

ABSTRAK

Usaha Mikro, Kecil, dan Menengah (UMKM) di Kecamatan Sampang memiliki peran penting dalam pertumbuhan ekonomi daerah, penyerapan tenaga kerja, dan peningkatan kesejahteraan masyarakat. Namun, perbedaan karakteristik usaha seperti omzet, aset, kapasitas produksi, jumlah tenaga kerja, legalitas usaha, serta pemanfaatan teknologi digital menyebabkan identifikasi potensi UMKM masih sulit dilakukan secara objektif. Kondisi ini mengakibatkan program pembinaan dan bantuan usaha belum sepenuhnya tepat sasaran. Penelitian ini bertujuan melakukan segmentasi dan klasifikasi potensi UMKM menggunakan algoritma K-Means dan XGBoost dengan pendekatan CRISP-DM yang meliputi business understanding, data understanding, data preparation, modeling, evaluation, dan deployment. Data yang digunakan merupakan data sekunder sebanyak 1.056 UMKM yang diperoleh dari Dinas Koperasi, Perindustrian, dan Perdagangan Kabupaten Sampang. Variabel penelitian meliputi omzet per tahun, aset, jumlah tenaga kerja, kapasitas produksi per tahun, NIB/SKU, penggunaan media sosial, kehadiran di marketplace, dan kepemilikan lahan. Hasil penelitian menunjukkan bahwa segmentasi terbaik diperoleh pada jumlah cluster $k = 8$ dengan nilai Silhouette Coefficient sebesar 0,7867 yang menunjukkan kualitas cluster baik. Pada tahap klasifikasi, model XGBoost Baseline dan model XGBoost dengan tambahan informasi cluster memperoleh nilai akurasi yang sama sebesar 0,9670. Namun, model baseline memberikan hasil lebih baik pada recall macro sebesar 0,9613 dan F1-score macro sebesar 0,9641, sedangkan model dengan tambahan cluster memperoleh recall macro 0,9601 dan F1-score macro 0,9636. Berdasarkan hasil tersebut, algoritma K-Means efektif digunakan untuk segmentasi potensi UMKM, sedangkan XGBoost Baseline menjadi model terbaik untuk klasifikasi potensi UMKM. Hasil penelitian ini diharapkan dapat mendukung pengambilan keputusan yang lebih tepat sasaran dalam pembinaan dan pengembangan UMKM di Kecamatan Sampang.

Kata Kunci: UMKM, K-Means, XGBoost, segmentasi, klasifikasi.

ABSTRACT

Micro, Small, and Medium Enterprises (MSMEs) in Sampang District play an important role in regional economic growth, employment generation, and improving community welfare. However, differences in business characteristics such as turnover, assets, production capacity, number of workers, business legality, and the use of digital technology make the identification of MSME potential still difficult to conduct objectively. This condition causes business development programs and assistance to be not yet fully well-targeted. This study aims to perform MSME potential segmentation and classification using the K-Means and XGBoost algorithms with the CRISP-DM approach, which includes business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The data used in this study were secondary data consisting of 1,056 MSMEs obtained from the Department of Cooperatives, Industry, and Trade of Sampang Regency. The research variables include annual turnover, assets, number of workers, annual production capacity, NIB/SKU, social media usage, marketplace presence, and land ownership. The results show that the best segmentation was obtained at $k = 8$ clusters with a Silhouette Coefficient value of 0.7867, indicating good cluster quality. In the classification stage, the Baseline XGBoost model and the XGBoost model with additional cluster information achieved the same accuracy value of 0.9670. However, the baseline model performed better in terms of macro recall, with a value of 0.9613, and macro F1-score, with a value of 0.9641, while the model with additional cluster information obtained a macro recall of 0.9601 and a macro F1-score of 0.9636. Based on these results, the K-Means algorithm is effective for MSME potential segmentation, while Baseline XGBoost is the best model for MSME potential classification. The results of this study are expected to support more targeted decision-making in the development and empowerment of MSMEs in Sampang District.

Keywords: MSMEs, K-Means, XGBoost, segmentation, classification.