

DAFTAR ISI

PRAKATA.....	iv
ARTI LAMBANG DAN SINGKATAN	vi
ABSTRACT	vii
INTISARI.....	ix
DAFTAR ISI	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL.....	xiv
BABI PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan masalah.....	4
1.3 Keaslian penelitian.....	5
1.4 Tujuan Penelitian	7
1.5 Batasan Masalah	7
1.6 Manfaat Penelitian.....	7
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI.....	8
2.1 Tinjauan Pustaka	8
2.2 Landasan Teori	8
2.2.1 Data mining	10
2.2.2 Text mining.....	11
2.2.3 <i>Web Scraping</i>	13
2.2.4 Pra-proses Teks.....	13
2.2.3.1 Case folding.....	14
2.2.3.2 Filtering	14
2.2.3.3 Stopword Removal	14
2.2.3.4 Stemming.....	14
2.2.4 Fitur Pembobotan Kata TF-IDF (<i>Term Frequency – Inverse Document Frequency</i>) dan VSM (<i>Vector Space Model</i>).....	16
2.2.2.3 Klasterisasi dokumen.....	17
2.2.4 Perhitungan Kesamaan Kosinus	18
2.2.5 <i>K-means Clustering</i>	18
2.2.6 Evaluasi Klasterisasi Dokumen	20
BAB III METODOLOGI.....	21
3.1 Alat dan Bahan.....	21
3.1.1 Alat.....	21
3.1.2 Bahan	21
3.2 Perancangan Sistem	24

3.2.1 Langkah-langkah penelitian.....	24
3.3 Perancangan Sistem	25
3.3.1 Penyiapan Data	25
3.3.2 Pra-proses Data.....	26
3.3.3 Pembobotan TF-IDF	29
3.3.4 Perhitungan Kesamaan Kosinus	30
3.3.5 <i>K-means clustering</i>	30
3.4 Tahapan Pengujian.....	30
3.5 Skenario Uji Coba	31
3.6 Cara Analisis.....	32
BAB IV HASIL DAN PEMBAHASAN	34
4.1 Hasil <i>Web Scraping</i>	34
4.2 Implementasi	36
4.2.1 Implementasi Klasterisasi dengan Fitur TF-IDF	36
4.2 Pelaksanaan Pengujian	40
4.2.1 Uji Coba Variasi Ukuran Pemilihan Fitur.....	41
4.2.2 Uji Coba Penentuan Klaster Terbaik	46
BAB V KESIMPULAN DAN SARAN.....	52
5.1 Kesimpulan	52
5.2 Saran	52
DAFTAR PUSTAKA	53