

## ABSTRAK

Penelitian ini dilakukan melalui serangkaian tahapan yang mencakup pra-pemrosesan data, reduksi dimensi menggunakan *Principal Component Analysis (PCA)* dan *Uniform Manifold Approximation and Projection (UMAP)*, serta penerapan algoritma *K-Means* dan *DBSCAN* untuk menentukan metode clustering yang paling optimal. Kedua algoritma diuji menggunakan teknik reduksi dimensi yang sama agar perbandingan kinerja dapat dilakukan secara adil dan objektif. Pemilihan model terbaik didasarkan pada nilai *Silhouette Score* tertinggi dan *Davies–Bouldin Index (DBI)* terendah sebagai indikator kualitas cluster. Hasil pengujian menunjukkan bahwa kombinasi *K-Means* dengan *PCA* memberikan performa paling unggul dengan *Silhouette Score* sebesar 0,9279 dan *DBI* sebesar 0,1038, yang menandakan bahwa cluster yang terbentuk memiliki tingkat kekompakan yang tinggi, pemisahan yang jelas antar cluster. Di sisi lain, *DBSCAN* menunjukkan performa terbaiknya ketika dipadukan dengan *UMAP*, dengan *Silhouette Score* sebesar 0,8649 dan *DBI* sebesar 0,2035. Hal ini mengindikasikan bahwa *UMAP* lebih efektif dalam mempertahankan struktur kepadatan lokal data dibandingkan dengan *PCA*. Meskipun demikian, secara keseluruhan *K-Means* tetap menunjukkan kualitas clustering yang lebih baik dibandingkan *DBSCAN* karena mampu menghasilkan nilai *Silhouette Score* yang lebih tinggi dan *DBI* yang lebih rendah, sehingga lebih andal dalam membentuk segmentasi yang stabil, terstruktur, dan mudah diinterpretasikan.

Hasil segmentasi yang diperoleh tidak hanya memiliki validitas kuantitatif, tetapi juga relevansi praktis yang kuat dalam mendukung strategi pemasaran *marketplace*. Melalui proses *profiling*, setiap cluster terbukti mencerminkan segmen pasar yang berbeda dengan karakteristik khusus, seperti kecenderungan terhadap brand tertentu, dominasi kategori produk, variasi rentang harga, serta proporsi produk *cruelty-free*. Informasi ini dapat dimanfaatkan untuk merancang strategi pemasaran yang lebih terfokus dan personal, mengoptimalkan pengelolaan inventori, menyusun strategi harga yang lebih kompetitif, mengembangkan program loyalitas pelanggan, serta mengidentifikasi peluang inovasi produk baru. Dengan demikian, kombinasi *K-Means* dan *PCA* ditetapkan sebagai pendekatan utama yang paling efektif dalam penelitian ini, sedangkan *DBSCAN* dengan *UMAP* berperan sebagai metode pendukung yang bermanfaat untuk menangkap pola kepadatan lokal dan mendeteksi produk-produk unik yang berpotensi membuka peluang pasar baru.

## ABSTRACT

This study was conducted through a series of stages including data preprocessing, dimensionality reduction using Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP), and the application of K-Means and DBSCAN algorithms to determine the most optimal clustering method. Both algorithms were evaluated using the same dimensionality reduction techniques to ensure a fair and objective performance comparison. The selection of the best model was based on the highest Silhouette Score and the lowest Davies–Bouldin Index (DBI) as indicators of cluster quality. The experimental results show that the combination of K-Means with PCA achieved the best performance, with a Silhouette Score of 0.9279 and a DBI of 0.1038, indicating that the resulting clusters are highly compact and clearly separated. On the other hand, DBSCAN achieved its best performance when combined with UMAP, obtaining a Silhouette Score of 0.8649 and a DBI of 0.2035. This suggests that UMAP is more effective than PCA in preserving the local density structure of the data. Nevertheless, overall, K-Means demonstrates superior clustering quality compared to DBSCAN, as it produces a higher Silhouette Score and a lower DBI, making it more reliable in forming stable, well-structured, and easily interpretable segmentation.

The resulting segmentation not only shows strong quantitative validity but also offers significant practical relevance in supporting marketplace marketing strategies. Through cluster profiling, each cluster represents a distinct market segment with specific characteristics, such as preferences toward certain brands, dominance of particular product categories, variations in price ranges, and the proportion of cruelty-free products. This information can be utilized to design more focused and personalized marketing strategies, optimize inventory management, develop more competitive pricing strategies, enhance customer loyalty programs, and identify opportunities for new product innovation. Therefore, the combination of K-Means and PCA is established as the most effective primary approach in this study, while DBSCAN with UMAP serves as a complementary method that is useful for capturing local density patterns and detecting unique products that may open new market opportunities.