



DAFTAR ISI

PRAKATA	vi
DAFTAR ISI	vii
DAFTAR RUMUS	xiii
Intisari	xiv
Abstract	xv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA	5
BAB III LANDASAN TEORI	10
3.1 Klasifikasi Teks Dokumen	10
3.2 Text Preprocessing	11
3.2.1 Tokenisasi	11
3.2.2 <i>Stemming</i>	12
3.2.3 Stop Words Removal	14
3.3 <i>Term Document Matrix</i>	15
3.4 TF-IDF	15
3.5 Support Vector Machine	18
3.6 Multi Class SVM	22
3.7 Naïve Bayes Classifier	23
3.7.1 Multinomial Naïve Bayes	23
3.8 Random Forest	26
3.9 Evaluasi Performa	27
3.9.1 Akurasi	28
3.9.2 Presisi	28
3.9.3 Recall	28
3.9.4 F1-Score	29



3.10	K-fold Cross Validation.....	29
BAB IV ANALISIS DAN RANCANGAN		31
4.1	Deskripsi Penelitian	31
4.2	Rancangan Alur Penelitian	31
4.2.1	Studi Literatur	32
4.2.2	Akuisisi Dataset	33
4.2.3	Pelabelan Kategori.....	34
4.2.4	Pra-Pemrosesan.....	35
4.2.5	Ekstraksi Fitur.....	40
4.2.6	Pembagian Dataset.....	40
4.2.7	Algoritma Pengklasifikasi	41
4.2.8	Grid Search K-Fold Cross Validation.....	41
4.3	Rancangan Skenario Penelitian	42
BAB V IMPLEMENTASI		44
5.1	Spesifikasi Sistem.....	44
5.1.1	Pengumpulan Data.....	45
5.2	Implementasi Pelabelan Data	45
5.3	Implementasi Loading Dataset	47
5.4	Implementasi <i>Preprocessing</i>	49
5.4.1	Case Folding	49
5.4.2	URL dan Non-alpha-numeric Removal.....	49
5.4.3	Stemming.....	50
5.5	Implementasi Ekstraksi fitur.....	53
5.6	Implementasi Pembagian <i>Dataset</i>	53
5.7	Deklarasi Parameter Pencarian	54
5.8	Grid Search K-fold Cross Validation	55
5.8.1	Naïve Bayes Grid Search.....	56
5.8.2	Random Forest Grid Search.....	57
5.8.4	Hasil Grid Search K-Fold Cross Validation	59
5.9	Hasil Klasifikasi.....	61
6.1	Hasil Pengumpulan dan Pelabelan Data	62



6.2	Hasil Pre-Processing Data	63
6.3	Hasil Ekstraksi Fitur	66
6.4	Grid Search K-Fold Cross Validation.....	67
6.4.1	Naïve Bayes Grid Search.....	68
6.4.2	Random Forest Grid Search.....	69
6.4.3	SVM Grid Search	71
6.5	Hasil Pencarian Parameter terbaik.....	72
6.5	Hasil Evaluasi Model.....	73
6.5.1	Confusion Matrix.....	73
6.5.2	Perbandingan Algoritma Pengklasifikasi	75
BAB VII KESIMPULAN DAN SARAN		80
7.1	Kesimpulan	80
7.2	Saran	80
DAFTAR PUSTAKA.....		81