

## DAFTAR ISI

ABSTRAK .....	<b>Error! Bookmark not defined.</b>
KATA PENGANTAR .....	<b>Error! Bookmark not defined.</b>
DAFTAR ISI .....	i
DAFTAR TABEL .....	iii
DAFTAR GAMBAR .....	iv
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang Masalah .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah .....	3
1.4 Tujuan Penelitian .....	3
1.5 Manfaat Penelitian .....	4
1.6 Metode Penelitian .....	4
1.7 Sistematika Penelitian .....	5
BAB II TINJAUAN PUSTAKA .....	7
2.1 <i>Data Mining</i> .....	7
2.2 Analisis Cluster .....	10
2.2.1 Tipe data dalam analisis <i>cluster</i> .....	11
2.2.2 Klasifikasi Algoritma Clustering .....	12
2.3 Text Mining .....	15
2.4 Text Preprocessing .....	17
2.5 Transformasi Data ( <i>Feature Generation</i> ) .....	18
2.5.1 Pembobotan TF-IDF .....	18
2.5.2 Model Ruang Vektor .....	20
2.6 Algoritma K-Means .....	22
2.7 Evaluasi Clustering .....	25
2.8 Agregator Berita .....	26
BAB III METODOLOGI PENELITIAN .....	33
3.1 Desain Penelitian .....	33
3.2 Metode Penelitian .....	34

3.2.1	Eksplorasi dan studi literatur.....	34
3.2.2	Metode Pengembangan Perangkat Lunak.....	34
3.2.3	Eksperimen terhadap Algoritma .....	38
3.3	Alat dan Bahan Penelitian .....	38
3.3.1	Alat Penelitian.....	38
3.3.2	Bahan Penelitian.....	39
<b>BAB IV</b>	<b>HASIL PENELITIAN</b> .....	<b>40</b>
4.1	Pembahasan Penelitian.....	40
4.2	Pengembangan Perangkat Lunak .....	63
4.2.1	Analisis Sistem.....	63
4.2.2	Perancangan Sistem .....	66
4.2.3	Implementasi Coding .....	68
4.2.4	Uji Coba .....	73
4.3	Eksperimen.....	74
<b>BAB V</b>	<b>PENUTUP</b> .....	<b>80</b>
5.1	Kesimpulan.....	80
5.2	Saran .....	80
	<b>DAFTAR PUSTAKA</b> .....	<b>82</b>
	<b>LAMPIRAN</b> .....	<b>84</b>
Lampiran 1	.....	84
Lampiran 2	.....	90
Lampiran 3	.....	101
Lampiran 4	.....	117
Lampiran 5	.....	132
Lampiran 6	.....	146
	<b>RIWAYAT HIDUP</b> .....	<b>149</b>

## DAFTAR TABEL

Tabel 4.1 Daftar Berita.....	40
Tabel 4.2 Tabel Daftar Token dari Berita setelah dilakukan Proses Tokenisasi dan dalam Proses Penghilangan Stopwords .....	45
Tabel 4.3 Tabel Nilai $tf$ per Dokumen, Nilai $Df$ per Term dan Nilai $idf$ per Term.....	49
Tabel 4.4 Daftar Nilai $tf.idf$ dalam Data Matriks .....	50
Tabel 4.5 Perhitungan titik centroid Dokumen ke-1 .....	51
Tabel 4.6 Perhitungan titik centroid Dokumen ke-2 .....	52
Tabel 4.7 Perhitungan titik centroid Dokumen ke-3 .....	52
Tabel 4.8 Perhitungan titik centroid Dokumen ke-4 .....	53
Tabel 4.9 Perhitungan titik centroid Dokumen ke-5 .....	53
Tabel 4.10 Perhitungan titik centroid Dokumen ke-6 .....	54
Tabel 4.11 Daftar Anggota Cluster.....	60
Tabel 4.12 Daftar Anggota Cluster Iterasi ke-1 .....	62
Tabel 4.13 Daftar Centroid Baru pada Iterasi ke-1 .....	62
Tabel 4.14 Daftar Anggota Cluster Iterasi ke-2 .....	62
Tabel 4.15 Daftar Centroid Baru pada Iterasi ke-2 .....	62
Tabel 4.16 Daftar Anggota Cluster Iterasi ke-3 .....	62
Tabel 4.17 Daftar Centroid Baru pada Iterasi ke-3 .....	63
Tabel 4.18 Daftar Anggota Cluster Iterasi ke-4 .....	63
Tabel 4.19 Daftar Centroid Baru pada Iterasi ke-4 .....	63
Tabel 4.20 Daftar Klasifikasi.....	65
Tabel 4.21 Implementasi Modul Tokenizer.....	68
Tabel 4.22 Implementasi Modul Indexer .....	69
Tabel 4.23 Implementasi Modul Clustering Engine .....	69
Tabel 4.24 Pelaksanaan Pengujian (Black Box).....	73
Tabel 4.25 Hasil Perhitungan Purity terhadap Hasil Clustering non-Klasifikasi.....	75
Tabel 4.26 Hasil Perhitungan Purity terhadap Hasil Clustering dengan Klasifikasi .....	76
Tabel 4.27 Hasil Clustering .....	77

## DAFTAR GAMBAR

Gambar 2.1 <i>Data mining sebagai tahapan dalam proses KDD</i> .....	8
Gambar 2.2 <i>Cluster</i> .....	11
Gambar 2.3 <i>Contoh Data Matrix</i> .....	12
Gambar 2.4 <i>Contoh Dissimilarity Matrix</i> .....	12
Gambar 2.5 <i>Exclusive Clustering</i> .....	14
Gambar 2.6 <i>Tahap-Tahap dalam Text Mining</i> .....	16
Gambar 2.7 <i>Contoh Penerapan Model Ruang Vektor Dokumen</i> .....	21
Gambar 2.8 <i>Proses Algoritma K-Means</i> .....	23
Gambar 2.9 <i>Anatomi dari Mesin Pencarian Berita</i> .....	27
Gambar 2.10 <i>Halaman News Google</i> .....	29
Gambar 2.11 <i>Halaman Yahoo! News</i> .....	29
Gambar 2.12 <i>Halaman Daylife</i> .....	30
Gambar 2.13 <i>Halaman News MagPortal</i> .....	31
Gambar 2.14 <i>Halaman Newslookup</i> .....	31
Gambar 2.15 <i>Halaman LexisNexis</i> .....	32
Gambar 2.16 <i>Halaman Topix.net</i> .....	32
Gambar 3.1 <i>Desain Penelitian</i> .....	33
Gambar 3.2 <i>Model Sekuensial Linier</i> .....	38
Gambar 4.1 <i>Data Matriks Hasil Transformasi Data</i> .....	48
Gambar 4.2 <i>Transformasi Vektor Multidimensi ke Vektor 2 Dimensi</i> .....	55
Gambar 4.3 <i>Diagram E-R dari News Clustering System (NCS)</i> .....	66
Gambar 4.4 <i>Struktur tabel Basis Data</i> .....	67
Gambar 4.5 <i>Arsitektur News Clustering System (NCS)</i> .....	68
Gambar 4.6 <i>Pengurutan Dokumen Lain Berdasarkan pada Dokumen Terpilih</i> .....	71
Gambar 4.7 <i>Tampilan Hasil dari Komponen Server Side Interface</i> .....	72
Gambar 4.8 <i>Tampilan Depan pada Sisi Client</i> .....	73