

ABSTRAK

Penelitian ini mengeksplorasi implementasi dari metode *Agglomerative Hierarchical Clustering* (AHC) pada data reservasi hotel yang berjumlah 36.275 dengan 18 atribut untuk mengelompokkan *preferensi* pelanggan. Dengan berfokus pada optimasi variasi parameter untuk menghasilkan jumlah *cluster* dengan kategori terbaik. Data yang berasal dari transaksi reservasi akan digunakan untuk mengidentifikasi pola preferensi pelanggan, memberikan informasi bagi strategi pengembangan manajemen yang cerdas dan efektif, serta berkontribusi pada pengembangan ilmu data mining. Pada penelitian ini, metode yang digunakan adalah *agglomerative hierarchical clustering*. Kemudian *distance measure* yang digunakan adalah Euclidean dan Manhattan serta menggunakan *Linkage Single, Average, Complete* dan *Centroid*. Hasil dari penelitian adalah pengelompokan dapat dilakukan dengan menggunakan metode *agglomerative hierarchical clustering* pada 5 atribut dominan dengan variasi parameter matriks Euclidean dan metode penggabungan menggunakan *complete linkage* memberikan hasil *silhouette coefficient* terbaik pada *cluster 2* yaitu 0.900940. Preferensi tamu pada 2 *cluster*; *cluster* pertama merupakan tamu baru dengan pemesanan jangka Panjang pada bulan Agustus, September dan Oktober, terdiri dari 2 orang, dan lebih menyukai pada akhir pekan dan hari kerja. *Cluster* kedua adalah tamu berulang dengan Riwayat pembatalan tinggi, durasi menginap relatif pendek. Saran dalam penelitian ini, dapat dilakukan penelitian dengan mencangkup pengetahuan yang lebih luas seperti tingkat okupansi, *high* dan *low season*, serta strategi *market* kemudian dapat menambahkan metode validasi.

Kata Kunci: *Agglomerative Hierarchical Clustering, linkage, matrik jarak, data mining, segmentasi pelanggan*

ABSTRACT

This research explores the implementation of Agglomerative Hierarchical *Clustering* (AHC) method on 36,275 hotel reservation data with 18 attributes to *cluster* customer preferences. By focusing on optimizing parameter variations to produce the number of *clusters* with the best category. Data derived from reservation transactions will be used to identify customer preference patterns, provide information for smart and effective management development strategies, and contribute to the development of data mining science. In this research, the method used is agglomerative hierarchical *clustering*. Then the Distance measure used is Euclidean and Manhattan and uses Linkage Single, Average, Complete and Centroid. The results of the study are *clustering* can be done using the agglomerative hierarchical *clustering* method on 5 dominant attributes with variations in Euclidean matrix parameters and the merging method using complete linkage gives the best silhouette coefficient results in *cluster 2* which is 0.900940. Guest preferences in 2 *clusters*, the first *cluster* is new guests with long-term bookings in August, September and October, consisting of 2 people, and prefer weekends and weekdays. The second *cluster* is repeat guests with high cancellation history, relatively short duration of stay. Suggestions in this study, research can be carried out by covering broader knowledge such as occupancy rates, high and low seasons, and market strategies and then adding validation methods.

Keywords: Agglomerative Hierarchical *Clustering*, linkage, distance matrix, data mining, customer segmentation.