

Perbandingan Kinerja Algoritma K-means dan Agglomerative Clustering Untuk Segmentasi Penjualan Online Pada Customer Retail

Ghanim Ramadhan¹⁾, Yuli Astuti^{2*)}

¹Prodi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta, Yogyakarta

²Prodi Manajemen Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta, Yogyakarta

^{1,2}Jl. Ringroad Utara, Condongcatur, Depok, Sleman, 55283, Yogyakarta, Indonesia

email: ganimramadhan01@students.amikom.ac.id, yuli@amikom.ac.id

Abstract – This research focuses on the comparison between two popular algorithms in data science, namely K-means and Agglomerative Clustering algorithms. The main context of this research is customer data segmentation, a very important process in the business world to understand and serve customers better. The main objective of this research is to evaluate and compare the performance of the two algorithms in generating effective and efficient customer segments. In this research, the dataset used is a retail customer dataset. This dataset includes various attributes that reflect customer characteristics and behavior. To measure the performance of both algorithms, this research uses the RFM (Recency, Frequency, Monetary) weighting method. This method is a commonly used method in customer analysis to identify the most valuable customers based on how recently they transacted (Recency), how often they transact (Frequency), and how much their transactions are worth (Monetary). In addition, this research also uses an evaluation metric known as silhouette score. This metric is used to measure how well an object fits into its own cluster compared to other clusters. The results of this study provide valuable insights into the quality of both algorithms in segmenting customer data. It was found that the K-Means algorithm produced a silhouette score value of 0.5087, while Agglomerative Clustering produced a higher value of 0.6363. This suggests that, in the context of this dataset, Agglomerative Clustering may be more effective compared to K-Means. However, further research is certainly needed to validate these findings and to further explore how these two algorithms can be optimized for customer data segmentation.

Abstrak – Penelitian ini berfokus pada perbandingan antara dua algoritma populer dalam bidang data science, yaitu algoritma K-means dan Agglomerative Clustering. Konteks utama dari penelitian ini adalah segmentasi data pelanggan, suatu proses yang sangat penting dalam dunia bisnis untuk memahami dan melayani pelanggan dengan lebih baik. Tujuan utama dari penelitian ini adalah untuk mengevaluasi dan membandingkan kinerja kedua algoritma tersebut dalam menghasilkan segmen pelanggan yang efektif dan efisien. Dalam penelitian ini, dataset yang digunakan adalah dataset pelanggan retail. Dataset ini mencakup berbagai atribut yang mencerminkan karakteristik dan perilaku pelanggan. Untuk mengukur kinerja kedua algoritma, penelitian ini menggunakan metode pembobotan RFM (Recency, Frequency, Monetary). Metode ini adalah metode yang umum digunakan dalam analisis pelanggan untuk mengidentifikasi pelanggan yang paling berharga berdasarkan seberapa baru mereka bertransaksi (Recency), seberapa sering mereka bertransaksi (Frequency), dan seberapa besar nilai

transaksi mereka (Monetary). Selain itu, penelitian ini juga menggunakan metrik evaluasi yang dikenal sebagai silhouette score. Metrik ini digunakan untuk mengukur seberapa baik suatu objek cocok dengan klusternya sendiri dibandingkan dengan kluster lain. Hasil penelitian ini memberikan wawasan yang berharga tentang kualitas kedua algoritma dalam melakukan segmentasi data pelanggan. Ditemukan bahwa algoritma K-Means menghasilkan nilai silhouette score sebesar 0.5087, sedangkan Agglomerative Clustering menghasilkan nilai yang lebih tinggi, yaitu 0.6363. Ini menunjukkan bahwa, dalam konteks dataset ini, Agglomerative Clustering mungkin lebih efektif dibandingkan dengan K-Means. Namun, penelitian lebih lanjut tentunya diperlukan untuk memvalidasi temuan ini dan untuk mengeksplorasi lebih lanjut tentang bagaimana kedua algoritma ini dapat dioptimalkan untuk segmentasi data pelanggan.

Kata Kunci – Analisis algoritma, K-means, Agglomerative Clustering, RFM, Silhouette Score.

I. PENDAHULUAN

Seiring dengan lonjakan perusahaan retail yang bermunculan, upaya untuk memberikan pelayanan terbaik menjadi semakin krusial untuk bersaing dengan pesaing lainnya. Namun, upaya ini telah mengalami tantangan karena kurangnya strategi pemasaran yang spesifik dan fokus yang tepat. Karena itulah, diperlukan suatu sistem segmentasi pelanggan yang inovatif untuk mengidentifikasi jenis-jenis pelanggan yang berbeda. Dengan memahami karakteristik unik dari setiap kelompok pelanggan, akan memungkinkan merumuskan strategi pemasaran yang sesuai dengan kebutuhan dan preferensi masing-masing kelompok. Proses segmentasi data berperan penting dalam mengidentifikasi kelompok-kelompok alami dari pelanggan berdasarkan atribut yang relevan. Segmentasi pelanggan digunakan untuk mengelompokkan data pelanggan berdasarkan atributnya. Sebagian besar algoritma clustering atau pengelompokan data membuat model dari beberapa rangkaian iterasi dan berhenti jika model tersebut berhasil mengelompokkan dan mengumpulkan data. Segmentasi data konsumen bertujuan untuk mendukung perusahaan ritel dalam menentukan strategi pemasaran yang cocok dengan profil konsumen mereka.[1].

Algoritma clustering adalah salah satu metode yang digunakan untuk melakukan segmentasi pelanggan. Clustering merupakan salah satu teknik dalam analisis data yang digunakan untuk mengelompokkan data berdasarkan kesamaan atau perbedaan karakteristik tertentu [2]. Dua

*) penulis korespondensi: Ghanim Ramadhan
Email: ganimramadhan01@students.amikom.ac.id

algoritma clustering yang umum digunakan adalah k-means dan agglomerative clustering. Kedua algoritma ini memiliki kelebihan dan kekurangan masing-masing, sehingga pemilihan algoritma clustering yang tepat sangat penting untuk mendapatkan hasil clustering yang akurat. Algoritma k-means merupakan algoritma clustering yang paling populer dan sering digunakan dalam analisis data. Algoritma ini bekerja dengan cara mengelompokkan data ke dalam k kelompok yang sudah ditentukan sebelumnya. Setiap data akan diatributkan ke kelompok yang memiliki jarak terdekat dengan pusat kelompok atau centroid. Algoritma k-means cukup cepat dan efisien untuk data dengan jumlah yang relatif kecil dan memiliki distribusi data yang berbentuk lingkaran atau bola [3]. Sementara itu, Agglomerative clustering adalah algoritma clustering berbasis hierarki. Algoritma ini memulai dengan setiap data sebagai cluster terpisah dan menggabungkan dua cluster yang paling dekat satu sama lain dalam setiap iterasi. Proses penggabungan ini berlanjut hingga semua data tergabung dalam satu cluster atau hingga batas yang ditentukan [4].

II. PENELITIAN YANG TERKAIT

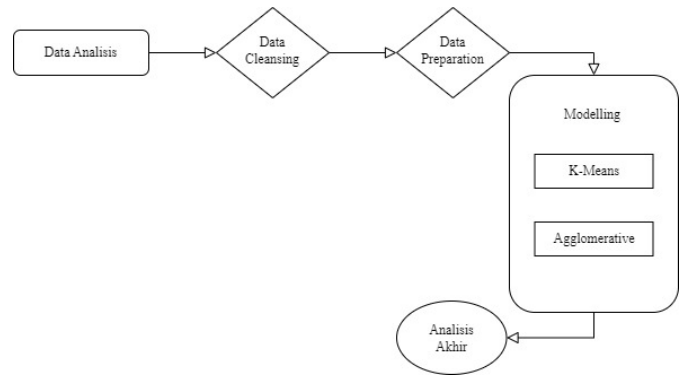
Sudah menjadi hal yang umum untuk melakukan clustering dan melabeli pelanggan dengan dua atau lebih label yang berbeda. Tujuan dari penelitian ini adalah untuk menemukan model pembelajaran mesin yang dapat menggeneralisasi dengan baik intuisi segmentasi data perusahaan.

Ozan, S.[5] melakukan proses segmentasi pelanggan dengan menggunakan tiga metode, yaitu metode normal equation (NEM), metode regresi linear multivariat (LiRM), dan metode regresi logistik (LoRM). Hasil penelitian menunjukkan bahwa metode LoRM memberikan hasil segmentasi yang baik. Aggarwal dan Yagav[6] menerapkan pembobotan Fuzzy-AHP dan mendapatkan perbedaan sebesar 5% untuk setiap kluster. Pramono, Isti dan Enrico Laoh[7] menggunakan pendekatan Customer Lifetime Value (CLV) dalam metode Two-Stage clustering. Muchardie, Gunawan, dan Aditya[8] melakukan segmentasi berdasarkan tingkat kepuasan dan retensi pelanggan. Torizuka, Oi, Saitoh, dan Ishizu[9] menggunakan algoritma random forest dalam proses segmentasi mereka.

III. METODE PENELITIAN

A. Alur Penelitian

Langkah langkah penelitian untuk melakukan segmentasi pelanggan adalah analisis data, data cleansing, data preparation/preprocessing data, modelling/clustering, dan analisis akhir. Langkah tersebut dapat dilihat pada Gbr 1.



Gbr. 1 Alur Penelitian.

B. Data Analysis

Data yang digunakan pada penelitian ini adalah data pelanggan perusahaan retail dan diperoleh dari instansi penyedia dataset yaitu Kaggle. Tahap data analisis ini peneliti membaca dan memahami isi dataset yang berisi 541.909 data terbagi menjadi 8 kolom, seperti yang ditunjukkan pada gambar 2.

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	01-12-2010 08:26	2.55	17850.0	United Kingdom
536365	71003	WHITE METAL LANTERN	6	01-12-2010 08:26	3.39	17850.0	United Kingdom
536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	01-12-2010 08:26	2.75	17850.0	United Kingdom
536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	01-12-2010 08:26	3.39	17850.0	United Kingdom
536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	01-12-2010 08:26	3.39	17850.0	United Kingdom

Gbr. 2 Data Pembelian Pelanggan

Dari gambar 2 diperlihatkan 8 kolom yaitu *InvoiceNo*, *StockCode*, *Description*, *Quantity*, *InvoiceDate*, *UnitPrice*, *CustomerID*, *Country*. Secara berurutan kolom tersebut berisi data kode tanda pembelian, kode stok barang, deskripsi barang atau nama barang, jumlah barang, tanggal pembelian, harga barang, nomor ID pelanggan, dan asal negara pelanggan.

C. Data Cleansing

Data cleansing, juga dikenal sebagai data cleaning atau data scrubbing, adalah proses mengidentifikasi, memperbaiki, dan menghapus kesalahan, inkonsistensi, dan ketidakakuratan dalam kumpulan data. Dalam konteks penelitian, data cleansing adalah langkah penting untuk memastikan keandalan dan kualitas data yang digunakan dalam analisis. Peneliti melakukan identifikasi dan pemeriksaan pada dataset, dalam penelitian ini ditemukan beberapa data null yang kemudian dibersihkan atau dihapus untuk meningkatkan kualitas data. Diperlihatkan pada tabel I dan gambar 3.

TABEL I JUMLAH DATA NULL

Description	1454
CustomerID	135080

InvoiceNo	0
StockCode	0
Description	1454
Quantity	0
InvoiceDate	0
UnitPrice	0
CustomerID	135080
Country	0

Gbr. 3 Jumlah data null dataset

Tabel I dan Gbr. 3 menunjukkan jumlah data null yang terdapat pada kolom *Description* dan *CustomerID* yang masing-masing berjumlah 1454 dan 135.080 data.

D. Data Preparation

Tahapan selanjutnya setelah dilakukan pembersihan data null adalah mempersiapkan data dengan melakukan pembobotan variabel Recency, Frecuency, Moneterly (RFM) menggunakan software *Google Colab*. Analisis RFM digunakan untuk menganalisis nilai pelanggan.

Analisis ini terdiri dari tiga aspek. Aspek pertama adalah Recency, yang memeriksa berapa lama yang lalu pelanggan membeli produk.

Aspek berikutnya adalah Frecuency, yaitu seberapa sering pelanggan melakukan pembelian. Aspek terakhir adalah Moneterly mengambil jumlah total yang dimiliki pelanggan dibelanjakan. Dengan menggunakan nilai yang diambil, skor dapat diberikan untuk pelanggan, satu adalah yang terendah dan lima adalah tertinggi. Segmen pelanggan dapat dibuat dengan membagi nilai ke dalam beberapa kuantil. Setiap paragraf dapat memberikan informasi mulai dari pelanggan yang paling berharga hingga yang paling tidak berharga[10].

E. Modelling

1. K-Means

Merupakan salah satu algoritma clustering yang popular digunakan. Lebih mudah dan lebih cepat dalam perhitungan. Namun, untuk jumlah cluster yang berbeda, itu menunjukkan tes cluster yang berbeda. K2 harus diinisialisasi agar mendapatkan jumlah cluster yang tepat. Angka k pusat juga harus diinisialisasi. Nilai yang berbeda akan menghasilkan cluster yang berbeda dari fokus awal[11]. K-mean clustering dihitung dengan menggunakan persamaan berikut

$$j = \sum_{i=1}^n \sum_{j=1}^k ||x^i - c^j|| \quad (1)$$

2. Agglomerative

Metode ini dilakukan dengan mengelompokkan sekumpulan objek secara bersama-sama (beberapa objek dalam satu kesatuan). Mulailah dengan menempatkan objek apa saja pada cluster dan kemudian gabungkan struktur clustering menjadi clustering yang lebih besar hingga semua objek berada pada cluster yang sama atau memenuhi kondisi yang telah ditentukan. Algoritme ini mengarah pada pendekatan dari bawah ke atas (bottom-up)[12]. Metode hierarki yang digunakan di penelitian ini adalah metode Average Linkage. Average Linkage menghitung jarak antara dua cluster sebagai jarak rata-rata antara pasangan cluster yang digabungkan. Langkah yang digunakan dalam average linkage adalah menghitung jarak antar objek dengan formula jarak Euclidian[13]. Average Linkage memiliki persamaan matematika sebagai berikut

$$d_{u,v}^{k+1} = \frac{\sum_{u \in U} \sum_{v \in V} d_{uv}}{N_u + N_v} \quad (2)$$

F. Analisis Akhir

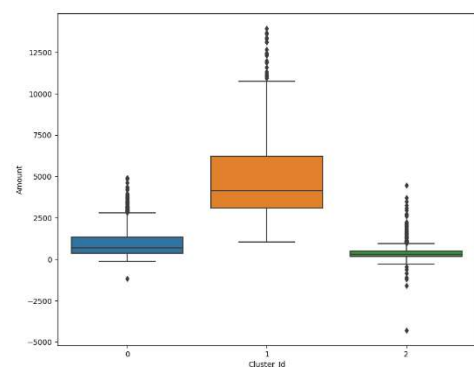
Pada tahap ini peneliti melakukan analisis dari hasil visualisasi cluster serta evaluasi terhadap hasil cluster dan mengambil kesimpulan dari analisis tersebut. Visualisasi pada penelitian ini menggunakan visualisasi boxplot. Boxplot adalah sejenis grafik statistik yang digunakan untuk menampilkan distribusi dari sekumpulan data. Dinamai sesuai dengan bentuknya dan digunakan dalam berbagai bidang, terutama manajemen kualitas, menyediakan metode penyaringan pencilan bagi peneliti[14]. Pada boxplot bagian tengah data ditunjukkan dengan kotak persegi panjang. Kuartil atas atau Q3 dan kuartil bawah atau Q1 dari data ditunjukkan oleh tepi atas persegi panjang dan tepi bawah persegi panjang. Garis median digambar di dalam kotak persegi panjang. 1,5 kali rentang antar kuartil (Q3 - Q1) mendefinisikan langkah. Dari pusat tepi atas kotak hingga nilai maksimum titik data yang berada di dalam langkah, sebuah garis lurus dibuat secara vertikal. Demikian pula, prosedur yang sama diulangi di tepi bawah persegi panjang hingga titik data minimum yang terletak di dalam anak tangga. Masing-masing titik data diplot yang tidak berada dalam rentang langkah. Titik-titik ini menjadi penting dan dapat berupa titik pencilan atau titik noise[15]. Evaluasi yang dipakai pada penelitian ini adalah silhouette score. Fungsi scikit-learn silhouette score adalah menghitung koefisien silhouette rata-rata dari semua sampel. Koefisien silhouette dihitung dengan memperhitungkan rata-rata jarak intra-klaster a dan rata-rata jarak klaster terdekat b untuk setiap titik data[16].

IV. HASIL DAN PEMBAHASAN

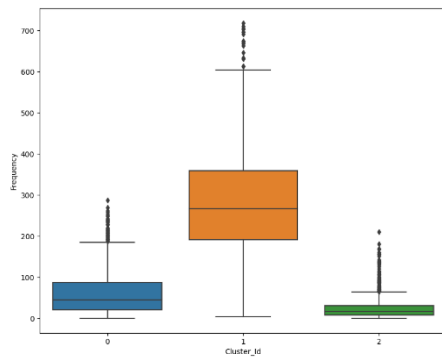
Penelitian ini menggunakan 2 metode algoritma yaitu agglomerative clustering dan K-means yang digunakan untuk melakukan analisis segmentasi terhadap data pelanggan retail sebanyak 541.909 data dan dikelompokkan menjadi 3 cluster data dengan pembobotan variabel RFM.

A. K-means

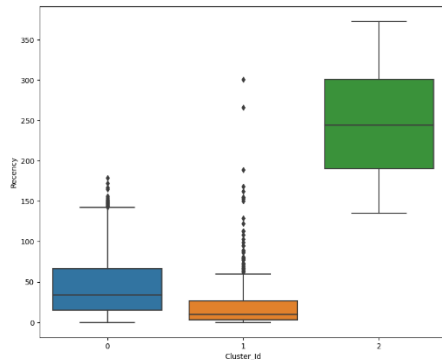
Algoritma K-means pada penelitian ini menggunakan metode elbow point dalam menentukan jumlah cluster, dan cluster yang dihasilkan berjumlah 3 yaitu cluster 0, 1, dan 2 sehingga hasil cluster dapat dilihat pada Gbr 4, Gbr 5 dan Gbr 6.



Gbr. 4 Boxplot cluster dengan variabel amount



Gbr. 5 Boxplot cluster dengan variabel frequency



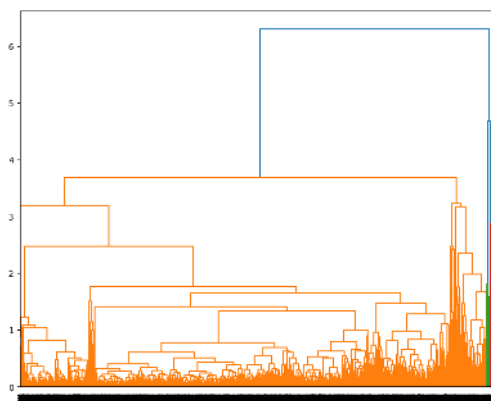
Gbr. 6 Boxplot cluster dengan variabel recency

Dari visualisasi boxplot K-means clustering didapatkan analisa sebagai berikut,

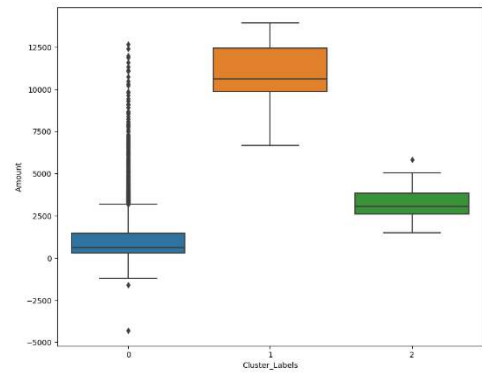
- a. Cluster_Id 1 adalah pelanggan dengan jumlah transaksi yang tinggi dibandingkan dengan pelanggan lainnya.
- b. Pelanggan dengan Cluster_id 1 adalah pembeli yang loyal.
- c. Pelanggan dengan Cluster_id 2 memiliki jumlah transaksi yang kecil serta bukan pembeli terkini oleh karena itu tidak terlalu penting dari sudut pandang bisnis.

B. Agglomerative clustering

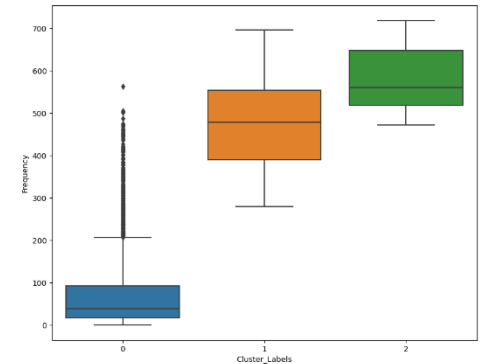
Algoritma agglomerative clustering yang digunakan pada penelitian ini adalah agglomerative clustering average linkage dengan memotong dendrogram. Dendrogram digunakan untuk melihat secara visual jumlah cluster yang terbentuk seperti pada Gbr 7. Hasil cluster dapat dilihat pada Gbr 8, Gbr 9, dan Gbr 10.



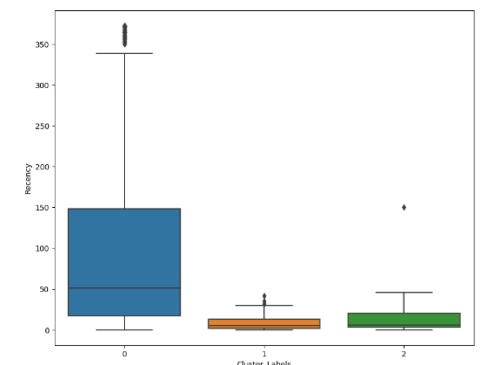
Gbr. 7 Dendrogram cluster



Gbr. 8 Boxplot cluster dengan variabel amount



Gbr. 9 Boxplot cluster dengan variabel frequency



Gbr. 10 Boxplot cluster dengan variabel recency

Peneliti melakukan analisis terhadap boxplot Agglomerative clustering dengan 3 cluster dan mendapatkan hasil sebagai berikut,

- a. Pelanggan dengan cluster_label 1 adalah pelanggan yang memiliki total transaksi tinggi dibanding pelanggan label lain.
- b. Pelanggan dengan cluster_label 1 dan 2 adalah pelanggan yang sering membeli barang.
- c. Pelanggan dengan cluster_label 0 adalah pelanggan yang memiliki jumlah transaksi rendah dan jarang membeli barang maka pelanggan dengan cluster_label 0 tidak menjadi sasaran penting dari sudut pandang bisnis.

Peneliti melakukan evaluasi dengan menggunakan metode silhouette score pada kedua algoritma yang telah berhasil membuat cluster data. Algoritma k-means mendapatkan nilai silhouette score sebesar 0.5087 dan untuk algoritma

agglomerative clustering mendapatkan nilai silhouette score sebesar 0.6363. Semakin besar nilai silhouette score maka kualitas cluster semakin baik.

V. KESIMPULAN

Penelitian perbandingan analisis 2 algoritma clustering untuk melakukan segmentasi data dengan menggunakan pembobotan variabel RFM. Algoritma yang digunakan pada penelitian ini adalah algoritma agglomerative clustering yang berbasis hierarki dengan pendekatan “bottom up” dan algoritma k-means dengan perhitungan yang cepat. Data yang digunakan adalah data pelanggan usaha retail.

Hasil menunjukkan bahwa kedua algoritma berhasil membuat cluster data sebanyak 3 cluster. Dengan menggunakan perhitungan silhouette score, algoritma agglomerative clustering memiliki kualitas cluster yang lebih baik daripada algoritma k-means dijelaskan dengan nilai silhouette score 0.6363 sedangkan algoritma k-means hanya mendapatkan nilai silhouette score 0.5087.

Untuk penelitian selanjutnya, kualitas cluster dapat ditingkatkan lagi dengan melakukan tahap preprocessing data yang lebih bersih dan pembobotan variabel yang lebih kompleks.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada pihak yang membantu ataupun memberikan dukungan terkait dengan penelitian yang dilakukan.

DAFTAR PUSTAKA

- [1] W. Widyawati, W. L. Y. Saptomo, and Y. R. W. Utami, “Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan,” *J. Ilm. SINUS*, vol. 18, no. 1, p. 75, Jan. 2020, doi: 10.30646/sinus.v18i1.448.
- [2] K. B., “A Comparative Study on K-Means Clustering and Agglomerative Hierarchical Clustering,” *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 5, pp. 1600–1604, 2020, doi: 10.30534/ijeter/2020/20852020.
- [3] N. K. Zuhail, “Study Comparison K-Means Clustering dengan Algoritma Hierarchical Clustering,” *Pros. Semin. Nas. Teknol. dan Sains*, vol. 1, pp. 200–205, 2022, [Online]. Available: <https://jurnal.dharmawangsa.ac.id/index.php/djtechno/article/view/966/867>
- [4] R. P. Justitia, N. Hidayat, and E. Santoso, “Implementasi Metode Agglomerative Hierarchical Clustering Pada Segmentasi Pelanggan Barbershop (Studi Kasus : RichDjoe Barbershop Malang),” 2021. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [5] Ş. Ozan, “A Case Study on Customer Segmentation by using Machine Learning Methods,” *2018 Int. Conf. Artif. Intell. Data Process. IDAP 2018*, pp. 0–5, 2019, doi: 10.1109/IDAP.2018.8620892.
- [6] A. G. Aggarwal and S. Yadav, “Customer Segmentation Using Fuzzy-AHP and RFM Model,” *ICRITO 2020 - IEEE 8th Int. Conf. Reliab. Infocom Technol. Optim. (Trends Futur. Dir.)*, pp. 77–80, 2020, doi: 10.1109/ICRITO48877.2020.9197903.
- [7] P. P. Pramono, I. Surjandari, and E. Laoh, “Estimating customer segmentation based on customer lifetime value using two-stage clustering method,” *2019 16th Int. Conf. Serv. Syst. Serv. Manag. ICSSSM 2019*, no. 1994, pp. 1–5, 2019, doi: 10.1109/ICSSSM.2019.8887704.
- [8] B. G. Muchardie, A. Gunawan, and B. Aditya, “E-Commerce Market Segmentation Based on the Antecedents of Customer Satisfaction and Customer Retention,” *Proc. 2019 Int. Conf. Inf. Manag. Technol. ICIMTech 2019*, vol. 1, no. August, pp. 103–108, 2019, doi: 10.1109/ICIMTech.2019.8843792.
- [9] K. Torizuka, H. Oi, F. Saitoh, and S. Ishizu, “Benefit Segmentation of Online Customer Reviews Using Random Forest,” *IEEE Int. Conf. Ind. Eng. Eng. Manag.*, vol. 2019-December, pp. 487–491, 2019, doi: 10.1109/IEEM.2018.8607697.
- [10] Y. Parikh and E. Abdelfattah, “Clustering Algorithms and RFM Analysis Performed on Retail Transactions,” *2020 11th IEEE Annu. Ubiquitous Comput. Electron. Mob. Commun. Conf. UEMCON 2020*, pp. 0506–0511, 2020, doi: 10.1109/UEMCON51285.2020.9298123.
- [11] S. S. Chikkond, R. Salagar, and S. S. Veni, “Segmentation Of Document Images Using Different Methods Like K-Mean Clustering And Fuzzy Clustering,” in *2021 International Conference on Computer Communication and Informatics (ICCCI)*, Jan. 2021, pp. 1–6. doi: 10.1109/ICCCI50826.2021.9402546.
- [12] A. Nugraha, M. Arista Harum Perdana, H. Agus Santoso, J. Zeniarja, A. Luthfiarta, and A. Pertiwi, “Determining the Senior High School Major Using Agglomerative Hierarchical Clustering Algorithm,” *Proc. - 2018 Int. Semin. Appl. Technol. Inf. Commun. Creat. Technol. Hum. Life, iSemantic 2018*, pp. 225–228, 2018, doi: 10.1109/ISEMANTIC.2018.8549834.
- [13] Kristiana, K. R. Sungkono, and R. Sarno, “Determine Types of Indonesian Hospital by Criteria-based Proses Model, K-means Cluster, and Hierarchical Average Linkage,” *Proc. - 2019 Int. Semin. Appl. Technol. Inf. Commun. Ind. 4.0 Retrospect. Prospect. Challenges, iSemantic 2019*, pp. 191–195, 2019, doi: 10.1109/ISEMANTIC.2019.8884299.
- [14] H. Yin, Z. Wang, P. Liu, Z. Zhang, and Y. Li, “Voltage Fault Diagnosis of Power Batteries based on Boxplots and Gini Impurity for Electric Vehicles,” *2019 Electr. Veh. Int. Conf. EV 2019*, pp. 1–5, 2019, doi: 10.1109/EV.2019.8892849.
- [15] K. Patil, N. K. Nagwani, and S. Tripathi, “A Parametric Study of Partitioning and Density Based Clustering Techniques for Boxplot Generation,” *2018 3rd Int. Conf. Conver. Technol. I2CT 2018*, pp. 1–5, 2018, doi: 10.1109/I2CT.2018.8529468.
- [16] K. R. Shahapure and C. Nicholas, “Cluster quality analysis using silhouette score,” *Proc. - 2020 IEEE 7th Int. Conf. Data Sci. Adv. Anal. DSAA 2020*, pp. 747–748, 2020, doi: 10.1109/DSAA49011.2020.00096.