

DAFTAR ISI

HALAMAN PENGESAHAN	iii
PERNYATAAN.....	iv
PRAKATA	v
DAFTAR ISI.....	vii
DAFTAR GAMBAR.....	ix
DAFTAR TABEL	x
INTISARI	xi
ABSTRACT	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian	3
1.6 Metodologi Penelitian	3
1.7 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
BAB III LANDASAN TEORI.....	12
3.1 Metadata	12
3.1.1 Fungsi Metadata	12
3.1.2 Jenis Metadata	12
3.2 <i>Term Frequency – Inverse Document Frequency (TF-IDF)</i>	13
3.3 Algoritme <i>K-means</i>	15
3.3.1 Pemilihan Pusat Kluster Awal.....	15
3.3.2 Pemilihan Pusat Kluster Awal dengan Metode Probabilitas	16
3.3.3 Pemilihan Pusat Kluster Awal dengan Metode Densitas ..	16
3.3.4 Proses Algoritme <i>K-means</i>	18
3.3.5 Jarak Euclidean	19
3.3.6 Pembaruan Pusat Kluster.....	20
3.3.7 Kelebihan dan Kekurangan Algoritme <i>K-means</i>	20
3.4 Evaluasi Internal Kluster	21
BAB IV ANALISIS DAN RANCANGAN.....	23
4.1 Analisis Permasalahan	23
4.1.1 Definisi Permasalahan.....	23
4.1.2 Analisis Kebutuhan	23
4.1.3 Pendekatan Penyelesaian Masalah	24
4.2 Analisis Data	25
4.2.1 Rancangan Penyiapan <i>Dataset</i>	25
4.3 Rancangan Penelitian	27
4.4 Rancangan Representasi Kata	28
4.5 Rancangan Pemilihan Pusat Kluster.....	32
4.5.1 Pemilihan Pusat Kluster Awal dengan Metode Densitas ..	32

4.5.2	Pemilihan Pusat Kluster Awal dengan Metode Probabilitas.	35
4.6	Rancangan Proses <i>Clustering</i>	36
4.7	Rancangan Pengujian	37
BAB V	IMPLEMENTASI	39
5.1	Lingkungan Implementasi	39
5.2	Implementasi Ekstraksi Metadata	39
5.3	Implementasi Penyiapan Data	40
5.4	Implementasi Representasi Kata dengan TF-IDF	41
5.5	Implementasi Pemilihan Pusat Kluster Awal dengan Metode Densitas	42
5.6	Implementasi Pemilihan Pusat Kluster Awal dengan Metode Probabilitas	44
5.7	Implementasi <i>Clustering</i> dengan Algoritme <i>K-means</i>	45
5.7.1	Implementasi Jarak <i>Euclidean</i>	46
5.7.2	Implementasi Pembaruan Pusat Kluster	47
5.8	Implementasi <i>Sum Squared Error</i> (SSE)	47
BAB VI	HASIL DAN PEMBAHASAN	49
6.1	Pengujian dengan 200 Jumlah Dokumen	49
6.1.1	Hasil <i>Clustering</i> untuk 200 Jumlah Dokumen	50
6.2	Pengujian dengan 400 Jumlah Dokumen	51
6.2.1	Hasil <i>Clustering</i> untuk 400 Jumlah Dokumen	53
6.3	Pengujian dengan 800 Jumlah Dokumen	54
6.3.1	Hasil <i>Clustering</i> untuk 800 Jumlah Dokumen	55
6.4	Pengujian dengan 1600 Jumlah Dokumen	56
6.4.1	Hasil <i>Clustering</i> untuk 1600 Jumlah Dokumen	58
6.5	Perbandingan Pengujian Nilai SSE untuk Semua Jumlah Dokumen	59
6.6	Perbandingan Hasil Kluster untuk Semua Jumlah Dokumen	62
BAB VII	PENUTUP	64
7.1	Kesimpulan	64
7.2	Saran	64
	DAFTAR PUSTAKA	65
	LAMPIRAN	67