



DAFTAR ISI

PENGESAHAN.....	ii
PERNYATAAN.....	iii
PRAKATA.....	iv
ARTI LAMBANG DAN SINGKATAN.....	vi
ABSTRACT.....	vii
INTISARI.....	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xi
DAFTAR TABEL.....	xiii
BAB I PENDAHULUAN.....	5
1.1 Latar Belakang.....	5
1.2 Perumusan Masalah.....	6
1.3 Keaslian Penelitian.....	7
1.4 Tujuan Penelitian.....	9
1.5 Manfaat Penelitian.....	9
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI.....	10
2.1 Tinjauan Pustaka.....	10
2.2 Landasan Teori.....	13
2.2.1 <i>Email Spam</i>	13
2.2.2 Klasifikasi Teks.....	14
2.2.3 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	17
2.2.4 <i>Term Frequency-Relevance Frequency (TF-RF)</i>	21
2.2.5 <i>Delta Term Frequency-Inverse Document Frequency (Delta TF-IDF)</i>	23
2.2.6 <i>Term Frequency-Inverse Category Frequency</i>	25
2.2.7 <i>Log Relative Risk (LRR)</i>	27
2.2.8 Metode Klasifikasi <i>Naïve Bayes</i>	30
2.2.9 Metode Klasifikasi <i>Support Vector Machines (SVM)</i>	31
2.3 Hipotesis.....	33
BAB III METODOLOGI PENELITIAN.....	34
3.1 Alat dan Bahan Penelitian.....	34
3.1.1 Alat Penelitian.....	34
3.1.2 Bahan Penelitian.....	34
3.2 Alur Penelitian.....	35
3.2.1 Pengumpulan Data.....	36
3.2.2 Menghilangkan <i>Email Header</i> dan Konversi ke <i>Plain Teks</i>	36
3.2.3 <i>Preprocessing</i>	37
3.2.4 Pembobotan Fitur.....	40
3.2.5 Konversi CSV menjadi Sparse ARFF.....	42



3.2.6	Klasifikasi.....	43
3.2.7	Evaluasi	43
3.2.8	Uji Signifikansi	44
BAB IV HASIL DAN PEMBAHASAN		48
4.1	<i>Preprocessing</i>	48
4.2	Konversi Format <i>File</i> CSV ke Sparse ARFF	49
4.3	Evaluasi Metode Pembobotan Fitur	50
4.3.1	Akurasi	51
4.3.2	Presisi	53
4.3.3	Sensitivitas.....	54
4.4	Evaluasi Metode Klasifikasi	56
4.5	Uji Signifikansi	58
4.5.1	Akurasi	58
4.5.2	Presisi	62
4.5.3	Sensitivitas.....	65
BAB V KESIMPULAN DAN SARAN		69
5.1	Kesimpulan	69
5.2	Saran.....	69
DAFTAR PUSTAKA		71
LAMPIRAN		69