

ABSTRAK

Kedai kopi telah menjadi bagian integral dari kehidupan masyarakat modern seiring dengan perubahan gaya hidup dan meningkatnya budaya konsumsi kopi. Pertumbuhan industri kedai kopi yang pesat telah memicu tingkat kompetisi yang semakin ketat di antara para pelaku usaha. Situasi ini menuntut penerapan strategi yang efektif untuk menarik minat pelanggan sekaligus mempertahankan loyalitas mereka. Salah satu pendekatan yang dapat digunakan dalam perumusan strategi tersebut adalah segmentasi pelanggan. Melalui segmentasi pelanggan, preferensi dan pola pembelian dapat diidentifikasi secara lebih mendalam, sehingga menjadi dasar yang kuat untuk merancang strategi penjualan yang lebih tepat sasaran. Strategi yang disusun berdasarkan hasil segmentasi ini diharapkan mampu meningkatkan kepuasan pelanggan, mengoptimalkan efisiensi penjualan, serta mendorong peningkatan penjualan pada kedai kopi di tengah kompetisi yang semakin dinamis.

Berbagai penelitian sebelumnya telah membahas segmentasi pelanggan dengan menerapkan algoritma seperti *K-means*, *hierarchical clustering*, dan DBSCAN, di mana *K-means* menunjukkan kinerja yang lebih optimal dibandingkan algoritma lainnya. Berdasarkan temuan tersebut, penelitian ini menggunakan algoritma *K-means* untuk melakukan segmentasi pelanggan. Penelitian sebelumnya juga menunjukkan kelemahan *K-means* berupa ketergantungan terhadap inisialisasi jumlah *cluster* (K) dan titik pusat awal (*centroid*). Untuk mengatasi masalah ini, penelitian ini memanfaatkan algoritma *firefly* sebagai metode inisialisasi *centroid* karena kemampuannya dalam mengeksplorasi ruang fitur secara menyeluruh, sehingga dapat mengurangi risiko konvergensi dini. Namun, karena algoritma *firefly* memiliki tingkat komputasi yang cukup tinggi dan masalah *curse of dimensionality* yang kerap terjadi pada *K-means*, teknik reduksi dimensi melalui *Principal Component Analysis* (PCA) diterapkan guna menyederhanakan kompleksitas data tanpa kehilangan informasi penting. Dataset yang digunakan dalam penelitian ini merupakan data transaksi harian dari kedai kopi Ruang Ngopi selama periode April hingga Juni 2024, yang terdiri dari 2380 baris data dengan 8 fitur yang akan ditransformasikan sebelum proses analisis dilakukan.

Hasil penelitian menunjukkan bahwa kombinasi algoritma *firefly* dan PCA berhasil mengoptimalkan *K-means clustering* dengan menghasilkan 5 *cluster*. Berdasarkan analisis nilai *cumulative variance*, komponen PCA yang digunakan adalah komponen dengan nilai > 0.9 , di mana diperoleh 4 komponen utama yang digunakan dalam proses optimasi. Untuk mengevaluasi kinerja model, digunakan metrik *silhouette score* dan *sum of squared errors* (SSE). Model *K-means* yang dioptimalkan dengan *firefly* + PCA menghasilkan *silhouette score* sebesar 0.489 dan SSE sebesar 4836, sementara metode *K-means* konvensional hanya mencapai *silhouette score* sebesar 0.448 dan SSE sebesar 5221.

Kata Kunci: kedai kopi, *clustering*, segmentasi pelanggan, *K-means*, *principal component analysis*, *firefly*

ABSTRACT

Coffee shops have become an integral part of modern society, driven by changing lifestyles and the growing coffee consumption culture. The rapid growth of the coffee shop industry has triggered increasingly tight competition among business players. This situation demands the implementation of effective strategies to attract customer interest while maintaining their loyalty. One approach that can be used in formulating such strategies is customer segmentation. Through customer segmentation, preferences and purchasing patterns can be identified in greater depth, providing a solid foundation for designing more targeted marketing strategies. Strategies developed based on segmentation results are expected to enhance customer satisfaction, optimize marketing efficiency, and drive increased sales in coffee shops amid an increasingly dynamic competitive environment.

Previous studies have discussed customer segmentation by applying algorithms such as K-means, hierarchical clustering, and DBSCAN, where K-means has shown more optimal performance compared to other algorithms. Based on these findings, this study employs the K-means algorithm to perform customer segmentation, although this method has weaknesses, particularly its dependence on the initialization of the number of clusters (K) and the initial centroids. To address this issue, this study utilizes the firefly algorithm as a method for centroid initialization due to its ability to explore the feature space comprehensively, thereby reducing the risk of premature convergence. However, because the firefly algorithm requires significant computational resources and problem of curse of dimensionality that usually happen in K-means with complex data, dimensionality reduction techniques using Principal Component Analysis (PCA) are applied to simplify data complexity without losing valuable information. The dataset used in this study consists of daily sales data from the coffee shop Ruang Ngopi for the period April to June 2024, comprising 2380 rows of data with 8 features that will be transformed before the analysis process is carried out.

The results of the study indicate that the combination of the firefly algorithm and PCA successfully optimizes K-means clustering, producing 5 clusters. Based on the cumulative variance analysis, the PCA components used are those with a cumulative variance value > 0.9 , where 4 main components were obtained for the optimization process. To evaluate the model's performance, the silhouette score and sum of squared errors (SSE) metrics were used. The K-means model optimized with firefly + PCA achieved a silhouette score of 0.489 and an SSE of 4836, whereas the conventional K-means method only achieved a silhouette score of 0.448 and an SSE of 5221.

Keywords: *coffee shop, clustering, customer segmentation, K-means, principal component analysis, firefly*