

ABSTRAK

Yogantara, Urip (2026), *Analisis Komparatif Algoritma K-Means, HDBSCAN, dan Self-Organizing Map untuk Profiling Siswa SMA/SMK Berdasarkan Nilai Akademik dan Status Sosial-Ekonomi*. Tesis, Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha.

Tesis ini sudah disetujui dan diperiksa oleh Pembimbing 1: I Gede Aris Gunadi, S.Si., M.Kom. dan Pembimbing II: Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs.

Kata Kunci: Profiling Siswa, *K-Means*, HDBSCAN, *Self-Organizing Map*, Sosial-Ekonomi.

Keberagaman kemampuan akademik dan latar belakang sosial-ekonomi siswa SMA/SMK menuntut pendekatan pembelajaran berdiferensiasi yang berbasis bukti. Namun, profiling siswa secara komprehensif yang mengintegrasikan data akademik dan sosial-ekonomi secara simultan belum banyak dilakukan. Penelitian ini bertujuan memetakan profil siswa SMA/SMK dengan mengintegrasikan data nilai akademik dan status sosial-ekonomi menggunakan metode *machine learning non-supervised*, sekaligus membandingkan tiga algoritma klasterisasi dengan karakteristik berbeda: *K-Means*, HDBSCAN, dan *Self-Organizing Map*. Data yang dianalisis mencakup 42.722 siswa dengan dua fitur utama, yaitu nilai akademik dan tingkat kesejahteraan sosio-ekonomi. Kinerja algoritma dievaluasi menggunakan *Silhouette Score* dan *Davies-Bouldin Index* (DBI). Hasil evaluasi menunjukkan *K-Means* sebagai algoritma paling efektif dengan *Silhouette Score* tertinggi (0,3658) dan DBI terendah (0,8589), diikuti SOM dengan *Silhouette* sebesar 0,3350 dan DBI sebesar 0,9185, sedangkan HDBSCAN memiliki kinerja terlemah dengan *Silhouette* 0,0192 dan DBI 3,3981. *K-Means* membentuk empat profil sosial-akademik yang seimbang dan distingtif yaitu Mampu–Akademik Tinggi (23,1%), Kurang Mampu–Akademik Rendah (25,5%), Mampu–Akademik Rendah (29,2%), dan Kurang Mampu–Akademik Tinggi (22,2%). Penelitian ini menyimpulkan bahwa *K-Means* merupakan algoritma paling sesuai untuk segmentasi profil siswa pada data sosial-akademik berdimensi rendah, dan peta profil yang dihasilkan dapat menjadi dasar bagi sekolah maupun pemerintah dalam merancang pembelajaran berdiferensiasi dan kebijakan pendidikan yang lebih tepat sasaran.

ABSTRACT

Yogantara, Urip (2026), *Comparative Analysis of K-Means, HDBSCAN, and Self-Organizing Map Algorithms for Profiling Senior and Vocational High School Students Based on Academic Performance and Socioeconomic Status*. Thesis, Computer Science, Postgraduate Program, Ganesha University of Education.

This thesis has been reviewed, examined, and approved by Supervisor I, I Gede Aris Gunadi, S.Si., M.Kom., and Supervisor II, Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs.

Keywords: Student Profiling, K-Means, HDBSCAN, Self-Organizing Map, Socio-Economic.

The diversity of academic abilities and socio-economic backgrounds among senior high school (SMA/SMK) students calls for an evidence-based differentiated learning approach. However, comprehensive student profiling that integrates academic and socio-economic data simultaneously has rarely been carried out. This study aims to map the profiles of SMA/SMK students by integrating academic performance and socio-economic status data using an unsupervised machine learning approach, while comparing three clustering algorithms with distinct characteristics: K-Means, HDBSCAN, and Self-Organizing Map. The analyzed data comprises 42,722 students with two main features, namely academic performance and socio-economic welfare level. Algorithm performance was evaluated using the Silhouette Score and the Davies-Bouldin Index (DBI). The results show that K-Means is the most effective algorithm, achieving the highest Silhouette Score (0.3658) and the lowest DBI (0.8589), followed by SOM with a Silhouette of 0.3350 and a DBI of 0.9185, while HDBSCAN performed the weakest with a Silhouette of 0.0192 and a DBI of 3.3981. K-Means produced four balanced and distinctive socio-academic profiles, namely Well-off-High Achieving (23.1%), Underprivileged-Low Achieving (25.5%), Well-off-Low Achieving (29.2%), and Underprivileged-High Achieving (22.2%). This study concludes that K-Means is the most suitable algorithm for segmenting student profiles on low-dimensional socio-academic data, and the resulting profile map can serve as a basis for schools and governments in designing differentiated learning and more precisely targeted education policies.