

Implementasi Metode *Agglomerative Hierarchical Clustering* Pada Segmentasi Pelanggan *Barbershop* (Studi Kasus : RichDjoe Barbershop Malang)

Rhayhana Putri Justitia¹, Nurul Hidayat², Edy Santoso³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya

Email: ¹rhayhanapjustitia@gmail.com, ²ntayadih@ub.ac.id, ³edy144@ub.ac.id

Abstrak

Strategi merupakan salah satu dari banyak hal yang berpengaruh dalam suksesnya seseorang menjalankan sebuah usaha. Menerapkan strategi yang benar menjadi hal krusial agar usaha berjalan sesuai dengan yang diinginkan. Salah satu strategi yang dapat diterapkan adalah *Customer Relationship Management* (CRM) yaitu dengan mengatur segmentasi dan cara perlakuan terhadap konsumen. Dalam melakukan analisa para pelaku usaha harus dengan teliti mempertimbangkan setiap konsumen dengan masing-masing kebiasaannya ketika melakukan transaksi untuk dikelompokkan ke dalam segmen-segmen tertentu. Tidak jarang konsumen berada di dalam segmen yang tidak sesuai dengan karakteristiknya karena kesalahan dalam analisis dengan cara yang masih manual. Oleh karena itu, penelitian ini akan menerapkan metode *Agglomerative Hierarchical Clustering* pada pelanggan *Barbershop* dalam melakukan segmentasi. Dalam penelitian ini menggunakan 489 data pelanggan dari RichDjoe Barbershops Malang yang diambil mulai dari bulan Mei hingga Oktober 2020 berupa tanggal kedatangan terbaru setiap pelanggan yang ditandai dengan nomor ID, jumlah kedatangan setiap ID pada kurun waktu tertentu, berapa uang yang dikeluarkan setiap pelanggan, dan akumulasi jarak tanggal terlama hingga terkini setiap nomor ID. Metode *Agglomerative Hierarchical Clustering* diketahui memiliki kemampuan yang baik dalam mengelompokkan data dengan membuat sebuah hirarki secara bertahap dari data yang berupa *singleton*/individu hingga seluruh data terkelompokkan. Metode pengukuran jarak menggunakan *manhattan distance* dengan Parameter jarak yang digunakan yaitu *Single Linkage*, *Complete Linkage* dan *Average Linkage*. Penelitian ini menggunakan *Silhouette Coefficient* dalam melakukan evaluasi pengujian dan menghasilkan hasil tertinggi dari 489 data pada titik potong 391 yaitu dengan nilai rata-rata *silhouette coefficient*-nya 0.968850698 pada parameter *single linkage* dan *average linkage* yang menghasilkan 8 *cluster* dan *silhouette coefficient* setiap datanya menghasilkan 376 dari 399 data yang tepat berada pada kelompoknya serta *silhouette coefficient* setiap *cluster* yang terbentuk mendapatkan hasil 6 dari 8 *cluster* terbentuk dengan tepat.

Kata kunci: *agglomerative hierarchical clustering, segmentasi pelanggan, barbershop*

Abstract

Strategy is one of many thing that influence the success of a person in running a business. Implementing the right strategy is crucial so that the business runs as desired. One of the strategies that can be applied is Customer Relationship Management (CRM), namely by regulating segmentation and how to treat consumers. In conducting analysis, business actors must carefully consider each consumer with their respective habits when conducting transactions to be grouped into certain segments. It is not uncommon for consumers to be in a segment that does not match their characteristics due to errors in the analysis because it is done manually. Therefore, this study will apply the Agglomerative Hierarchical Clustering method to Barbershop customers in segmenting. In this study, using 489 customer data from

RichDJoe Barbershops Malang which were taken from May to October 2020 in the form of the latest arrival date of each customer marked with an ID number, the number of arrivals of each ID during a certain period of time, how much money was spent by each customer, and the accumulation distance from the oldest to the most recent date each ID number. The Agglomerative Hierarchical Clustering method is known to have a good ability in grouping data by making a hierarchy gradually from data in the form of singleton / individual until all data are grouped. The distance measurement method uses Manhattan Distance with the distance parameters used are Single Linkage, Complete Linkage and Average Linkage. This study uses the Silhouette Coefficient in evaluating the test and produces the highest results from 489 data at the 391 cut point, with an average silhouette coefficient value of 0.968850698 on the single linkage parameter and the average linkage resulting in 8 clusters and silhouette coefficient for each data resulting in 376 of 399 the right data are in the group and the silhouette coefficient of each cluster formed shows that 6 out of 8 clusters are formed correctly.

Keywords : *agglomerative hierarchical clustering, customer segmentation, barbershop*

1. PENDAHULUAN

Dalam menjalankan sebuah usaha, pelanggan merupakan elemen yang sangat penting. Agar usaha dapat tetap berjalan, pelaku usaha harus bisa menarik pelanggan dengan terus melakukan segala cara untuk mempertahankan pelanggannya dalam persaingan bisnis salah satunya dengan menerapkan *Customer Relationship Management* (CRM). Menurut Buttle (2004) CRM merupakan salah satu dari beberapa strategi bisnis untuk mendapatkan keuntungan, pendapatan, dan kepuasan dari konsumen dengan upaya mengatur segmentasi dan cara memperlakukan konsumen. Tujuan yang didapat dari penerapan segmentasi pelanggan yaitu untuk memahami perilaku dari konsumen kemudian menerapkan mengimplementasikan strategi pemasaran yang cocok (Nugraheni, 2011). Hal ini perlu diterapkan oleh berbagai jenis usaha, salah satunya *barbershop*.

Salah satu *Barbershop* yang sudah tidak asing lagi di telinga masyarakat Malang, Jember dan Surabaya yaitu RichDJoe Barbershops. Berdiri pada tahun 2015, RichDJoe merupakan tempat pangkas rambut bergaya *modern-classic* menjadi salah satu *barbershop* yang paling terkenal hingga saat ini. Namun, sampai saat ini RichDJoe sendiri dalam menganalisis strategi pemasaran yang khusus dan tepat sasaran masih secara manual dengan mengelompokkan setiap pelanggan dengan masing-masing kebiasaan dalam bertransaksi yang berbeda-beda sehingga harus memberikan *effort* dan ketelitian serta waktu yang banyak. Sehingga berawal dari masalah ini Penerapan metode *Agglomerative Hierarchical Clustering* (AHC) dapat membantu

merealisasikan segmentasi pelanggan pada RichDJoe *Barbershops* secara otomatis. Metode AHC mempunyai kemampuan menggabungkan data dengan membuat hirarki, dimana yang memiliki kemiripan akan ditempatkan di hirarki yang berdekatan dan yang tidak memiliki kemiripan ditempatkan pada hirarki berjauhan (Donny, 2013). Pendekatan yang dilakukan oleh metode AHC berupa *bottom-up* yang berarti proses pengelompokan data akan dimulai dari yang terkecil hingga ke pengelompokan yang besar.

Terdapat penelitian lain mengenai implementasi metode ini yaitu Implementasi Algoritma *Agglomerative Hierarchical Clustering* Untuk Mengelompokkan Capaian Belajar Siswa SD oleh Aloysius Ari Kurniawan pada tahun 2017. Dalam penelitian ini menjelaskan dari 3 metode yang memberikan hasil paling seimbang yaitu metode *average linkage*. Hasil perhitungan *silhouette index* mendapatkan hasil sebanyak seluruh data sebanyak 14 berada tepat di *cluster*, kemudian *cluster* 2 ada sebanyak 10 data yang tepat dari 22 data dan *cluster* 3 sebanyak 4 dari 31 data. Disimpulkan bahwa metode *Agglomerative Hierarchical Clustering* dengan baik mengelompokkan data dari nilai siswa dan menghasilkan *cluster* yang cukup seimbang untuk digunakan sebagai acuan pengelompokan kelas siswa.

Penelitian lain tentang penerapan *Agglomerative Hierarchical Clustering* untuk segmentasi pelanggan oleh Widyawati yang dipublikasi pada Januari 2020. Dijelaskan bahwa dalam penelitian ini metode yang digunakan yaitu *agglomerative hierarchical clustering*

berhasil dalam melakukan pengelompokan pelanggan dan dapat dikaitkan pada strategi pemasaran.

Dari penelitian-penelitian tersebut maka dipilih metode *Agglomerative Hierarchical Clustering* untuk menyelesaikan permasalahan dalam penelitian segmentasi pelanggan ini.

2. TINJAUAN PUSTAKA

2.1 *Barbershop*

Barbershop berasal dari bahasa latin yaitu “barba” yang artinya “janggut”. Dahulu, para pemangkas selalu dihubungkan dengan pembedahan dan klinik gigi karena orang-orang pada masanya tidak hanya datang untuk memangkas rambut akan tetapi meminta pelayanan operasi bedah kecil seperti mencabut gigi atau mengobati luka. Lain halnya dengan masyarakat Indonesia yang lebih mengenal *barbershop* dengan tempat cukur rambut yang biasa ditemukan di bawah pohon sebagai tempat memangkas dan merapikan rambut pria. Terdapat ratusan *barbershop* yang tersebar di seluruh Indonesia khususnya Malang mulai dari yang kecil hingga besar.

2.2 Segmentasi Pelanggan

Segmentasi pelanggan melakukan pengelompokan karakteristik, kebutuhan, dan minat konsumen/pelanggan yang sama ke dalam satu kelompok dengan tujuan untuk menerapkan model bisnis yang mampu memenuhi setiap kebutuhan dari segmen-segmen tersebut. Dilakukannya hal ini agar strategi pemasaran dari perusahaan lebih tepat sasaran dan efisien. Dengan dilakukannya segmentasi, pelaku usaha dapat memahami bahwa pelanggan pun memiliki karakteristik khusus saat berinteraksi dengan produk.

2.3 *Data Mining*

Data mining merupakan sebuah proses untuk menggali informasi tersembunyi pada suatu data dengan menggunakan teknik tertentu. Seperti yang dikatakan oleh Turban, et al. (2005) *data mining* merupakan suatu cara dengan memanfaatkan teknik-teknik dalam kecerdasan buatan untuk melakukan penggalian informasi yang bermanfaat di dalam suatu *database* yang besar. *Data mining* bisa mendapat suatu fakta yang belum ditemukan sebelumnya di dalam suatu data menggunakan metode baru mudah yang mudah dimengerti dan berguna

untuk pengguna data tersebut. Karakteristik dari *data mining* yang paling menonjol adalah dapat menemukan pola data yang rumit dan menemukan informasi yang potensial. Menurut Larose (2005) *data mining* memiliki beberapa proses yang dapat dilakukan yaitu *description*, *estimation*, *prediction*, *classification*, *clustering*, dan *association*.

2.4 *Preprocessing*

Menurut Feldman dan Sanger (2007) *preprocessing* merupakan sebuah tahapan yang dilakukan untuk mengekstraksi atau mengambil sebuah informasi dari suatu data yang tidak terstruktur. Dalam tahap ini data yang masih dalam bentuk mentah akan dipersiapkan untuk dapat diproses pada tahap selanjutnya. Terdapat 4 tahapan dalam *preprocessing* menurut Han dan Kamber (2006) yaitu : *data cleaning*, *data integration*, *data selection*, *data transformation*.

2.5 Pembobotan

Pembobotan merupakan sebuah proses pemberian bobot terhadap data yang memiliki variasi sehingga pada akhirnya memiliki keseragaman. Pada penentuan bobot akan menggunakan prinsip variabel yang memiliki jumlah penyesuaian paling kecil terhadap objek maka akan diberikan bobot paling besar. Begitu juga sebaliknya jika variabelnya memiliki jumlah penyesuaian paling besar maka akan diberikan bobot paling kecil.

2.6 *Agglomerative Hierarchical Clustering*

Agglomerative Hierarchical Clustering merupakan salah satu teknik dalam mengelompokan data. Teknik ini akan mengelompokan data secara berulang-ulang berdasarkan tingkat kemiripan data satu sama lain dan membentuk sebuah hirarki. Pada teknik pengelompokan ini menggunakan pendekatan *bottom-up* dengan mengelompokan data mulai dari data yang berupa *singleton* atau individu kemudian digabung menjadi satu *cluster*/kelompok secara berulang sehingga seluruh data terkelompokan. Saat proses pengelompokan menggunakan teknik ini inisialisasi jumlah *cluster* dan *centroid* tidak diperlukan. Yang diperlukan adalah acuan dengan menggunakan jarak terdekat setiap data. Terdapat beberapa cara menghitung jarak tersebut salah satunya menggunakan metode *manhattan distance*. Rumus dari *manhattan*

distance dapat dilihat pada Persamaan 1.

$$D_{man}(x, y) = \sum_{j=1}^d |x_j - y_j| \quad (1)$$

Setelah hasil dari perhitungan *manhattan distance* setiap data diperoleh selanjutnya data akan masuk proses penghitungan menggunakan parameter jarak yaitu *single linkage*, *complete linkage*, dan *average linkage*.

2.5.1 Single Linkage

Single linkage merupakan sebuah parameter jarak berdasarkan jarak paling kecil tiap data. Algoritma parameter jarak *single linkage* dimulai dengan memilih jarak terkecil dalam matriks data. Rumus dari *single linkage* dapat dilihat pada Persamaan 2.

$$d_{(UV)W} = \min(d_{UW}, d_{VW}) \quad (2)$$

2.5.2 Complete Linkage

Complete linkage merupakan sebuah parameter jarak berdasarkan jarak paling besar tiap data. Algoritma parameter jarak *complete linkage* dimulai dengan memilih jarak terbesar dalam matriks data. Rumus dari *complete linkage* pada Persamaan 3.

$$d_{(UV)W} = \max(d_{UW}, d_{VW}) \quad (3)$$

2.5.3 Average Linkage

Average linkage merupakan sebuah parameter jarak berdasarkan jarak rata-rata tiap data. Algoritma parameter jarak *average linkage* dimulai dengan memilih jarak rata-rata dalam matriks data. Rumus dari *average linkage* pada Persamaan 4.

$$d_{(UV)W} = \frac{d_{(UW)} + d_{(VW)}}{n_{(UV)} n_W} \quad (4)$$

2.7 Silhouette Coefficient (SC)

Silhouette Coefficient (SC) merupakan salah satu *method* dalam melakukan evaluasi yang berguna untuk mengetahui kualitas dan kekuatan suatu *cluster* beserta suatu data dalam *cluster* tersebut (Handoyono, Mangkudjaja, Nasution, 2014). Menghitung nilai SC dapat dilihat pada Persamaan 5.

$$SC = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (5)$$

Nilai *silhouette coefficient* mempunyai *range* nilai antara -1 hingga 1. *Cluster* dapat dikatakan terbentuk dengan baik jika memiliki nilai yang positif dalam arti mendekati atau sama dengan 1. Nilai *silhouette coefficient* positif

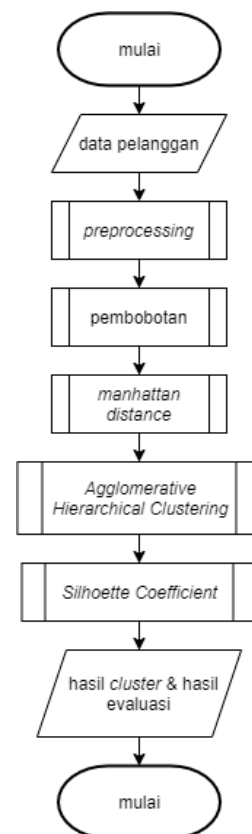
dapat diartikan bahwa data berada di dalam *cluster* yang tepat atau *cluster* terbentuk dengan baik. Sebaliknya jika nilai *silhouette coefficient* mendekati atau sama dengan -1, maka *cluster* yang terbentuk bernilai negatif dalam arti data yang terbentuk tidak berada di *cluster* yang tepat atau *cluster* tidak terbentuk dengan baik.

3 METODOLOGI PENELITIAN

3.1 Data Penelitian

Data yang digunakan merupakan data pelanggan RichDJoe Barbershops Malang dengan jumlah 489 data pada periode bulan Mei sampai dengan bulan Oktober 2020. Data berupa nomor ID, nama serta tanggal setiap kedatangan pelanggan.

3.2 Diagram Alir Sistem



Gambar 1 Diagram Alir Sistem

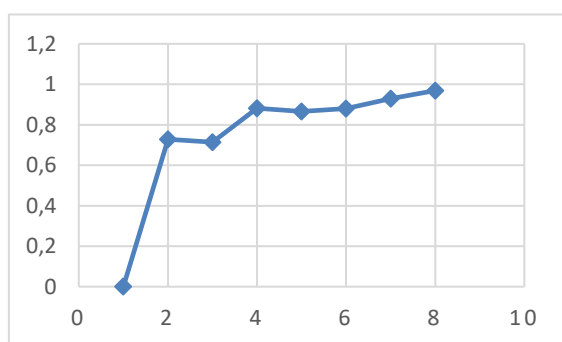
Berdasarkan Gambar 1. Tahap pertama dalam penelitian ini adalah melakukan *input* data berupa nomor ID, tanggal dan nama. Kemudian data akan melalui tahap *preprocessing* untuk mempersiapkan data sebelum dilakukannya proses *clustering*. Setelah tahap *preprocessing* akan dilakukan pembobotan pada setiap data. Tahap selanjutnya *manhattan distance* untuk

mengetahui nilai jarak pada setiap data. Lalu setiap data yang telah melalui proses-proses tersebut akan dilakukan pengelompokan dengan metode *single linkage*, *complete linkage*, dan *average linkage*. Tahap terakhir yaitu dilakukannya tahap evaluasi menggunakan *silhouette coefficient*.

4 HASIL DAN PEMBAHASAN

4.1 Pengujian Single Linkage

Pada pengujian menggunakan parameter jarak *Single Linkage* untuk mendapatkan 8 *cluster* maka titik potong yang digunakan akan dimulai dari 398 hingga 391.

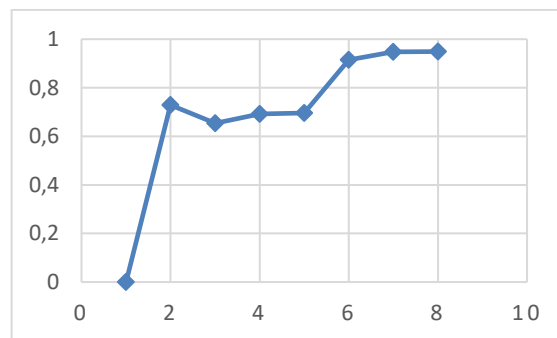


Gambar 2 Grafik Pengujian Single Linkage

Grafik hasil dari pengujian *clustering* data pelanggan menggunakan *single linkage* dapat dilihat pada Gambar 2. Hasil paling tinggi terdapat di titik potong 391 berjumlah 8 *cluster* dan memiliki nilai rata-rata dari *silhouette coefficient* keseluruhan datanya 0.968850698. Sedangkan paling rendah terdapat pada titik potong 398 dengan jumlah 1 *cluster* dan memiliki nilai rata-rata *silhouette coefficient* keseluruhan data yaitu 0. Pada titik potong 391 dapat diartikan memiliki hasil paling optimal dalam mengelompokkan data pelanggan *barbershop* karena memiliki *silhouette coefficient* tertinggi. Sedangkan pada titik potong 398 memiliki hasil paling kurang optimal karena memiliki *silhouette coefficient* terendah.

4.2 Pengujian Complete Linkage

Pada pengujian menggunakan parameter jarak *Complete Linkage* untuk mendapatkan 8 *cluster* maka titik potong yang digunakan akan dimulai dari 398 hingga 391.

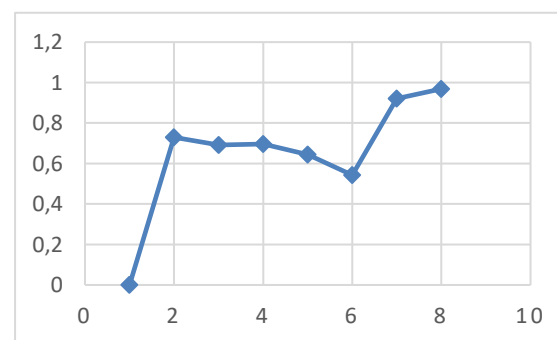


Gambar 3 Grafik Pengujian Complete Linkage

Grafik hasil dari pengujian *clustering* data pelanggan menggunakan *single linkage* dapat dilihat pada Gambar 6.3. Hasil paling tinggi terdapat di titik potong 391 berjumlah 8 *cluster* dan memiliki nilai rata-rata dari *silhouette coefficient* keseluruhan datanya 0.949077239. Sedangkan paling rendah terdapat pada titik potong 398 dengan jumlah 1 *cluster* dan memiliki nilai rata-rata *silhouette coefficient* keseluruhan data yaitu 0. Pada titik potong 391 dapat diartikan memiliki hasil paling optimal dalam mengelompokkan data pelanggan *barbershop* karena memiliki *silhouette coefficient* tertinggi. Sedangkan pada titik potong 398 memiliki hasil paling kurang optimal karena memiliki *silhouette coefficient* terendah.

4.3 Pengujian Average Linkage

Pada pengujian menggunakan parameter jarak *Average Linkage* untuk mendapatkan 8 *cluster* maka titik potong yang digunakan akan dimulai dari 398 hingga 391.



Gambar 4 Grafik Pengujian Average Linkage

Grafik hasil dari pengujian *clustering* data pelanggan menggunakan *average linkage* dapat dilihat pada Gambar 6.3. Hasil paling tinggi terdapat di titik potong 391 berjumlah 8 *cluster* dan memiliki nilai rata-rata dari *silhouette coefficient* keseluruhan datanya 0.968850698.

Sedangkan paling rendah terdapat pada titik potong 398 dengan jumlah 1 *cluster* dan memiliki nilai rata-rata *silhouette coefficient* keseluruhan data yaitu 0. Pada titik potong 391 dapat diartikan memiliki hasil paling optimal dalam mengelompokkan data pelanggan *barbershop* karena memiliki *silhouette coefficient* tertinggi. Sedangkan pada titik potong 398 memiliki hasil paling kurang optimal karena memiliki *silhouette coefficient* terendah.

4.4 Evaluasi

Analisis pengujian akan dilakukan dengan membandingkan ketiga parameter jarak berdasarkan rata-rata nilai *silhouette coefficient* tertingginya. Standar nilai *silhouette coefficient* berada di *range* -1 sampai dengan 1. Jika nilai *silhouette coefficient*-nya mendekati atau sama dengan 1 berarti setiap data yang diproses berada di *cluster* yang benar. Sedangkan jika nilai *silhouette coefficient* semakin mendekati atau sama dengan 0 maka data berada di *cluster* yang kurang tepat. Sementara, jika nilai mendekati atau sama dengan -1 maka data berada di *cluster* yang tidak tepat.

Tabel 1 Silhouette Coefficient Tertinggi setiap Parameter Jarak

Parameter Jarak	Silhouette Coefficient
Single Linkage	0.968850698
Complete Linkage	0.949077239
Average Linkage	0.968850698

Berdasarkan Tabel 1 dapat diketahui bahwa nilai rata-rata *silhouette coefficient* tertinggi terdapat pada parameter jarak *single linkage* dan *average linkage* dengan sama-sama memiliki nilai 0.968850698. Nilai *silhouette coefficient* tersebut sangat mendekati 1 sehingga dapat diartikan bahwa *cluster-cluster* data terbentuk dengan tepat. Pada pengujian ini diketahui bahwa nilai *silhouette coefficient* tertinggi pada masing-masing parameter berada di jumlah *cluster* 8 yang merupakan *cluster* terbanyak pada percobaan titik potong. Jadi dapat diartikan bahwa semakin tinggi *cluster* yang dihasilkan maka nilai *silhouette coefficient* akan semakin tinggi yang berarti data-data yang dikelompokkan semakin akurat. Nilai *silhouette coefficient* dipengaruhi oleh jarak kedekatan antar data yang pada penelitian ini dihitung menggunakan *manhattan distance*. Semakin dekat jarak antar

setiap data maka akan menghasilkan nilai *silhouette coefficient* yang tinggi. Variabel-variabel yang dapat mempengaruhi tingkat ketepatan suatu data pada *cluster* pada penelitian ini yaitu jumlah data yang diuji dan tingkat keberagaman data.

Parameter jarak dengan nilai *silhouette coefficient* tertinggi yaitu *single linkage* dan *average linkage* berada di titik potong 391. Pada titik potong ini nilai *silhouette coefficient* setiap data dan setiap *cluster* pada parameter jarak *single linkage* dan *average linkage* memiliki hasil yang sama. Nilai rata-rata *silhouette coefficient* tertinggi pada setiap data adalah 1 untuk 376 data, sedangkan yang terendah bernilai 0 pada 1 data. Sementara untuk nilai rata-rata *silhouette coefficient* setiap *cluster* yang paling memiliki nilai 1 pada *cluster* 1 hingga *cluster* 6, sedangkan yang terendah pada *cluster* 0 dengan nilai 0. Dapat diartikan bahwa hampir seluruh data pada *cluster* dikelompokkan dengan benar.

5 PENUTUP

5.1 Kesimpulan

Berdasarkan hasil pengujian dan analisis pada implementasi metode *agglomerative hierarchical clustering* pada segmentasi pelanggan *barbershop* dapat ditarik kesimpulannya sebagai berikut :

1. Penerapan metode *Agglomerative Hierarchical Clustering* pada penelitian ini menghasilkan 2 parameter jarak terbaik yaitu *single linkage* dan *average linkage*. Kedua parameter jarak ini dalam proses pengelompokan data memberikan hasil paling optimal pada titik potong 391 dengan jumlah 8 *cluster*.
2. Tingkat akurasi metode *Agglomerative Hierarchical Clustering* yang diuji menggunakan *Silhouette Coefficient* pada *single linkage* dan *average linkage* yang merupakan parameter jarak terbaik mendapatkan nilai rata-rata paling tinggi yaitu 0.968850698. Pada pengujian pada setiap data mendapatkan hasil 376 dari 399 data berada tepat pada *cluster*-nya dan pengujian pada setiap *cluster*-nya mendapatkan hasil 6 dari 8 *cluster* yang terbentuk tepat.

5.2 Saran

Berdasarkan kesimpulan yang didapatkan, peneliti memiliki beberapa saran yaitu :

1. Pada penelitian selanjutnya dapat menerapkan metode pengukuran dan parameter jarak yang lain untuk meningkatkan kemungkinan akurasi yang lebih tinggi dari hasil yang didapatkan.
2. Menggunakan strategi pengelompokan *Hierarchical Clustering* yang lain untuk melakukan pengelompokan data.
Dapat menggunakan variasi fitur dan jumlah data yang lebih banyak sehingga meningkatkan kompleksitas data yang akan digunakan untuk segmentasi.

6 DAFTAR PUSTAKA

- Aloysius, A. K. 2017. Implementasi Algoritma *Agglomerative Hierarchical Clustering* Untuk Mengelompokkan Capaian Belajar Siswa SD. Universitas Sanata Dharma Yogyakarta.
- Buttle, F. 2004. *Customer relationship management*. Routledge.
- Donny, I. S., Ema, R., Warih, M. 2013. Analisis dan Implementasi *Hierarchical Agglomerative Clustering* Pada Dokumen Berita Berbahasa Indonesia. Fakultas Teknik Informatika, Universitas Telkom.
- Feldman, R., & Sanger, J. 2007. *The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data*. Cambridge: Cambridge University Press.
- Han, Jiawei and Kamber, Micheline. 2006. *Data Mining: Concepts and Technologies. Second Edition*. (Ebook). Oxford: Elsevier.
- Larose, T. Daniel & Larose, D. Chantal (2005). *Discovery Knowledge in Data : An Introduction to Data mining 2nd Edition*. John Willey & Sons, Inc.
- Nugraheni, Y. 2011. Data mining dengan metode fuzzy untuk *customer relationship managment* (CRM) pada perusahaasn retail. Universitas Udayana.
- Widyawati, Saptomo W. L. Y, Utami Y. R. W. 2020. Penerapan *Agglomerative Hierarchical Clustering* Untuk Segmentasi Pelanggan. *STMIK Sinar Nusantara Surakarta. Vol : 18*.