

ABSTRAK

Usaha Mikro, Kecil, dan Menengah (UMKM) merupakan sektor penting dalam mendukung pertumbuhan ekonomi Indonesia, termasuk di Kabupaten Sidoarjo. Namun, dengan jumlah UMKM yang terus berkembang pesat, tanpa adanya pengelompokan yang tepat, sulit bagi Dinas Koperasi dan Usaha Mikro Kabupaten Sidoarjo untuk memahami perbedaan potensi, dan tantangan yang dihadapi oleh setiap kelompok UMKM. Penelitian ini bertujuan untuk mengelompokkan UMKM di Kabupaten Sidoarjo menggunakan algoritma *DBSCAN* dengan pendekatan metodologi *CRISP-DM* dimulai dari pemahaman bisnis hingga penerapan hasil. Data yang digunakan terdiri dari 1.479 UMKM dengan atribut meliputi modal, omset perbulan, tenaga kerja, izin usaha, marketplace, jenis usaha, dan kecamatan. Data diolah melalui tahapan transformasi dengan metode *label encoding*, dan normalisasi menggunakan metode *MaxAbs*. Algoritma *DBSCAN* digunakan dengan membandingkan tiga metode pengukuran jarak yaitu *euclidean*, *manhattan*, dan *hamming* kemudian dikelompokkan dengan melakukan eksperimen menggunakan parameter *epsilon* dan *minpts* yang divariasikan untuk masing-masing metode. Evaluasi hasil klasterisasi menggunakan *silhouette coefficient* dan *davies-bouldin index*. Hasil uji coba menunjukkan bahwa metode *euclidean Distance* memberikan hasil yang paling baik dibandingkan *euclidean* dan *hamming*, dengan nilai *epsilon* 0,1 dan *minpts* 36 menghasilkan 115 *outlier*, mendapatkan nilai *silhouette coefficient* sebesar 0,99, dan *davies-bouldin index* sebesar 0,02, menunjukkan struktur yang sangat kuat dan pemisahan antar klaster yang sangat jelas. Klaster yang terbentuk terdiri dari *cluster* 1, 3, 4, dan 5 berskala mikro, serta *cluster* 2 berskala kecil. Hasil penelitian ini diharapkan dapat menjadi dasar rekomendasi dalam pengambilan keputusan dan kebijakan berbasis data untuk pengembangan ekonomi lokal secara lebih tepat sasaran.

Kata Kunci: UMKM, *DBSCAN*, *CRISP-DM*, *Clustering*, Perbandingan Jarak.

ABSTRACT

Micro, Small, and Medium Enterprises (MSMEs) are an important sector in supporting Indonesia's economic growth, including in Sidoarjo Regency. However, with the rapidly increasing number of MSMEs, without proper grouping, it is difficult for the Sidoarjo Regency Cooperative and Micro Business Office to understand the differences in potential and challenges faced by each MSME group. This study aims to cluster MSMEs in Sidoarjo Regency using the DBSCAN algorithm with the CRISP-DM methodology approach, starting from business understanding to result implementation. The data used consists of 1,479 MSMEs with attributes including capital, monthly turnover, workforce, business permits, marketplaces, types of businesses, and sub-districts. The data was processed through transformation stages using the label encoding method, and normalization using the MaxAbs method. The DBSCAN algorithm was used by comparing three distance measurement methods: Euclidean, Manhattan, and Hamming, and then grouped by conducting experiments using varied epsilon and minpts parameters for each method. Evaluation of the clustering results was conducted using the silhouette coefficient and the Davies-Bouldin index. The experimental results showed that the Euclidean Distance method provided the best results compared to the Euclidean and Hamming methods, with an epsilon value of 0.1 and minpts of 36 resulting in 115 outliers, achieving a silhouette coefficient of 0.99 and a Davies-Bouldin index of 0.02, indicating a very strong structure and a very clear separation between clusters. The formed clusters consist of clusters 1, 3, 4, and 5 on a micro scale, and cluster 2 on a small scale. The results of this research are expected to serve as a basis for recommendations in data-based decision-making and policies for more targeted local economic development.

Keywords: MSMEs, DBSCAN, CRISP-DM, Clustering, Distance Comparison.