

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN	iii
PRAKATA	iv
DAFTAR ISI	vi
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
INTISARI	xii
ABSTRACT	xiii
I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Batasan Masalah	3
1.3. Tujuan Penelitian	4
1.4. Tinjauan Pustaka	5
1.5. Metodologi Penelitian	7
1.6. Sistematika Penelitian	7
II DASAR TEORI	9
2.1. <i>Data Mining</i>	9
2.2. <i>Machine Learning</i>	10
2.3. Analisis Kluster	11
2.4. Indeks Validitas Kluster	14
2.5. Matriks Data Multivariat	16
2.6. Ukuran Pemusatan Data	16
2.6.1. Rata-Rata	16
2.6.2. Median	17
2.7. Ukuran Penyebaran Data	17
2.7.1. Jangkauan	17
2.7.2. Variansi	17
2.7.3. Standar deviasi	17
2.8. Kovariansi dan Korelasi	18
2.8.1. Kovariansi	18
2.8.2. Korelasi	18
2.9. Standardisasi	19

2.10. Matriks	19
2.11. Jarak Euclidean	23
III ALGORITMA OPTICS DENGAN TEKNIK EKSTRAKSI KLASSTER OTOMATIS	24
3.1. Analisis Komponen Utama	24
3.1.1. Korelasi Antar Fitur	26
3.1.2. Uji Bartlett	27
3.1.3. Uji KMO	27
3.2. <i>Density-Based Spatial Clustering of Applications with Noise</i>	29
3.3. <i>Ordering-Points to Identify Clustering Structures</i>	31
3.3.1. Plot Keterjangkauan	33
3.3.2. Ekstraksi Kluster dengan Teknik Otomatis	46
3.3.3. Ekstraksi Kluster dengan Pendekatan DBSCAN	51
3.4. Metrik Evaluasi CDbw	52
3.5. Diagram Alur Penelitian	56
IV STUDI KASUS	59
4.1. Deskripsi Data	59
4.1.1. Dataset <i>Zero Hunger</i>	59
4.1.2. Dataset <i>Mall Customers</i>	63
4.1.3. Dataset <i>Seeds</i>	64
4.1.4. Dataset <i>Blobs</i>	64
4.2. Eksplorasi Data	65
4.2.1. Duplikasi Data	65
4.2.2. <i>Missing Values</i>	65
4.2.3. Penskalaan fitur	67
4.2.4. Distribusi Data	68
4.2.5. Korelasi antar Variabel	69
4.3. Analisis Komponen Utama	71
4.3.1. Uji Bartlett	72
4.3.2. Uji KMO	72
4.3.3. Analisis Komponen Utama	73
4.4. OPTICS	76
4.4.1. Parameter dan Atribut OPTICS	76
4.4.2. Plot Keterjangkauan	79
4.4.3. Ekstraksi Kluster dengan Teknik Otomatis	82
4.4.4. Ekstraksi Kluster dengan Pendekatan DBSCAN	89

4.5. DBSCAN	93
4.6. Evaluasi Model	95
4.7. Perbandingan Metode	96
4.8. Visualisasi Anggota Kluster	97
4.8.1. Dataset <i>Zero Hunger</i>	97
4.8.2. Dataset <i>Mall Customers</i>	103
4.8.3. Dataset <i>Seeds</i>	104
4.8.4. Dataset <i>Blobs</i>	105
V PENUTUP	106
5.1. Kesimpulan	106
5.2. Saran	107
DAFTAR PUSTAKA	108
A Syntax Analisis Data	111
B Data <i>Zero Hunger</i>	153
C Data <i>Blobs</i>	168

DAFTAR TABEL

2.1	Tabel data multivariat	16
3.1	Data visualisasi penyusunan plot keterjangkauan	35
3.2	Keterjangkauan titik A	36
3.3	Jarak inti titik A	37
3.4	<i>SeedList</i> pada iterasi pertama	37
3.5	Keterjangkauan titik A dan D	37
3.6	Jarak inti titik A dan D	38
3.7	<i>SeedList</i> pada iterasi kedua	39
3.8	Keterjangkauan titik A, D, dan C	39
3.9	Jarak inti titik A, D, dan C	39
3.10	<i>SeedList</i> pada iterasi kedua	40
3.11	Hasil Iterasi ke-11	42
4.1	Sumber data	62
4.2	Variabel pada dataset <i>Mall Customers</i>	63
4.3	Variabel pada dataset <i>Seeds</i>	64
4.4	Variabel pada dataset <i>Blobs</i>	64
4.5	Data duplikat	65
4.6	<i>Missing values</i> keseluruhan	66
4.7	<i>Missing values</i> berdasarkan <i>Region</i>	67
4.8	Uji Bartlett	72
4.9	Uji KMO	73
4.10	Analisis komponen utama untuk dataset <i>Zero Hunger</i>	73
4.11	Analisis komponen utama untuk dataset <i>Seeds</i>	74
4.14	Struktur hierarki kluster	84
4.15	Evaluasi dari beberapa level hierarki kluster	87
4.16	Komponen indeks CDbw	95
4.17	Negara-negara dalam kluster 0	98
4.18	Negara-negara dalam kluster 1	100
4.19	Negara-negara dalam kluster 2	101
4.20	Negara-negara dalam kluster 3	102
4.21	Negara-negara dalam kluster 3	102

DAFTAR GAMBAR

3.1	<i>Kondisi titik inti dan density-reachable</i> langsung [Ester <i>dkk.</i> , 1996] .	29
3.2	<i>Density-reachable</i> dan <i>density-connected</i> [Ester <i>dkk.</i> , 1996]	30
3.3	$core(o)$, $r(p_1, o)$ dan $r(p_2, o)$ untuk $MinPts = 4$ [Ankerst <i>dkk.</i> , 1999] .	33
3.4	Ilustrasi plot keterjangkauan [Ankerst <i>dkk.</i> , 1999]	34
3.5	<i>Scatterplot</i> data visualisasi penyusunan plot keterjangkauan	35
3.6	Visualisasi iterasi pertama	36
3.7	Visualisasi iterasi kedua	38
3.8	Visualisasi iterasi ketiga	40
3.9	Visualisasi iterasi ke-11	41
3.10	Plot keterjangkauan	42
3.11	Hasil simulasi menggunakan titik insiasi berbeda	44
3.12	Hasil simulasi dari algoritma modifikasi menggunakan titik insiasi berbeda	45
3.13	Ilustrasi sebuah kluster dalam plot keterjangkauan [Ankerst <i>dkk.</i> , 1999]	46
3.14	Ilustrasi kluster dalam data nyata [Ankerst <i>dkk.</i> , 1999]	47
3.15	Ilustrasi penentuan titik awal dan akhir kluster [Ankerst <i>dkk.</i> , 1999] .	49
3.16	Diagram alur penelitian	58
4.1	<i>Box plot</i> data <i>Zero Hunger</i> hasil standarisasi	68
4.2	<i>Box plot</i> data <i>Mall Customers</i> hasil standarisasi	68
4.3	<i>Box plot</i> data <i>Seeds</i> hasil standarisasi	69
4.4	<i>Box plot</i> data <i>Blobs</i> hasil standarisasi	69
4.5	Korelasi antar variabel data <i>Zero Hunger</i>	70
4.6	Korelasi antar variabel data <i>Mall Customers</i>	70
4.7	Korelasi antar variabel data <i>Seeds</i>	71
4.8	<i>Scree plot</i> nilai eigen komponen utama data <i>Zero Hunger</i>	75
4.9	<i>Scree plot</i> nilai eigen komponen utama data <i>Seeds</i>	75
4.10	Pengaruh parameter ϵ terhadap plot keterjangkauan	80
4.11	Pengaruh parameter $MinPts$ terhadap plot keterjangkauan	81
4.12	Plot keterjangkauan Model 1 data <i>Zero Hunger</i>	84
4.13	Hasil klastering Model 1 data <i>Zero Hunger</i>	84
4.14	Plot keterjangkauan Model 1 dengan level hierarki 0	85

4.15	Plot keterjangkauan Model 1 dengan level hierarki 1	85
4.16	Plot keterjangkauan Model 1 dengan level hierarki 2	85
4.17	Plot keterjangkauan Model 1 dengan level hierarki 3	86
4.18	Plot keterjangkauan Model 1 dengan level hierarki 4	86
4.19	Plot keterjangkauan Model 1 dengan level hierarki 5	86
4.20	Plot keterjangkauan Model 1 data <i>Mall Customers</i>	88
4.21	Plot keterjangkauan Model 1 data <i>Seeds</i>	88
4.22	Plot keterjangkauan Model 1 data <i>Blobs</i>	89
4.23	Plot keterjangkauan Model 2 data <i>Zero Hunger</i>	90
4.24	Pengaruh parameter <i>eps</i> terhadap plot keterjangkauan	91
4.25	Plot keterjangkauan Model 2 data <i>Mall Customers</i>	92
4.26	Plot keterjangkauan Model 2 data <i>Seeds</i>	93
4.27	Plot keterjangkauan Model 2 data <i>Blobs</i>	93
4.28	Visualisasi hasil klastering terbaik untuk data <i>Zero Hunger</i>	97
4.29	Visualisasi hasil klastering Model 1	103
4.30	Visualisasi hasil klastering terbaik untuk data <i>Mall Customers</i> . . .	104
4.31	Visualisasi hasil klastering terbaik untuk data <i>Seeds</i>	104
4.32	Visualisasi hasil klastering terbaik untuk data <i>Blobs</i>	105