

**IMPLEMENTASI K-MODES PADA CLUSTERING DATA
KATEGORI MENGGUNAKAN NEW DISSIMILARITY
MEASURE**

SKRIPSI

KONSENTRASI KOMPUTASI CERDAS DAN VISUALISASI

**Untuk memenuhi sebagian persyaratan untuk
Mencapai gelar Sarjana Komputer**



Disusun Oleh:

R. YISKA DEVIARANI SUWARSA

0610963045

**KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
PROGRAM TEKNOLOGI INFORMASI DAN ILMU KOMPUTER
MALANG
2013**

DAFTAR ISI

KATA PENGANTAR.....	i
ABSTRAK.....	iii
DAFTAR ISI.....	v
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Penelitian.....	2
1.4 Batasan Masalah.....	3
1.5 Manfaat Penelitian.....	3
1.6 Metodologi Penelitian.....	3
1.7 Sistematika Penulisan.....	4
BAB II KAJIAN PUSTAKA DAN DASAR TEORI.....	5
2.1 Kajian Pustaka.....	5
2.2 Data.....	5
2.2.1 Data Numerik.....	7
2.2.2 Data Kategori.....	7
2.3 Data Mining.....	8
2.3.1 Pengertian <i>Data Mining</i>	8
2.3.2 Proses <i>Data Mining</i>	9
2.3.3 Metode <i>Data Mining</i>	10
2.3.4 Fungsi pada <i>Data Mining</i>	10
2.3.4.1 Pengklasteran (<i>Clustering</i>).....	10
2.4 Algoritma <i>K-Modes</i>	12
2.5 <i>Similarity Measure, Dissimilarity Measure dan New Dissimilarity Measure</i>	13
2.5.1 <i>Similarity Measure</i>	13
2.5.2 <i>Dissimilarity Measure</i>	14
2.5.3 <i>New Dissimilarity Measure</i>	16

2.6 Purity Measure.....	16
2.7 Pengujian Accuracy.....	18
2.8 Pengujian F-Measure.....	18

BAB III METODOLOGI DAN PERANCANGAN SISTEM..... 22

3.1 Deskripsi Data.....	22
3.2 Analisis Perangkat Lunak.....	22
3.2.1 Deskripsi Umum Perangkat Lunak.....	22
3.2.2 Batasan Perangkat Lunak.....	23
3.3 Perancangan Perangkat Lunak.....	23
3.3.1 Perancangan Proses.....	23
3.4 Perancangan Antarmuka.....	32
3.5 Perancangan Uji Coba	33
3.6 Hitung Manual	33
3.6.1 Hitung K-Modes New Dissimilarity Measure.....	36
3.6.2 Hitung Purity.....	48
3.6.3 Contoh Perhitungan Metode Evaluasi F-Measure.....	49

BAB IV IMPLEMENTASI..... 53

4.1 Lingkungan Implementasi.....	53
4.2.1 Lingkungan Perangkat Keras.....	53
4.2.2 Lingkungan Perangkat Lunak	53
4.2 Implementasi Program.....	53
4.2.1 Implementasi Clustering.....	53
4.2.2 Penentuan Pusat Cluster Awal.....	54
4.2.3 Penentuan Jarak.....	54
4.2.4 Penentuan Purity Measure.....	64
4.2.5 Evaluasi F-Measure.....	66
4.3 Implementasi Antarmuka.....	68

BAB V ANALISA DAN PEMBAHASAN..... 71

5.1 Pengujian.....	71
--------------------	----

5.2 Hasil Uji.....	71
5.3 Pengujian untuk mengetahui Nilai <i>Purity</i> pada <i>K-Modes</i> yang menggunakan <i>new dissimilarity measure</i> dan <i>K-Modes</i> Konvensional.....	71
1. <i>Balance Scale</i>	72
2. <i>Car Evaluation</i>	78
3. <i>Lenses</i>	86
4. <i>Tic Tac Toe</i>	89
5.3.1 Analisis hasil pengujian Nilai <i>Purity</i>	99
5.4 Pengujian untuk mengetahui <i>F-Measure</i> pada <i>K-Modes</i> yang menggunakan <i>new dissimilarity measure</i> dan <i>K-Modes</i> Konvensional.....	100
1. Evaluasi <i>F-Measure</i> pada dataset <i>Balance Scale</i>	100
2. Evaluasi <i>F-Measure</i> pada dataset <i>Car Evaluation</i>	105
3. Evaluasi <i>F-Measure</i> pada dataset <i>Lenses</i>	111
4. Evaluasi <i>F-Measure</i> pada dataset <i>Tic Tac Toe</i>	114
Analisa hasil pengklasteran metode <i>K-Modes</i>	120
a. Analisa pola nilai <i>precision</i> , <i>recall</i> , dan <i>F-Measure</i>	120
b. Analisa banyak <i>class</i> dalam <i>dataset</i> terhadap nilai <i>precision</i> , <i>recall</i> , dan <i>F-Measure</i>	121
BAB VI KESIMPULAN DAN SARAN	122
6.1 Kesimpulan.....	122
6.2 Saran.....	122
DAFTAR PUSTAKA	124

DAFTAR GAMBAR

Gambar 2.1	Proses <i>Knowledge discovery in databases</i> (KDD) (Usama Fayyad, 1996).....	5
Gambar 2.2	Proses-proses dalam <i>data mining</i> (Kantardzic, 2003).....	9
Gambar 2.3	Model dan Teknik <i>Data Mining</i>	21
Gambar 3.1	Tahapan-tahapan Penelitian.....	21
Gambar 3.2	Flowchart Proses Perangkat Lunak <i>K-Modes New Dissimilarity Measure</i> menggunakan evaluasi <i>Purity Measure</i>	25
Gambar 3.3	Flowchart Proses Perangkat Lunak <i>K-Modes New Dissimilarity Measure</i> menggunakan evaluasi <i>F-Measure</i>	26
Gambar 3.4	Flowchart Proses Perangkat Lunak <i>K-Modes Konvensional</i> menggunakan evaluasi <i>Purity Measure</i>	26
Gambar 3.5	Flowchart Proses Perangkat Lunak <i>K-Modes Konvensional</i> menggunakan evaluasi <i>F-Measure</i>	27
Gambar 3.6	Flowchart Algoritma <i>Clustering K-Modes New Dissimilarity Measure</i>	28
Gambar 3.7	Flowchart Algoritma <i>Clustering K-Modes Konvensional</i>	29
Gambar 3.8	Flowchart Hitung <i>Purity Measure</i>	30
Gambar 3.9	Flowchart Evaluasi <i>F-Measure</i>	31
Gambar 3.10	Rancangan Antarmuka.....	32
Gambar 3.11	<i>Cluster</i> dan anggota <i>cluster</i>	50
Gambar 4.1	Antarmuka system <i>K-Modes Clustering</i>	69
Gambar 4.2	Informasi Antarmuka system <i>K-Modes Clustering</i>	69
Gambar 5.1	Data <i>Balance Scale</i>	72
Gambar 5.2	Grafik Hasil <i>Purity</i> Percobaan <i>New K-Modes</i> dan <i>K-Modes Konvensional</i> pada <i>dataset Balance Scale</i>	77
Gambar 5.3	Grafik Hasil <i>Error</i> Percobaan <i>New K-Modes</i> dan <i>K-Modes Konvensional</i> pada <i>dataset Balance Scale</i>	78
Gambar 5.4	Data <i>Car Evaluation</i>	79
Gambar 5.5	Grafik Hasil <i>Purity</i> Percobaan <i>New K-Modes</i> dan <i>K-Modes Konvensional</i> pada <i>dataset Car Evaluation</i>	84

Gambar 5.6	Grafik Hasil <i>Error</i> Percobaan <i>New K-Modes</i> dan <i>K-Modes</i> Konvensional pada <i>dataset Car Evaluation</i>	85
Gambar 5.7	Data <i>Lenses</i>	86
Gambar 5.8	Grafik Hasil <i>Purity</i> Percobaan <i>New K-Modes</i> dan <i>K-Modes</i> Konvensional pada <i>dataset Lenses</i>	88
Gambar 5.9	Grafik Hasil <i>Erro</i> Percobaan <i>New K-Modes</i> dan <i>K-Modes</i> Konvensional pada <i>dataset Lenses</i>	89
Gambar 5.10	Data <i>Tic Tac Toe</i>	90
Gambar 5.11	Grafik Hasil <i>Purity</i> Percobaan <i>New K-Modes</i> dan <i>K-Modes</i> Konvensional pada <i>dataset Tic Tac Toe</i>	95
Gambar 5.12	Grafik Hasil <i>Erro</i> Percobaan <i>New K-Modes</i> dan <i>K-Modes</i> Konvensional pada <i>dataset Tic Tac Toe</i>	96
Gambar 5.13	Grafik Nilai <i>Purity</i> pada pengujian <i>Clustering</i> pada Data Kategori.....	97
Gambar 5.14	Grafik Nilai <i>Error</i> pada pengujian <i>Clustering</i> pada Data Kategori.....	98
Gambar 5.15	Grafik Evaluasi <i>F-Measure</i> pada Data <i>Balance Scale</i> menggunakan <i>New K-Modes</i> dan <i>K-Modes</i> Konvensional.....	105
Gambar 5.16	Grafik Evaluasi <i>F-Measure</i> pada Data <i>Car Evaluation</i> menggunakan <i>New K-Modes</i> dan <i>K-Modes</i> Konvensional.....	111
Gambar 5.17	Evaluasi <i>F-Measure</i> pada Data <i>Lenses</i> menggunakan <i>New K-Modes</i> dan <i>K-Modes</i> Konvensional.....	113
Gambar 5.18	Evaluasi <i>F-Measure</i> pada Data <i>Tic Tac Toe</i> menggunakan <i>New K-Modes</i> dan <i>K-Modes</i> Konvensional.....	119
Gambar 5.19	Evaluasi <i>F-Measure</i> pada Data Kategori pada <i>K-Modes New Dissimilarity</i>	120

DAFTAR TABEL

Tabel 2.1	Contoh Hasil <i>Clustering</i>	17
Tabel 2.2	Matriks <i>Confusion</i>	18
Tabel 3.1	Deskripsi <i>Dataset</i>	22
Tabel 3.2	Uji Coba.....	33
Tabel 3.3	<i>Dataset Lenses</i>	33
Tabel 3.4	<i>Dataset Lenses</i> yang telah dipilih <i>random</i>	35
Tabel 3.5	Pusat <i>Cluster</i>	36
Tabel 3.6	Jarak dan <i>Cluster</i> (Iterasi-1).....	39
Tabel 3.7	Hasil <i>Cluster</i> Semua <i>Record</i>	40
Tabel 3.8	Pusat <i>Cluster</i> yang Baru.....	40
Tabel 3.9	Jarak dan <i>Cluster</i> (Iterasi-2).....	43
Tabel 3.10	Hasil <i>Cluster</i> Semua <i>Record</i>	44
Tabel 3.11	Pusat <i>Cluster</i> Baru.....	44
Tabel 3.12	Jarak dan <i>Cluster</i> (Iterasi-3).....	47
Tabel 3.13	Hasil <i>Cluster</i> Semua <i>Record</i>	47
Tabel 3.14	Pusat <i>Cluster</i> yang Baru.....	47
Tabel 3.15	<i>Dataset Lenses</i>	48
Tabel 3.16	Hasil <i>Cluster</i> Semua <i>Record</i>	48
Tabel 3.17	Hasil <i>Clustering</i>	49
Tabel 3.18	<i>Cluster</i> dan <i>Class</i> per <i>record</i>	50
Tabel 3.19	<i>Precision</i> , <i>Recall</i> , dan <i>F-Measure</i>	52
Tabel 5.1	<i>Dataset</i> yang digunakan.....	71
Tabel 5.2	Hasil <i>Purity</i> Percobaan <i>New K-Modes</i> pada <i>dataset Balance Scale</i>	73
Tabel 5.3	Hasil <i>Purity</i> dan <i>Error</i> Percobaan <i>New K-Modes</i> pada <i>dataset Balance Scale</i>	74
Tabel 5.4	Hasil <i>Purity</i> Percobaan <i>K-Modes Konvensional</i> pada <i>dataset Balance Scale</i>	75
Tabel 5.5	Hasil <i>Purity</i> dan <i>Error</i> Percobaan <i>K-Modes Konvensional</i> pada <i>dataset Balance Scale</i>	76

Tabel 5.6	Hasil <i>Purity</i> Percobaan <i>New K-Modes</i> dan <i>K-Modes</i> Konvensional pada <i>dataset Balance Scale</i>	76
Tabel 5.7	Hasil <i>Error</i> Percobaan <i>New K-Modes</i> dan <i>K-Modes</i> Konvensional pada <i>dataset Balance Scale</i>	77
Tabel 5.8	Hasil <i>Purity</i> Percobaan <i>New K-Modes</i> pada <i>dataset Car Evaluation</i>	80
Tabel 5.9	Hasil <i>Purity</i> dan <i>Error</i> Percobaan <i>New K-Modes</i> pada <i>dataset Car Evaluation</i>	81
Tabel 5.10	Hasil <i>Purity</i> Percobaan <i>K-Modes</i> Konvensional pada <i>dataset Car Evaluation</i>	82
Tabel 5.11	Hasil <i>Purity</i> dan <i>Error</i> Percobaan <i>K-Modes</i> Konvensional pada <i>dataset Car Evaluation</i>	83
Tabel 5.12	Hasil <i>Purity</i> Percobaan <i>New K-Modes</i> dan <i>K-Modes</i> Konvensional pada <i>dataset Car Evaluation</i>	83
Tabel 5.13	Hasil <i>Error</i> Percobaan <i>New K-Modes</i> dan <i>K-Modes</i> Konvensional pada <i>dataset Car Evaluation</i>	85
Tabel 5.14	Hasil <i>Purity</i> Percobaan <i>New K-Modes</i> dan <i>K-Modes</i> Konvensional pada <i>dataset Lenses</i>	87
Tabel 5.15	Hasil <i>Purity</i> Percobaan <i>New K-Modes</i> pada <i>dataset Tic Tac Toe</i>	91
Tabel 5.16	Hasil <i>Purity</i> dan <i>Error</i> Percobaan <i>New K-Modes</i> pada <i>dataset Tic Tac Toe</i>	92
Tabel 5.17	Hasil <i>Purity</i> Percobaan <i>K-Modes</i> Konvensional pada <i>dataset Tic Tac Toe</i>	93
Tabel 5.18	Hasil <i>Purity</i> dan <i>Error</i> Percobaan <i>K-Modes</i> Konvensional pada <i>dataset Tic Tac Toe</i>	94
Tabel 5.19	Hasil <i>Purity</i> Percobaan <i>New K-Modes</i> dan <i>K-Modes</i> Konvensional pada <i>dataset Tic Tac Toe</i>	94
Tabel 5.20	Hasil <i>Error</i> Percobaan <i>New K-Modes</i> dan <i>K-Modes</i> Konvensional pada <i>dataset Tic Tac Toe</i>	95
Tabel 5.21	Nilai <i>Purity</i> pada pengujian <i>Clustering</i> pada Data Kategori....	97
Tabel 5.22	<i>Error</i> pada pengujian <i>Clustering</i> pada Data Kategori.....	98

Tabel 5.23	Pengujian <i>Clustering</i> pada Data Kategori.....	99
Tabel 5.24	Pengujian <i>Precision</i> pada Data <i>Balance Scale New K-Modes</i> ..	100
Tabel 5.25	Pengujian <i>Recall</i> pada Data <i>Balance Scale New K-Modes</i>	100
Tabel 5.26	Pengujian <i>F-Measure</i> pada Data <i>Balance Scale New K-Modes</i> .	101
Tabel 5.27	Pengujian <i>F-Measure</i> pada Data <i>Balance Scale New K-Modes</i> .	101
Tabel 5.28	Pengujian <i>Precision</i> pada Data <i>Balance Scale K-Modes</i> Konvensional.....	102
Tabel 5.29	Pengujian <i>Recall</i> pada Data <i>Balance Scale K-Modes</i> Konvensional.....	102
Tabel 5.30	Pengujian <i>F-Measure</i> pada Data <i>Balance Scale K-Modes</i> Konvensional.....	103
Tabel 5.31	Pengujian Rata-rata <i>F-Measure</i> pada Data <i>Balance Scale</i> <i>K-Modes Konvensional</i>	104
Tabel 5.32	Evaluasi <i>F-Measure</i> pada Data <i>Balance Scale</i> menggunakan <i>New K-Modes dan K-Modes Konvensional</i>	104
Tabel 5.33	Pengujian <i>Precision</i> pada Data <i>Car Evaluation Scale New</i> <i>K-Modes</i>	105
Tabel 5.34	Pengujian <i>Recall</i> pada Data <i>Car Evaluation Scale New K-Modes</i>	106
Tabel 5.35	Pengujian <i>F-Measure</i> pada Data <i>Car Evaluation Scale New</i> <i>K-Modes</i>	106
Tabel 5.36	Pengujian Rata-rata <i>F-Measure</i> pada Data <i>Car Evaluation New</i> <i>K-Modes</i>	107
Tabel 5.37	Pengujian <i>Precision</i> pada Data <i>Car Evaluation K-Modes</i>	108
Tabel 5.38	Pengujian <i>Recall</i> pada Data <i>Car Evaluation K-Modes</i> Konvensional.....	108
Tabel 5.39	Pengujian <i>F-Measure</i> pada Data <i>Car Evaluation K-Modes</i> Konvensional.....	109
Tabel 5.40	Pengujian Rata-rata <i>F-Measure</i> pada Data <i>Car Evaluation K-</i> <i>Modes Konvensional</i>	109
Tabel 5.41	Evaluasi <i>F-Measure</i> pada Data <i>Car Evaluation</i> menggunakan <i>New K-Modes dan K-Modes Konvensional</i>	110

Tabel 5.42	Pengujian <i>F-Measure</i> pada Data <i>Lenses New K-Modes</i>	111
Tabel 5.43	Pengujian <i>F-Measure</i> pada Data <i>Lenses K-Modes Konvensional</i>	112
Tabel 5.44	Evaluasi <i>F-Measure</i> pada Data <i>Lenses</i> menggunakan <i>New K-Modes dan K-Modes Konvensional</i>	112
Tabel 5.45	Pengujian <i>Precision</i> pada Data <i>Tic Tac Toe New K-Modes</i>	114
Tabel 5.46	Pengujian <i>Recall</i> pada Data <i>Tic Tac Toe New K-Modes</i>	114
Tabel 5.47	Pengujian <i>F-Measure</i> pada Data <i>Tic Tac Toe New K-Modes</i> ..	115
Tabel 5.48	Pengujian Rata-rata <i>F-Measure</i> pada Data <i>Tic Tac Toe New K-Modes</i>	115
Tabel 5.49	Pengujian <i>Precision</i> pada Data <i>Tic Tac Toe K-Modes Konvensional</i>	116
Tabel 5.50	Pengujian <i>Recall</i> pada Data <i>Tic Tac Toe K-Modes Konvensional</i>	116
Tabel 5.51	Pengujian <i>F-Measure</i> pada Data <i>Tic Tac Toe K-Modes Konvensional</i>	117
Tabel 5.52	Pengujian Rata-rata <i>F-Measure</i> pada Data <i>Tic Tac Toe K-Modes Konvensional</i>	117
Tabel 5.53	Evaluasi <i>F-Measure</i> pada Data <i>Tic Tac Toe</i> menggunakan <i>New K-Modes dan K-Modes Konvensional</i>	118
Tabel 5.54	Pengujian Rata-rata <i>F-Measure</i> pada Data <i>Kategori</i> pada <i>K-Modes New Dissimilarity Measure</i>	119

DAFTAR SOURCE CODE

Source Code 4.1 Menentukan Pusat <i>Cluster</i> Awal.....	54
Source Code 4.2 Menentukan Jarak (<i>K-Modes</i>).....	55
Source Code 4.3 Menghitung Hasil <i>Purity Measure</i>	65
Source Code 4.4 Evaluasi <i>F-Measure</i>	66

