoks.katadata.co.id/datapublish/2022/04/29/ditinggal-200-ribu-pelanggan-pendapatan-netflix-masih-tumbuh>.
- [5] A. Rizal, *Manajemen Pemasaran*. Deepublish, 2020.
- [6] R. Shanthi, *Customer Relationship Management*. MJP Publisher, 2019.
- [7] KataData, “Okupansi Mal Nyaris 100%, Durasi Kunjungan Mencapai Dua Jam Lebih,” *KataData*, 2023. <https://katadata.co.id/tiakomaliasari/berita/63f7124ac6cff/okupansi-mal-nyaris-100-durasi-kunjungan-mencapai-dua-jam-lebih>.
- [8] M. G. Pradana and H. T. Ha, “Maximizing Strategy Improvement in Mall Customer Segmentation using K-means Clustering,” *J. Appl. Data Sci.*, vol. 2, no. 1, pp. 19–25, 2021, doi: 10.47738/jads.v2i1.18.
- [9] M. G. Pradana, A. C. Nurcahyo, P. H. Saputro, U. A. Ata, and U. A. Ata, “Penerapan Metode K-Means Klustering untuk Menentukan Kepuasan Pelanggan,” *Citec J.*, vol. 7, no. 1, pp. 42–50, 2020.
- [10] A. Estetikha, D. H. Gutama, and M. G. Pradana, “Comparison of K-Means Clustering & Logistic Regression on University data to differentiate between Public and Private University,” *Int. J. Informatics Inf. Syst.*, vol. 4, no. 1, pp. 21–29, 2021.
- [11] C. Briggs, Z. Fan, and P. Andras, “Federated learning with hierarchical clustering of local updates to improve training on non-IID data,” in *2020 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2020, pp. 1–9, doi: 10.1109/IJCNN48605.2020.9207469.
- [12] K. Zeng, M. Ning, Y. Wang, and Y. Guo, [13] “Hierarchical Clustering With Hard-Batch Triplet Loss for Person Re-Identification,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020, pp. 13654–13662, doi: 10.1109/CVPR42600.2020.01367.
- [14] L. He, B. Agard, and M. Trépanier, “A classification of public transit users with smart card data based on time series distance metrics and a hierarchical clustering method,” *Transp. A Transp. Sci.*, vol. 16, no. 1, pp. 56–75, Dec. 2020, doi: 10.1080/23249935.2018.1479722.
- [15] S. Balakrishna, M. Thirumaran, V. K. Solanki, and E. R. Núñez-valdez, “Incremental Hierarchical Clustering driven Automatic Annotations for Unifying IoT Streaming Data,” *Int. J. Interact. Multimed. Artif. Intell.*, vol. 6, pp. 56–70, 2020, doi: 10.9781/ijimai.2020.03.001.
- [16] M. F. Edy Irwansyah, *Advanced Clustering : Teori dan Aplikasi*. DEEPUBLISH, 2015.
- [17] G. E. Y. Rani Rotul Muhima, Muchamad Kurniawan, Septiyawan Rosetya Wardhana, Anton Yudhana, Sunardi, Weny Mistarika Rahmawati, *Kupas Tuntas Algoritma Clustering : Konsep, Perhitungan Manual,dan Program*. Andi, 2022.

22%

SIMILARITY INDEX

22%

INTERNET SOURCES

6%

PUBLICATIONS

5%

STUDENT PAPERS

PRIMARY SOURCES

1	wandalistathea.medium.com Internet Source	6%
2	Submitted to Universitas Amikom Student Paper	4%
3	jurnal.una.ac.id Internet Source	2%
4	www.researchgate.net Internet Source	2%
5	www.coursehero.com Internet Source	2%
6	www.megabagus.id Internet Source	1%
7	zh.scribd.com Internet Source	1%
8	katadata.co.id Internet Source	1%
9	123dok.com Internet Source	1%

10	jurnal.polgan.ac.id Internet Source	1 %
11	openaccess.thecvf.com Internet Source	<1 %
12	repository.uin-malang.ac.id Internet Source	<1 %
13	digilib.isi.ac.id Internet Source	<1 %
14	e-jurnal.ippmunsera.org Internet Source	<1 %
15	ejurnal.untag-smd.ac.id Internet Source	<1 %
16	doku.pub Internet Source	<1 %
17	id.scribd.com Internet Source	<1 %

Exclude quotes Off

Exclude bibliography On

Exclude matches Off