

## DAFTAR ISI

PRAKATA .....	vii
DAFTAR ISI .....	ix
DAFTAR GAMBAR .....	xii
DAFTAR TABEL .....	xiv
INTISARI .....	xv
ABSTRACT .....	xvi
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah .....	4
1.4 Tujuan Penelitian .....	4
1.5 Manfaat Penelitian .....	4
1.6 Sistematika Penulisan .....	5
BAB II TINJAUAN PUSTAKA .....	7
BAB III LANDASAN TEORI .....	13
3.1 <i>Text mining</i> .....	13
3.2 Klasifikasi Teks .....	14
3.3 <i>Preprocessing</i> .....	14
3.3.1 <i>Transform Case</i> .....	15
3.3.2 Tokenisasi .....	15
3.3.3 Penghilangan <i>Stopword</i> .....	15
3.3.4 <i>Stemming</i> .....	16
3.4 <i>Term Frequency – Inverse Document Frequency</i> .....	21
3.5 <i>Term Document Matrix</i> .....	23
3.6 <i>Naïve Bayes Classifier</i> .....	23
3.6.1 <i>Multinomial Naïve Bayes Classifier</i> .....	24
3.6.2 <i>Multivariate Bernoulli Naïve Bayes Classifier</i> .....	26
3.7 Evaluasi .....	27
3.7.1 Akurasi .....	28
3.7.2 Presisi .....	28
3.7.3 <i>Recall</i> .....	29
3.7.4 <i>F-Measure</i> .....	29
3.8 Cross Validation .....	29
BAB IV ANALISIS DAN PERANCANGAN .....	31
4.1 Analisis Permasalahan .....	31
4.2 Rancangan Umum Sistem .....	31
4.2.1 Proses <i>Training</i> .....	33
4.2.2 Proses <i>Testing</i> .....	34
4.2.3 Proses Evaluasi Nilai Performa .....	35
4.3 Data .....	37
4.3.1 Berita .....	37
4.3.2 Data <i>Stopword</i> .....	38
4.3.3 Data Kata Dasar .....	38
4.3.4 Pelabelan Data .....	38

4.4	<i>Preprocessing Data</i> .....	41
4.4.1	Rancangan <i>Transform Case</i> .....	41
4.4.2	Rancangan Tokenisasi .....	42
4.4.3	Perancangan Penghilangan <i>Stopword</i> .....	42
4.4.4	Perancangan <i>Stemming</i> .....	43
4.5	Perancangan Seleksi Fitur .....	46
4.5.1	Seleksi Fitur Metode TF-IDF .....	47
4.5.2	Seleksi Fitur Metode <i>Most Common</i> .....	50
4.6	Klasifikasi .....	51
4.6.1	Perancangan <i>Training Naïve Bayes</i> .....	52
4.6.2	Perancangan <i>Testing Naïve Bayes</i> .....	56
4.6.3	Perancangan Evaluasi dan Validasi Nilai Performa .....	59
BAB V IMPLEMENTASI .....		60
5.1	Lingkungan Implementasi .....	60
5.2	Implementasi Pengumpulan Data .....	60
5.3	Implementasi Pelabelan Data .....	62
5.4	Implementasi <i>Preprocessing</i> .....	63
5.4.1	Implementasi <i>Transform Case</i> .....	63
5.4.2	Implementasi Proses Tokenisasi .....	64
5.4.3	Implementasi Proses Penghilangan <i>Stopword</i> .....	64
5.4.4	Implementasi Proses <i>Stemming</i> .....	66
5.5	Implementasi Seleksi Fitur .....	66
5.5.1	Implementasi <i>Term Frequency-Inverse Document Frequency</i> .....	66
5.5.2	Implementasi <i>Most Common</i> .....	68
5.6	Implementasi <i>Training</i> .....	68
5.6.1	Implementasi <i>Training Model Multinomial</i> .....	69
5.6.2	Implementasi <i>Training Model Bernoulli</i> .....	70
5.7	Implementasi <i>Testing</i> dan Evaluasi .....	72
5.7.1	Implementasi <i>Testing</i> untuk Uji Performa Klasifikasi Data Baru ....	73
5.7.2	Implementasi <i>Testing</i> untuk Prediksi .....	76
5.8	Implementasi <i>K-Fold Cross Validation</i> .....	76
BAB VI HASIL DAN ANALISA .....		79
6.1	Pengujian Sistem dan Hasil .....	79
6.2	Hasil <i>Preprocessing</i> .....	79
6.2.1	Hasil <i>Transform Case</i> .....	80
6.2.2	Hasil Tokenisasi .....	80
6.2.3	Hasil Penghilangan <i>Stopword</i> .....	81
6.2.4	Hasil <i>Stemming</i> .....	81
6.3	Hasil Seleksi Fitur .....	82
6.4	Hasil Pembuatan <i>Term Document Matrix</i> .....	83
6.5	Hasil Proses <i>Training</i> .....	85
6.6	Hasil Proses <i>Testing</i> .....	88
6.6.1	Hasil Proses <i>Testing</i> untuk Uji Performa Klasifikasi Data Baru .....	88
6.6.2	Hasil Proses <i>Testing</i> untuk Prediksi .....	90
6.7	Hasil Evaluasi Nilai Performa .....	91
6.7.1	Hasil Evaluasi Nilai Performa dengan Metode <i>Most Common</i> .....	93

6.7.2 Hasil Evaluasi Nilai Performa dengan metode TF-IDF.....	93
BAB VII SARAN DAN KESIMPULAN .....	95
7.1 Kesimpulan .....	95
7.2 Saran .....	96
DAFTAR PUSTAKA .....	97
LAMPIRAN.....	101