

**ANALISIS KOMPARATIF ALGORITMA
K-MEANS, HDBSCAN, DAN SELF-ORGANIZING
MAP UNTUK PROFILING SISWA SMA/SMK
BERDASARKAN NILAI AKADEMIK DAN STATUS
SOSIAL-EKONOMI**

TESIS



**PROGRAM STUDI ILMU KOMPUTER (S2)
PROGRAM PASCASARJANA
UNIVERSITAS PENDIDIKAN GANESHA**



- UU ITE No. 11 Tahun 2008 Pasal 5 Ayat 1 "Informasi Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"
- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSrE - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSrE
- Cetakan dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.



- UU ITE No. 11 Tahun 2008 Pasal 5 Ayat 1 "Informasi Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"
- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSrE - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSrE
- Cetak dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.



Balai
Sertifikasi
Elektronik

TESIS

DIAJUKAN UNTUK MELENGKAPI TUGAS DAN MEMENUHI SYARAT-SYARAT UNTUK MENCAPAI GELAR MAGISTER KOMPUTER

	
Pembimbing I	Dr. I Gede Aris Gunadi, S.Si. M.Kom. NIP.197703182008121004
Pembimbing II	Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs. NIP.198307252008011008



- UU ITE No. 11 Tahun 2008 Pasal 5 Ayat 1 "Informasi Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"
- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSrE - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSrE
- Cetakn dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.

Tesis oleh I Gusti Putu Urip Yogantara telah dipertahankan di depan tim penguji dan dinyatakan diterima sebagai salah satu persyaratan untuk memperoleh gelar Magister Komputer di Ilmu Komputer (S2), Program Pascasarjana, Universitas Pendidikan Ganesha
Pada tanggal 03 Agustus 2026

Dewan Penguji

Ketua	Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs. NIP.198307252008011008
Anggota	Dr. I Gede Aris Gunadi, S.Si. M.Kom. NIP.197703182008121004
Anggota	Prof. Dr. I Made Candiasa, M.I.Kom. NIP.196012311986011004
Anggota	Prof. Dr. Luh Joni Erawati Dewi, S.T., M.Pd. NIP.197606252001122001

Mengetahui Direktur Program Pascasarjana
Universitas Pendidikan Ganesha,



Prof. Dr. I Nyoman Jampel, M.Pd.

NIP.195910101986031003



- UU ITE No. 11 Tahun 2008 Pasal 5 Ayat 1 "Informasi Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"
- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSrE - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSrE
- Cetakan dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.

LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya bahwa tesis yang saya susun sebagai syarat untuk memperoleh gelar Magister Ilmu Komputer dari Pascasarjana Universitas Pendidikan Ganesha seluruhnya merupakan hasil karya saya sendiri. Bagian-bagian tertentu dalam penulisan tesis yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas dan sesuai dengan norma, kaidah, serta etika akademis.

Apabila di kemudian hari ditemukan seluruh atau sebagian tesis ini bukan hasil karya saya sendiri atau adanya plagiat dalam bagian-bagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sandang dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku di wilayah Negara Kesatuan Republik Indonesia.

Singaraja, 3 Agustus 2026
Yang memberi pernyataan,

I Gusti Putu Urip Yogantara



PRAKATA

Puji dan syukur dipanjatkan kepada Tuhan Yang Maha Esa atas anugerah-Nya, sehingga tesis yang berjudul: “Analisis Komparatif Algoritma K-Means, HDBSCAN, dan Self-Organizing Map untuk Profiling Siswa SMA/SMK Berdasarkan Nilai Akademik dan Status Sosial-Ekonomi”, dapat diselesaikan sesuai dengan yang direncanakan. Tesis ini ditulis untuk memenuhi salah satu persyaratan untuk memperoleh gelar Magister Ilmu Komputer Pascasarjana Universitas Pendidikan Ganesha pada Program Studi Ilmu Komputer. terselesaikannya tesis ini telah banyak memperoleh uluran tangan dari berbagai pihak. Untuk itu, izinkan penulis menyampaikan terima kasih dan penghargaan kepada pihak-pihak berikut.

- I Gede Aris Gunadi, S.Si., M.Kom., sebagai pembimbing I yang telah dengan sabar membimbing, mengarahkan, dan memberikan motivasi yang demikian bermakna, sehingga penulis mampu melewati berbagai hambatan dalam perjalanan studi dan penyelesaian tesis ini;
- I Made Gede Sunarya, S.Kom., M.Cs., sebagai pembimbing II, yang dengan gaya dan pola komunikasi yang khas, telah melecut semangat, motivasi, dan harapan penulis selama penelitian dan penulisan naskah laporan tesis ini, sehingga tesis ini dapat terwujud dengan baik sesuai harapan;
- Prof. Dr. I Made Candiasa, M.I.Kom. dan Prof. Dr. Luh Joni Erawati Dewi, S.T., M.Pd. sebagai dosen penguji yang telah banyak memberikan masukan-masukan yang bermanfaat untuk penyempurnaan tesis ini;
- Koordinator Program Studi Ilmu Komputer dan staf dosen pengajar yang telah banyak membantu dan memotivasi penulis selama penyusunan tesis ini;
- Direktur Pascasarjana Undiksha dan staf, yang telah banyak membantu selama penulis menyelesaikan tesis ini;
- Rektor Universitas Pendidikan Ganesha, yang telah memberikan bantuan secara moral dan memfasilitasi berbagai kepentingan penulis dalam menyelesaikan tesis ini;
- Bapak I Gusti Gede Putu Urip dan Ibu Ini Putu Armini selaku orang tua penulis, serta Ni Putu Era Prasistya Damayanti, S.P. selaku istri penulis, yang telah banyak memberikan bantuan, dukungan material, dan motivasi moral selama proses penyelesaian tesis ini.

Semoga semua bantuan yang telah mereka berikan dalam menyelesaikan studi ini, mereka diberkati imbalan yang sepadan oleh Tuhan Yang Maha Esa, kesehatan, dan keharmonisan dalam menjalani kehidupan. Penulis menyadari bahwa tesis ini belum sempurna. Namun, kehadirannya dalam konstelasi masyarakat akademis akan menambah perbendaharaan ilmu dalam perkembangan ilmu pengetahuan. Semoga tesis ini bermanfaat bagi masyarakat akademis,

terutama mereka yang menyatakan diri bernaung di bawah kebesaran panji-panji pendidikan.

Singaraja, 3 Agustus 2026
Penulis,



DAFTAR ISI

PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN.....	xiv
ABSTRAK.....	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN.....	17
1.1 Latar Belakang.....	17
1.2 Identifikasi Masalah.....	23
1.3 Batasan Masalah.....	24
1.4 Rumusan Masalah.....	25
1.5 Tujuan Penelitian.....	25
1.6 Manfaat Penelitian.....	26
1.6.1 Manfaat Akademik.....	26
1.6.2 Manfaat Praktis.....	27
BAB II LANDASAN TEORI.....	28
2.1 Kajian Teori.....	28
2.1.1 <i>Data Mining</i> dalam Bidang Pendidikan.....	28
2.1.2 Konsep Klasterisasi (<i>Clustering</i>).....	30
2.1.3 Algoritma <i>K-Means Clustering</i>	33
2.1.4 Algoritma HDBSCAN.....	36
2.1.5 <i>Self-Organizing Map</i> (SOM).....	39
2.1.6 Evaluasi Kualitas Klaster.....	41
2.1.7 Pra-Pemrosesan Data untuk <i>Clustering</i>	43
2.1.8 Integrasi Data Akademik dan Sosial-Ekonomi.....	46
2.1.9 Profil Sosial-Akademik Siswa.....	48
2.1.10 Taksonomi Metode Klasterisasi dan Rasional Pemilihan Algoritma Pembanding.....	50
2.1.11 Pembelajaran Berdiferensiasi dan Diferensiasi Layanan Pendidikan.....	53
2.2 Kajian Hasil Penelitian yang Relevan.....	55
BAB III METODE PENELITIAN.....	63
3.1 Rancangan Penelitian.....	63
3.2 Pengumpulan Data.....	66
3.3 <i>Data Selection</i>	68
3.4 <i>Data Preprocessing</i>	70
3.4.1 Integrasi Data.....	70
3.4.2 Seleksi Data.....	72
3.4.3 Transformasi Data.....	74
3.5 Modeling.....	74
3.5.1 Perhitungan <i>K-Means Clustering</i>	75
3.5.2 Perhitungan HDBSCAN.....	79
3.5.3 Perhitungan SOM.....	82
3.6 Evaluasi.....	87

3.6.1	Evaluasi Kualitas Klaster	87
3.6.2	<i>Davies-Bouldin Index</i> (DBI)	88
3.6.3	<i>Silhouette coefficient</i>	89
BAB IV HASIL PENELITIAN DAN PEMBAHASAN		92
4.1	Prapemrosesan Data	92
4.2	Proses Klasterisasi	93
4.2.1	<i>K-Means Clustering</i>	93
4.2.2	<i>HDBSCAN Clustering</i>	95
4.2.3	<i>Self-Organizing Map</i> (SOM)	97
4.3	Hasil Klaster Masing-Masing Metode	100
4.3.1	Hasil <i>K-Means</i> ($K = 4$)	100
4.3.2	Hasil HDBSCAN (11 Klaster)	101
4.3.3	Hasil SOM (5 Klaster)	102
4.4	Evaluasi Klaster	103
4.5	Analisis Komparatif	106
4.5.1	Analisis Teoretis Keunggulan <i>K-Means</i>	109
4.6	Profiling Klaster Ketiga Algoritma	112
4.6.1	Cakupan Data dan Konvensi Pelaporan	113
4.6.2	Rata-rata Nilai Total per Klaster	114
4.6.3	Rata-rata Desil Provinsi per Klaster	119
4.6.4	Komposisi Desil Provinsi per Klaster	124
4.6.5	Komposisi Jenjang Pendidikan per Klaster	130
4.6.6	Komposisi Status Sekolah per Klaster	135
4.6.7	Komposisi Kabupaten/Kota per Klaster	140
4.6.8	Komposisi Klaster per Kabupaten/Kota	146
4.6.9	Ketidaklengkapan Data Kemiskinan per Klaster	152
4.6.10	Komposisi Jenis Lantai Terluas per Klaster	155
4.6.11	Komposisi Daya Listrik Terpasang per Klaster	161
4.6.12	Komposisi Bahan Bakar Utama Memasak per Klaster	167
4.6.13	Komposisi Sumber Air Minum Utama per Klaster	174
4.6.14	Ringkasan Profil Klaster per Algoritma	181
4.7	Implikasi Diferensiasi Layanan dan Pembelajaran Berbasis Profil Klaster 185	
BAB V PENUTUP		190
5.1	Simpulan	190
5.2	Saran	191
DAFTAR PUSTAKA		193
LAMPIRAN		202

DAFTAR TABEL

Tabel 2.1 <i>State of the Art</i> Penelitian Terdahulu	56
Tabel 3.1 Deskripsi Kolom <i>Dataset</i>	67
Tabel 3.2 Sampel Data	68
Tabel 3.3 Interpretasi Nilai <i>Silhouette (Rousseeuw, 1987)</i>	91
Tabel 4.1 Ringkasan <i>Missing Values Dataset</i>	92
Tabel 4.2 Hasil Tuning Parameter HDBSCAN	95
Tabel 4.3 Konfigurasi Pelatihan SOM	97
Tabel 4.4 Profil Klaster <i>K-Means</i> ($K = 4$)	100
Tabel 4.5 Profil Klaster HDBSCAN (11 Klaster + <i>Noise</i>)	101
Tabel 4.6 Profil Klaster SOM (5 Klaster)	102
Tabel 4.7 Perbandingan Metrik Evaluasi Klasterisasi	103
Tabel 4.8 Analisis Komparatif Komprehensif Ketiga Algoritma	108
Tabel 4.9 Cakupan Data Profiling Ketiga Algoritma	113
Tabel 4.10 Statistik Nilai Total per Klaster <i>K-Means</i>	114
Tabel 4.11 Statistik Nilai Total per Klaster HDBSCAN	116
Tabel 4.12 Statistik Nilai Total per Klaster SOM	117
Tabel 4.13 Statistik Desil Provinsi per Klaster <i>K-Means</i>	119
Tabel 4.14 Statistik Desil Provinsi per Klaster HDBSCAN	121
Tabel 4.15 Statistik Desil Provinsi per Klaster SOM	123
Tabel 4.16 Komposisi Desil Provinsi per Klaster <i>K-Means</i> (%)	125
Tabel 4.17 Komposisi Desil Provinsi per Klaster HDBSCAN (%)	127
Tabel 4.18 Komposisi Desil Provinsi per Klaster SOM (%)	128
Tabel 4.19 Komposisi Jenjang Pendidikan per Klaster <i>K-Means</i> (%)	131
Tabel 4.20 Komposisi Jenjang Pendidikan per Klaster HDBSCAN (%)	133
Tabel 4.21 Komposisi Jenjang Pendidikan per Klaster SOM (%)	134
Tabel 4.22 Komposisi Status Sekolah per Klaster <i>K-Means</i> (%)	136
Tabel 4.23 Komposisi Status Sekolah per Klaster HDBSCAN (%)	137
Tabel 4.24 Komposisi Status Sekolah per Klaster SOM (%)	139
Tabel 4.25 Komposisi Kabupaten/Kota per Klaster <i>K-Means</i> (%)	141
Tabel 4.26 Komposisi Kabupaten/Kota per Klaster HDBSCAN (%)	143
Tabel 4.27 Komposisi Kabupaten/Kota per Klaster SOM (%)	144
Tabel 4.28 Komposisi Klaster <i>K-Means</i> per Kabupaten/Kota (%)	147
Tabel 4.29 Komposisi Klaster HDBSCAN per Kabupaten/Kota (%)	149
Tabel 4.30 Komposisi Klaster SOM per Kabupaten/Kota (%)	150
Tabel 4.31 Persentase Data Sosial-Ekonomi Tidak Lengkap per Klaster <i>K-Means</i> (%)	153
Tabel 4.32 Persentase Data Sosial-Ekonomi Tidak Lengkap per Klaster HDBSCAN (%)	154
Tabel 4.33 Persentase Data Sosial-Ekonomi Tidak Lengkap per Klaster SOM (%)	155
Tabel 4.34 Komposisi Jenis Lantai Terluas per Klaster <i>K-Means</i> (%)	156
Tabel 4.35 Komposisi Jenis Lantai Terluas per Klaster HDBSCAN (%)	158
Tabel 4.36 Komposisi Jenis Lantai Terluas per Klaster SOM (%)	160
Tabel 4.37 Komposisi Daya Listrik Terpasang per Klaster <i>K-Means</i> (%)	162

Tabel 4.38 Komposisi Daya Listrik Terpasang per Klaster HDBSCAN (%).....	164
Tabel 4.39 Komposisi Daya Listrik Terpasang per Klaster SOM (%)	166
Tabel 4.40 Komposisi Bahan Bakar Utama Memasak per Klaster K-Means (%)	168
Tabel 4.41 Komposisi Bahan Bakar Utama Memasak per Klaster HDBSCAN (%)	170
Tabel 4.42 Komposisi Bahan Bakar Utama Memasak per Klaster SOM (%)	172
Tabel 4.43 Komposisi Sumber Air Minum Utama per Klaster K-Means (%)....	174
Tabel 4.44 Komposisi Sumber Air Minum Utama per Klaster HDBSCAN (%)	177
Tabel 4.45 Komposisi Sumber Air Minum Utama per Klaster SOM (%).....	179
Tabel 4.46 Ringkasan Profil Klaster K-Means	182
Tabel 4.47 Ringkasan Profil Klaster HDBSCAN	183
Tabel 4.48 Ringkasan Profil Klaster SOM	184
Tabel 4.49 Matriks Diferensiasi Layanan dan Pembelajaran Berdasarkan Profil Klaster K-Means	186



DAFTAR GAMBAR

Gambar 2.1 Ilustasi Pendekatan <i>Clustering</i>	31
Gambar 2.2 Ilustasi <i>K-Means</i>	33
Gambar 2.3 Ilustasi Pendekatan HDBSCAN	37
Gambar 2.4 Ilustasi Pendekatan SOM	39
Gambar 3.1 Alur Penelitian Komparatif <i>K-Means</i> , HDBSCAN, dan SOM	65
Gambar 4.1 Penentuan K Optimal: <i>Elbow Method</i> , Jarak Tegak Lurus, dan <i>Silhouette Score</i>	93
Gambar 4.2 Hasil Klastering <i>K-Means</i> : PCA 2D, Sebaran Nilai Total vs Desil, dan Distribusi Klaster	94
Gambar 4.3 <i>Heatmap</i> Profil Rata-rata Klaster K-Means	94
Gambar 4.4 HDBSCAN <i>Condensed Tree</i> : Hierarki Pembentukan Klaster	96
Gambar 4.5 Hasil Klastering HDBSCAN: PCA 2D, Sebaran Nilai Total vs Desil, dan Distribusi Klaster	96
Gambar 4.6 Analisis U-Matrix SOM: U-Matrix Raw, Batas <i>Threshold</i> , dan Metrik Kualitas	98
Gambar 4.7 Penentuan K Optimal Klaster Neuron SOM (<i>Elbow Method</i> + <i>Silhouette</i>)	98
Gambar 4.8 Hasil Klastering SOM: Neuron Map, PCA 2D, dan Sebaran Nilai Total vs Desil	99
Gambar 4.9 SOM <i>Feature Maps</i> : Distribusi Bobot Neuron per Fitur	99
Gambar 4.10 Perbandingan <i>Silhouette Score</i> dan <i>Davies-Bouldin Index</i> Ketiga Algoritma	104
Gambar 4.11 <i>Silhouette Plot</i> per Algoritma Klastering	105
Gambar 4.12 Radar <i>Chart</i> Komparasi Algoritma (Normalisasi 0–1)	106
Gambar 4.13 Perbandingan Hasil Klastering Ketiga Algoritma pada Ruang PCA 2D	107
Gambar 4.14 Distribusi Desil Provinsi per Klaster pada Ketiga Algoritma	107
Gambar 4.15 Proporsi Anggota Klaster per Algoritma	108
Gambar 4.16 Rata-rata Nilai Total per Klaster (K-Means)	115
Gambar 4.17 Rata-rata Nilai Total per Klaster (HDBSCAN)	117
Gambar 4.18 Rata-rata Nilai Total per Klaster (SOM)	119
Gambar 4.19 Rata-rata Desil Provinsi per Klaster (K-Means)	120
Gambar 4.20 Rata-rata Desil Provinsi per Klaster (HDBSCAN)	122
Gambar 4.21 Rata-rata Desil Provinsi per Klaster (SOM)	124
Gambar 4.22 Komposisi Desil Provinsi per Klaster (K-Means)	126
Gambar 4.23 Komposisi Desil Provinsi per Klaster (HDBSCAN)	128
Gambar 4.24 Komposisi Desil Provinsi per Klaster (SOM)	130
Gambar 4.25 Komposisi Jenjang Pendidikan per Klaster (K-Means)	132
Gambar 4.26 Komposisi Jenjang Pendidikan per Klaster (HDBSCAN)	133
Gambar 4.27 Komposisi Jenjang Pendidikan per Klaster (SOM)	135
Gambar 4.28 Komposisi Status Sekolah per Klaster (K-Means)	137
Gambar 4.29 Komposisi Status Sekolah per Klaster (HDBSCAN)	138
Gambar 4.30 Komposisi Status Sekolah per Klaster (SOM)	140
Gambar 4.31 Komposisi Kabupaten/Kota per Klaster (K-Means)	142

Gambar 4.32 Komposisi Kabupaten/Kota per Klaster (HDBSCAN).....	144
Gambar 4.33 Komposisi Kabupaten/Kota per Klaster (SOM)	146
Gambar 4.34 Komposisi Klaster per Kabupaten/Kota (K-Means)	148
Gambar 4.35 Komposisi Klaster per Kabupaten/Kota (HDBSCAN).....	150
Gambar 4.36 Komposisi Klaster per Kabupaten/Kota (SOM)	152
Gambar 4.37 Komposisi Jenis Lantai Terluas per Klaster (K-Means)	157
Gambar 4.38 Komposisi Jenis Lantai Terluas per Klaster (HDBSCAN).....	159
Gambar 4.39 Komposisi Jenis Lantai Terluas per Klaster (SOM)	161
Gambar 4.40 Komposisi Daya Listrik Terpasang per Klaster (K-Means).....	163
Gambar 4.41 Komposisi Daya Listrik Terpasang per Klaster (HDBSCAN)	165
Gambar 4.42 Komposisi Daya Listrik Terpasang per Klaster (SOM).....	167
Gambar 4.43 Komposisi Bahan Bakar Utama Memasak per Klaster (K-Means)	169
Gambar 4.44 Komposisi Bahan Bakar Utama Memasak per Klaster (HDBSCAN)	171
Gambar 4.45 Komposisi Bahan Bakar Utama Memasak per Klaster (SOM).....	173
Gambar 4.46 Komposisi Sumber Air Minum Utama per Klaster (K-Means)	176
Gambar 4.47 Komposisi Sumber Air Minum Utama per Klaster (HDBSCAN)	178
Gambar 4.48 Komposisi Sumber Air Minum Utama per Klaster (SOM)	181



DAFTAR LAMPIRAN

Lampiran 1 Surat Permohonan Ijin Penelitian	202
Lampiran 2 Surat Balasan Rekomendasi Penelitian	203
Lampiran 3 Surat Keterangan Telah Melaksanakan Penelitian	204
Lampiran 4 Tampilan Dashboard Tes Kemampuan Akademik.....	205
Lampiran 5 Tampilan Aplikasi Data Pokok Pendidikan (DAPODIK).....	206
Lampiran 6 Tampilan Aplikasi Cek Bansos	207
Lampiran 7 Riwayat Hidup Penulis	208

