



DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN	iii
HALAMAN PERSEMBAHAN	iv
HALAMAN MOTTO	v
PRAKATA	vi
DAFTAR ISI	ix
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xv
DAFTAR LAMBANG	xvi
INTISARI	xx
ABSTRACT	xxii
I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	6
1.3 Batasan Masalah	6
1.4 Tujuan dan Manfaat Penelitian	6
1.4.1 Tujuan Penelitian	6
1.4.2 Manfaat Penelitian	7
1.5 Tinjauan Pustaka	7
1.6 Metodologi Penelitian	13
1.7 Sistematika Penulisan	13
II DASAR TEORI	15
2.1 Data Pencilan (<i>Outliers</i>)	15
2.2 Deteksi <i>Outliers</i>	15
2.3 Statistik <i>Robust</i> atau <i>Robust Statistics</i>	17
2.4 Ukuran Jarak Standar	17
2.4.1 Jarak Euclidean	18
2.4.2 Jarak Euclidean Terdistandardisasi	18
2.4.3 Jarak Manhattan	18
2.4.4 Jarak Mahalanobis	18
2.4.5 Jarak Minkowski	19
2.4.6 Jarak Chebyshev	19
2.4.7 Jarak Canberra	20



2.5	Metode Klasterisasi (<i>Clustering Method</i>)	20
2.5.1	Metode <i>K-Means Clustering</i>	24
2.5.2	Metode <i>Trimmed K-Means Clustering</i>	25
2.5.3	Metode <i>Robust Sparse K-Means Clustering</i>	26
2.6	Validitas Klaster	27
2.6.1	Indeks <i>Silhouette</i> - max	28
2.6.2	Indeks BH (Ball-Hall) - max	29
2.6.3	Indeks CH (Calinski-Harabasz) - max	29
2.6.4	Indeks Dunn - max	30
2.6.5	Indeks DB (Davies-Bouldin) - min	31
2.6.6	Indeks <i>Determinat-Ratio</i> - max	32
2.6.7	Indeks PC (<i>Partition Coefficient</i>) - max	33
2.6.8	Indeks FS (Fukuyama dan Sugeno) - min	33
2.6.9	Indeks XB (Xie-Beni) - min	34
2.7	M-Estimator	35
2.7.1	Definisi M-Estimator	35
2.7.2	Fungsi Kerugian dalam M-Estimator	35
2.8	Estimasi Skala dengan <i>Median Absolute Deviation</i> (MAD)	37
2.9	Estimasi <i>Fast Minimum Covariance Determinant</i> (FMCD)	37
2.10	Analisis Kompleksitas Algoritma	40
2.10.1	Kompleksitas Waktu Asimtotik	41
2.10.2	Notasi Asimtotik <i>Big-O</i>	42

**III PENGEMBANGAN METODE *K-MEANS CLUSTERING* BERBASIS
JARAK *ROBUST* DAN PENGEMBANGAN INDEKS VALIDITAS KLAS-
TER YANG *ROBUST* UNTUK KLASTERISASI DATA MENGANDUNG
OUTLIERS 46**

3.1	Deteksi <i>Outlier</i> Multivariat	48
3.2	Metode <i>K-Means Clustering</i> menggunakan Jarak <i>Robust</i>	50
3.2.1	Ukuran Jarak <i>Robust</i>	50
3.2.2	Modifikasi Algoritma <i>K-Means Clustering</i> Menggunakan Jarak <i>Robust</i>	54
3.2.3	Kompleksitas Waktu Algoritma <i>K-means Clustering</i> dengan Jarak <i>Robust</i>	55
3.3	Indeks Validitas Klaster yang <i>Robust</i> pada <i>K-Means Clustering</i>	58
3.3.1	Indeks Validitas pada setiap titik data tunggal pada <i>hard clustering</i>	59
3.3.2	Indeks Validitas Berdasarkan <i>M-Estimator</i>	60



3.3.3	Indeks Validitas Tipe Rata-Rata berdasarkan <i>M-estimator</i> pada <i>K-Means Clustering</i>	61
3.3.4	Indeks Validitas Tipe Median, <i>Trimmed Means</i> , <i>Winsorized Means</i> , <i>Huber Means</i> , dan <i>MCD Means</i> pada <i>K-Means Clustering</i>	63
3.3.5	Algoritma Evaluasi <i>K-Means Clustering</i> dengan Indeks Validitas Kluster	64
IV KAJIAN APLIKASI METODE K-MEANS CLUSTERING BERBASIS JARAK ROBUST DAN INDEKS VALIDITAS KLASTER YANG ROBUST PADA DATA RIIL 66		
4.1	Aplikasi Deteksi <i>Outlier</i> Data Multivariat	66
4.1.1	Deteksi <i>Outlier</i> berdasarkan Jarak Mahalanobis	67
4.1.2	Deteksi <i>Outlier</i> Spasial berdasarkan Jarak Mahalanobis menggunakan Algoritma Rata-Rata (<i>Mean</i>)	68
4.2	Aplikasi Metode <i>K-Means Clustering</i> dan Variannya pada Data Gempa	71
4.2.1	Aplikasi Metode <i>K-Means Clustering</i>	72
4.2.2	Aplikasi Metode <i>Trimmed K-Means Clustering</i>	73
4.2.3	Aplikasi Metode <i>Robust Sparse K-Means Clustering</i>	75
4.2.4	Klaster Optimal Berdasarkan Nilai Indeks Internal Validitas	77
4.3	Aplikasi Metode <i>K-Means Clustering</i> menggunakan Jarak <i>robust</i> pada data gempa	79
4.3.1	Penentuan Klaster Optimal pada Data Gempa menggunakan Indeks Validitas Klaster Internal	79
4.3.2	Penerapan Metode <i>K-Means Clustering</i> menggunakan Jarak <i>robust</i> pada data gempa	83
4.4	Aplikasi untuk Menentukan Klaster Optimal dengan Menggunakan Metode Indeks Validitas Klaster yang <i>Robust</i>	88
4.4.1	Aplikasi pada Data <i>Iris</i>	89
4.4.2	Aplikasi pada Data <i>Wine</i>	93
4.4.3	Aplikasi pada Data <i>Breast Cancer</i>	98
4.4.4	Aplikasi pada Data <i>Pima Indian Diabetes</i>	101
4.4.5	Aplikasi pada Data <i>Glass Identification</i>	105
4.4.6	Aplikasi pada Data <i>Ecoli</i>	108
4.4.7	Aplikasi pada Data COVID	112
4.4.8	Rekapitulasi Metode Terbaik untuk Indeks Validitas Kluster	116
V KESIMPULAN DAN SARAN 118		
5.1	Kesimpulan	118



5.2	Saran	119
DAFTAR PUSTAKA	120