



## Implementasi Algoritma Hierarchical Clustering untuk Klasterisasi Data Pelanggan Mall

*(Implementation of Hierarchical Clustering Algorithm for Mall Customer Data Clustering)*

**Risnandy Maulana<sup>1</sup>, Dimas Adi Putra Pratama<sup>2</sup>, Nisvy Sya`bana Nugraha<sup>3</sup>, Adellia Rahmasari<sup>4</sup>**

<sup>1</sup>Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050101@student.uinsgd.ac.id

<sup>2</sup>Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050030@student.uinsgd.ac.id

<sup>3</sup>Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050080@student.uinsgd.ac.id

<sup>4</sup>Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050004@student.uinsgd.ac.id

### Abstrak

Di kota-kota tertentu terdapat mall besar yang pasti menyimpan informasi pelanggan yang sering berlangganan di mall tersebut dan mempunyai kartu keanggotaan. Di dalam kartu keanggotaan mereka terdapat beberapa informasi diantaranya : jenis kelamin, usia, skor pengeluaran dan pendapatan tahunan. Pelanggan tersebut menggunakan kartu keanggotaan itu ketika akan melakukan berbagai pembelian di mall, sehingga secara otomatis mall tersebut memiliki riwayat pembelian dari semua pelanggan yang datang. Kami harus menyegmentasikan pelanggan tersebut berdasarkan detail pelanggan yang diberikan. Jadi kita akan membuat cluster pelanggan menggunakan Hierarchical Clustering. Dengan menggunakan clustering dalam penelitian ini sehingga dapat menghasilkan strategi yang sangat tepat untuk sekelompok pelanggan. Berkat teknik pembelajaran tanpa pengawasan ini, perusahaan juga dapat menyesuaikan strategi pemasaran dengan lebih efisien dan fokus kepada pelanggan yang memiliki pendapatan paling tinggi. Pelanggan disegmentasikan menjadi beberapa kelompok yang berbeda sesuai dengan skor pengeluaran dan pendapatan tahunan mereka.

**Kata kunci:** algoritma hierarchical clustering, data mining, klasterisasi

### Abstract

*In certain cities there are large malls that must store information on customers who often come to the mall and have cards that can. In their information card there is some information including: gender, age, expenditure and annual income. The customer uses the information card when he makes various purchases at the mall, so that the mall automatically has the spending spirit of all customers who come. We have to segment these customers based on the customer details provided. So we will create a cluster customer using Hierarchical Clustering. By using the grouping in this study so as to produce a very appropriate strategy for a group of customers. Thanks to this unsupervised learning technique, companies can also adjust their marketing strategies more efficiently and focus on the customers with the highest revenue. Customers are segmented into different groups according to their annual expenses and income.*

**Keywords:** clustering, data mining, hierarchical clustering

## 1 Pendahuluan

Dengan meningkatnya jumlah perusahaan yang ada pada jaman ini, tentu saja membuat persaingan global yang akan terus semakin meningkat termasuk persaingan pasar menjadi sangat sengit. Peningkatan perusahaan tersebut melahirkan pemahaman mengenai pentingnya pelanggan dalam membangun model bisnis dan juga analisis data transaksi pelanggan untuk membantu perencanaan komersial dari sebuah perusahaan. Agar tidak kalah saing dengan perusahaan luar, perusahaan dalam negeri pun mengembangkan pasar sesuai perkembangan yang saat ini terjadi dengan menerapkan strategi bisnis yang dihasilkan dari analisis kebutuhan pelanggan mereka.

Ada salah satu perusahaan dalam negeri yang sudah mengembangkan sistem terbaru mereka untuk keperluan pelanggan dalam proses transaksi. Maka dari itu, selain fokus dengan produk yang mereka jual perusahaan pun harus sangat memperhatikan kebutuhan pelanggannya sendiri. Untuk menciptakan nilai bagi perusahaan dan pelanggan terdapat pemahaman mengenai pelanggan yang terdapat dalam pengelolaan pelanggan yang menggambarkan strategi komprehensif dalam proses memperoleh dan mempertahankan. Perusahaan memiliki tugas penting selain mencari pelanggan, yaitu perusahaan harus mampu mempertahankan pelanggan yang ada karena jika mereka kehilangan pelanggan yang ada maka mereka akan kehilangan seluruh aliran bisnis dengan 2 pelanggan sekaligus. Untuk mempertahankan pelanggan tersebut perusahaan dalam cakupan lingkungan yang dinamis harus mencoba untuk memperoleh wawasan mengenai kebutuhan pelanggan, sikap serta perilaku pelanggan yang akan mereka pertahankan.

Tetapi pada kenyataan yang ada, perusahaan pasti mempunyai pelanggan yang berbeda beda perilaku dalam melakukan proses transaksi. Idealnya setiap pelanggan harus diperhatikan satu persatu oleh perusahaan dan harus diperlakukan sebagai individu, tapi dengan pendekatan yang seperti itu justru sangatlah tidak mungkin terjadi. Maka dari itu, ada alternatif yang lebih efisien yaitu melakukan segmentasi terhadap pelanggan. Segmentasi pelanggan merupakan sebuah model yang dibangun dalam mengelompokkan pelanggan sesuai dengan standar tertentu untuk digunakan menjadi variabel pengelompokan. Pelanggan akan menjadi satu kelompok yang sama apabila memiliki kesamaan tertentu, sementara kelompok atau segmen yang berbeda merupakan pelanggan memiliki karakteristik yang berbeda. Setelah dilakukannya analisis terhadap perilaku pelanggan melalui segmentasi tersebut, maka dapat dilakukan perencanaan untuk proses pemasaran di masa yang akan datang dalam pengiriman pesan yang paling tepat dari setiap segmen pelanggan.

Data mining adalah salah satu Teknik menemukan *insight knowledge* dari data yang besar [1], [2], yang dapat digunakan untuk menganalisis segmentasi pelanggan untuk target pasar dan pelayanan yang baik sesuai kondisi pelanggan. Salah satu pendekatan data mining adalah klasterisasi dengan menggunakan algoritma hierarchial clustering. Terdapat beberapa penelitian terdahulu yang terkait dengan penelitian ini, antara lain: (1) menentukan sumber pasokan kopi secara optimal dengan menggunakan algoritma hierarchial clustering [3]; (2) implementasi algoritma hierarchial clustering untuk integrasi teks otomatis [4], [5]; dan pengelompokkan komunitas media sosial dengan algoritma agglomerative hierarchial clustering [6].

## 2 Metodologi

Segmentasi atau Clustering adalah metode untuk mengidentifikasi kelompok alami dari sebuah kasus yang didasarkan pada sebuah kelompok atribut. Segmentasi mengelompokkan data yang memiliki kemiripan atribut. Clustering bersifat unsupervised, karena tidak ada satu atribut pun yang digunakan untuk memandu proses pembelajaran, jadi seluruh atribut input diperlakukan sama. Sebagian besar algoritma clustering akan membangun sebuah model melalui serangkaian iterasi dan berhenti ketika

model tersebut telah memusat atau berkumpul (batasan dari segmentasi telah stabil. Sedangkan pelanggan didefinisikan sebagai sebuah individu atau kelompok yang melakukan pembelian atas sebuah produk atau jasa berdasarkan keputusan akan pertimbangan harga dan penawaran yang berkomunikasi dengan perusahaan melalui surat, panggilan telepon, dan email yang dikirimkan secara berkala. Proses segmentasi pelanggan akan membagi pelanggan menjadi beberapa kelompok yang disesuaikan dengan tingkat keloyalitasannya.

### **Algoritma Hierarchical Clustering**

Pengelompokan, yang juga disebut sebagai analisis kluster, adalah penting subjek dalam penambangan data. Kumpulan data dipartisi menjadi beberapa grup dan titik data di setiap grup, yaitu cluster lebih terkait satu sama lain dibandingkan dengan yang ada di cluster lain. Ini titik data dikumpulkan bersama dengan mendeteksi korespondensi sesuai dengan atribut yang ditemukan dalam data mentah, namun Tujuan utamanya adalah untuk menemukan jumlah cluster yang sesuai relevan serta berwawasan untuk tujuan analisis. Proses ini tugas berulang dan berulang di mana sejumlah besar data mentah berada memindai persamaan dan pola. Data yang tidak terorganisir adalah mencari pengetahuan yang penting dan kemudian titik datanya ditugaskan. Untuk mendapatkan hasil yang menguntungkan, pengelompokan tertentu algoritma bersama dengan parameter tertentu mungkin paling cocok di a domain pasar. clustering adalah jenis teknik data mining digunakan dalam sejumlah aplikasi, yang melibatkan area seperti mesin pembelajaran, klasifikasi dan pengenalan pola. Ada bermacam-macam algoritma pengelompokan yang berbeda satu sama lain dalam hal pendekatan yang mereka ikuti untuk melakukan pengelompokan objek sesuai dengan karakteristiknya

Clustering dengan pendekatan hirarki akan mengelompokan data yang mirip dalam hirarki yang sama dan yang tidak mirip di hirarki yang agak jauh [7], [8]. Terdapat dua metode yang sering digunakan yaitu Agglomerative Hierarchical Clustering dan Divisive Hierarchical clustering. Agglomerative melakukan clustering dari N cluster menjadi kesatuan cluster, dimana N adalah jumlah data, sebaliknya divisive melakukan proses clustering dari satu cluster menjadi N cluster. Beberapa metode hierarchical clustering yang sering digunakan dibedakan menurut cara perhitungan tingkat kemiripan. Ada yang menggunakan Centroid, Single Linkage dan Complete Linkage. Seperti juga halnya dengan partition-based clustering, jarak dapat digunakan untuk menghitung tingkat kemiripan antar data. Hierarchical Clustering dapat digambarkan melalui dendogram. Dendogram disusun dengan membuat similarity matrix yang memuat tingkat kemiripan antar data yang dikelompokan. Tingkat kemiripan bisa dihitung dengan berbagai macam cara seperti dengan Euclidean Distance Space dan Manhattan Distance.

## **3 Hasil dan Pembahasan**

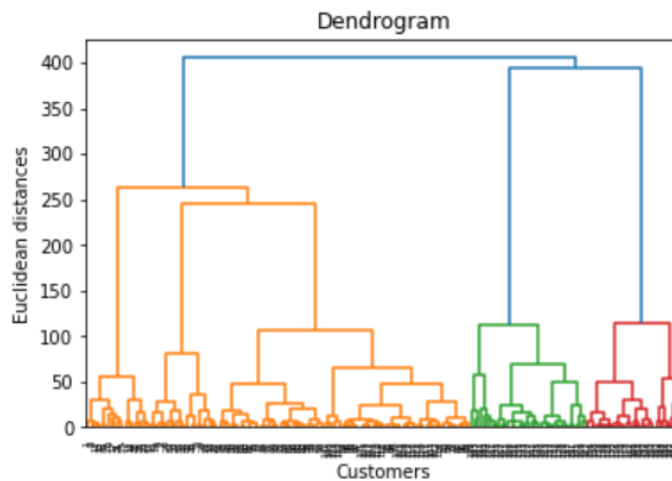
Penelitian ini dilaksanakan melalui beberapa tahapan antara lain:

### **1. Data pre-processing**

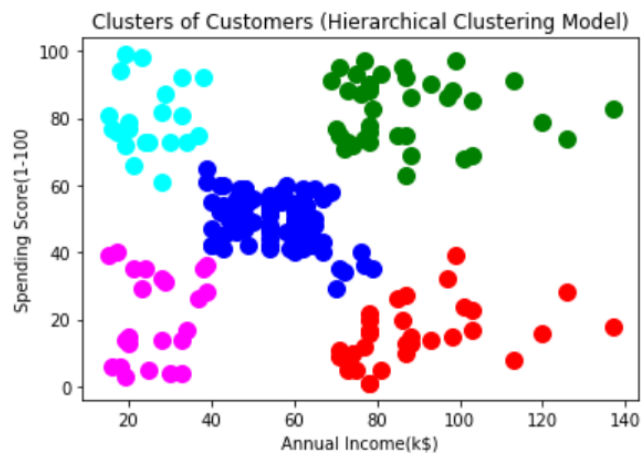
Di bagian ini kami melakukan sedikit eksplorasi data, memeriksa nilai null, tipe data objek, dan hal lain yang mungkin kami pertimbangkan untuk menjaga data kami tetap bersih dan terstruktur dengan baik. Seperti yang telah kita amati, nama beberapa kolom cukup kompleks dan dapat diubah menjadi nama yang lebih sederhana sehingga kita dapat mengakses data dengan lebih mudah. Sebagai bagian data pra-pemrosesan, kami hanya akan mempertimbangkan fitur-fitur berikut : Annual Income (k\$) dan Spending Score (1-100).

### **2. Metode Ward**

Metode ward merupakan salah satu linkage, yang merupakan pengembangan dari centroid yaitu, menggabungkan 2 klaster kecil menjadi 1 dengan memperhitungkan total jarak dari centroid. Gambar 1 menunjukkan contoh dendrogram dimana sumbu x terdiri dari pelanggan dan sumbu y terdiri dari jarak *Euclidean* antar klaster.



Gambar 1 Dendrogram hasil klasterisasi



Gambar 2 Visualisasi hasil klaster

Gambar 2 memvisualisasikan setiap cluster yang terlihat menempati range yang menentukan segmentasi customer yang layak. Segmentasi pelanggan adalah cara yang baik untuk memahami perilaku pelanggan yang berbeda dan merencanakan strategi pemasaran yang sesuai. Dimana 5 klaster tersebut terdiri dari:

1. Merah adalah segmentasi pelanggan yang lebih berhati-hati dalam berbelanja dengan pendapatan tinggi namun pengeluarannya sedikit.
2. Biru tua adalah segmentasi pelanggan dengan pengeluaran dan pemasukan yang normal.
3. Hijau adalah kluster segmentasi yang menjadi target pasar dengan pendapatan yang tinggi dan pengeluaran belanja yang tinggi pula.

4. Biru muda adalah segmentasi pelanggan dengan pendapatan yang sedikit namun pengeluaran belanja yang tinggi
5. Ungu merupakan segmentasi pelanggan dengan pendapatan dan pengeluaran belanja yang rendah

#### 4 Simpulan

Setelah mengembangkan solusi untuk masalah ini, kami sampai pada kesimpulan bahwa Hierarchical Clustering adalah teknik yang dapat digunakan untuk menganalisis segmentasi pelanggan. Segmentasi pelanggan adalah cara yang baik untuk memahami perilaku pelanggan yang berbeda dan merencanakan strategi pemasaran yang sesuai. Hasil penelitian ini menunjukkan adanya lima segmentasi pelanggan yang tentunya memiliki kondisi berbeda, sehingga perlakuannya pun harus berbeda untuk memenuhi target pasar dan pelayanan yang baik.

#### Referensi

- [1] H. Jiawei, M. Kamber, J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2006.
- [2] P.-N. Tan, M. Steinbach, V. Kumar, T. Pang-Ning, M. Steinbach, and V. Kumar, *Introduction to data mining: Instructor's*. 2006.
- [3] A. Saputra, "OPTIMALISASI PENENTUAN SUMBER PASOKAN KOPI ARABIKA GAYO MELALUI PENDEKATAN HIRARCHICAL CLUSTERING DATA MINING," *J. Optim.*, vol. 3, no. 4, 2018, doi: 10.35308/jopt.v3i4.214.
- [4] G. S. Budhi, A. I. Rahardjo, and H. Taufik, "Hierarchical Clustering untuk Aplikasi Automated Text Integration," in *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, 2008, vol. 1, no. 1.
- [5] G. A. Pradnyana, "Perancangan dan Implementasi Automated Document Integration dengan Menggunakan Algoritma Complete Linkage Agglomerative Hierarchical Clustering," *J. Ilmu Komput.*, vol. 5, no. 2, 2012.
- [6] I. W. Rahayu, I. Atastina, and A. Herdiani, "Analisis Dan Implementasi Algoritma Agglomerative Hierarchical Clustering Untuk Deteksi Komunitas Pada Media Sosial Facebook," *eProceedings Eng.*, vol. 5, no. 1, 2018.
- [7] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: an overview," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 2, no. 1, pp. 86–97, 2012.
- [8] S. C. Johnson, "Hierarchical clustering schemes," *Psychometrika*, vol. 32, no. 3, pp. 241–254, 1967.