

## DAFTAR ISI

HALAMAN PENGESAHAN .....	iii
PERNYATAAN .....	iv
HALAMAN MOTTO DAN PERSEMBAHAN .....	v
PRAKATA.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR .....	xiv
DAFTAR PERSAMAAN .....	xvi
INTISARI .....	xvii
ABSTRACT.....	xviii
BAB I.....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	5
1.6 Sistematika Penulisan.....	5
BAB II.....	7
BAB III .....	11
3.1 <i>Machine Learning</i> .....	11
3.2 <i>Text Mining</i> .....	11
3.3 Klasifikasi Teks .....	13
3.4 Proses Klasifikasi .....	14
3.5 <i>Supervised Learning</i> .....	15
3.6 <i>Preprocessing</i> .....	16
3.6.1 <i>Cleansing</i> .....	17
3.6.2 <i>Case Folding</i> .....	17
3.6.3 Tokenisasi .....	17
3.6.4 Pembuangan <i>Stopword</i> .....	18
3.6.5 Pemotongan imbuhan ( <i>Stemming</i> ) .....	18
3.6.6 N-gram .....	23
3.7 <i>Indexing</i> .....	24
3.7.1 <i>Binary Weighting</i> .....	25

3.7.2	<i>Term Frequency</i> .....	26
3.7.3	<i>Term Frequency Inverse Document Frequency (TF-IDF)</i> .....	26
3.8	Naïve Bayes.....	29
3.8.1	Model Naïve Bayes.....	30
3.8.2	Multinomial Naïve Bayes .....	32
3.9	Evaluasi .....	33
3.9.1	<i>Accuracy</i> .....	34
3.9.2	<i>Precision</i> .....	34
3.9.3	<i>Recall</i> .....	35
3.9.4	<i>F-Measure</i> .....	35
3.9.5	<i>K-Fold Cross Validation</i> .....	35
3.10	Berita .....	36
3.10.1	Berita Ekonomi .....	36
3.11	Produk Domestik Bruto .....	37
3.11.1	Klasifikasi Baku Lapangan Usaha Indonesia (KBLI).....	41
BAB IV	.....	43
4.1	Perancangan dan Analisis Penelitian.....	43
4.2	Deskripsi Sistem .....	45
4.3	Data .....	47
4.3.1	Berita .....	47
4.3.2	Data <i>Stopword</i> .....	49
4.3.3	Data Kata Dasar .....	49
4.4	<i>Preprocessing</i> (Prapemrosesan) .....	50
4.4.1	Perancangan <i>Cleansing</i> .....	51
4.4.2	Perancangan <i>Stemming</i> (Pemotongan Imbuhan Kata) .....	52
4.4.3	Perancangan <i>Case Folding</i> .....	56
4.4.4	Perancangan Tokenisasi .....	57
4.4.5	Perancangan Pembuangan <i>Stopword</i> .....	58
4.4.6	<i>Running Time Preprocessing</i> .....	59
4.5	Perancangan Pembuatan <i>Vocabulary</i> .....	60
4.6	Perancangan Pembobotan <i>TF-IDF</i> .....	60

4.7	Klasifikasi.....	60
4.7.1	Pelabelan Data .....	61
4.6.2	Perancangan <i>Training</i> .....	62
4.6.3	Perancangan <i>Testing</i> .....	64
4.6.4	Perancangan Evaluasi dan Validasi Klasifikasi .....	65
BAB V	.....	66
5.1	Spesifikasi Sistem .....	66
5.2	Implementasi Sistem .....	66
5.3	Implementasi Pelabelan Data .....	67
5.4.	Implementasi <i>Preprocessing</i> .....	68
5.4.1	Implementasi <i>Cleansing</i> .....	68
5.4.2	Implementasi <i>Stemming</i> .....	69
5.4.3	Implementasi <i>Case Folding</i> .....	70
5.4.4	Implementasi Tokenisasi dan Pembuangan <i>Stopword</i> .....	71
5.3.5	Implementasi <i>Running Time Preprocessing</i> .....	72
5.5.	Imlementasi Pembuatan <i>Vocabulary</i> .....	72
5.6.	Implementasi Pembobotan TF-IDF .....	73
5.7	Implementasi Klasifikasi Multinomial Naïve Bayes .....	75
5.7.1	Implementasi <i>Training</i> .....	75
5.7.2	Implementasi <i>Testing</i> .....	76
5.8	Implementasi Pengujian Model Klasifikasi .....	77
BAB VI	.....	79
6.1	Pengujian Sistem dan Hasil .....	79
6.2	Hasil <i>Preprocessing</i> .....	80
6.3	Hasil <i>Vocabulary</i> .....	81
6.4	Hasil Pembobotan TF-IDF .....	83
6.5	Hasil Klasifikasi .....	84
6.7	Hasil Pengujian .....	86
BAB VII	.....	88
7.1	Kesimpulan.....	88
7.2	Saran.....	88



DAFTAR PUSTAKA .....	89
LAMPIRAN.....	93
A. Aturan Pemenggalan Awalan Algoritma <i>ECS Stemmer</i> .....	93
B. Kategori Kegiatan Ekonomi Berdasarkan KBLI 2015 .....	95
C. Daftar <i>Stopwords</i> Bahasa Indonesia (Tala, 2003).....	101