

## DAFTAR ISI

<b>HALAMAN PENGESAHAN</b> .....	<b>i</b>
<b>PERNYATAAN</b> .....	<b>ii</b>
<b>HALAMAN MOTTO DAN PERSEMBAHAN</b> .....	<b>iii</b>
<b>PRAKATA</b> .....	<b>v</b>
<b>DAFTAR ISI</b> .....	<b>vii</b>
<b>DAFTAR TABEL</b> .....	<b>x</b>
<b>DAFTAR GAMBAR</b> .....	<b>xi</b>
<b>INTISARI</b> .....	<b>xiii</b>
<b>ABSTRACT</b> .....	<b>xiv</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1.Latar Belakang Masalah .....	1
1.2.Rumusan Masalah .....	3
1.3.Batasan Masalah .....	3
1.4.Tujuan Penelitian .....	3
1.5.Manfaat Penelitian .....	4
1.6.Metodologi Penelitian .....	4
1.7.Sistematika Penulisan .....	5
<b>BAB II TINJAUAN PUSTAKA</b> .....	<b>7</b>
<b>BAB III LANDASAN TEORI</b> .....	<b>13</b>
3.1.Berita .....	13
3.2.Text Mining .....	13
3.3.Klasifikasi Teks .....	14
3.4.Preprocessing .....	14
3.4.1 Transform Case .....	15
3.4.2 Tokenisasi .....	15
3.4.3 Penghilangan <i>Stopword</i> .....	15
3.4.4 <i>Stemming</i> .....	15
3.5. <i>Normalized Term Frequency</i> .....	17
3.6. <i>Term Frequency – Inverse Document Frequency</i> .....	18
3.7. <i>Chi Square</i> .....	19
3.8. <i>Term Document Matrix</i> .....	19
3.9. <i>K-Nearest Neighbors dengan Decision Rules</i> .....	20
3.10.Naïve Bayes Classifier .....	23
3.10.1 <i>Multinomial Naïve Bayes Classifier</i> .....	23
3.11.Evaluasi .....	24
3.11.1 Akurasi .....	25
3.11.2 Presisi .....	25
3.11.3 <i>Recall</i> .....	26
3.11.4 <i>F-measure</i> .....	26
3.12.Cross Validation .....	26
<b>BAB IV ANALISIS DAN PERANCANGAN</b> .....	<b>28</b>
4.1.Analisis Permasalahan .....	28
4.2.Rancangan Umum Sistem .....	28

4.2.1	Proses Training.....	29
4.2.2	Proses Testing .....	31
4.3.	Data.....	32
4.3.1	Berita .....	32
4.3.2	Data Stopword.....	32
4.3.3	Data Kata Dasar .....	33
4.4.	Preprocessing Data Berita .....	33
4.4.1	Rancangan <i>Transform Case</i> .....	33
4.4.2	Rancangan Tokenisasi.....	33
4.4.3	Perancangan Penghapusan Stopword.....	34
4.4.4	Perancangan Stemming.....	34
4.5.	Perancangan Seleksi Fitur .....	35
4.5.1	Seleksi Fitur Chi-Square .....	35
4.5.2	Seleksi Fitur Metode TF-IDF .....	37
4.6.	Klasifikasi.....	37
4.6.1	Perancangan Training <i>K-Nearest Neighbors with Decision Rules</i> .....	38
4.6.2	Perancangan Testing <i>K-Nearest Neighbors with Decision Rules</i> .....	39
4.6.3	Perancangan Training <i>Multinomial Naive Bayes</i> .....	42
4.6.4	Perancangan Testing <i>Multinomial Naive Bayes</i> .....	42
4.7.	Perancangan Evaluasi dan Validasi Nilai Performa .....	45
<b>BAB V</b>	<b>IMPLEMENTASI.....</b>	<b>47</b>
5.1.	Lingkungan Implementasi .....	47
5.2.	Implementasi Preprocessing .....	47
5.2.1	Implementasi <i>Transform Case</i> .....	47
5.2.2	Implementasi Tokenisasi.....	48
5.2.3	Implementasi penghapusan <i>stopwords</i> .....	48
5.2.4	Implementasi <i>Stemming</i> .....	49
5.3.	Implementasi Seleksi Fitur .....	49
5.3.1	Implementasi <i>Chi-Square</i> .....	49
5.3.2	Implementasi <i>Term Frequency-Inverse Document Frequency</i> .....	51
5.4.	Implementasi training .....	52
5.4.1	Implementasi training <i>K-Nearest Neighbors with Decision Rules</i> .....	52
5.4.2	Implementasi training model <i>Multinomial Naïve Bayes</i> .....	53
5.5.	Implementasi <i>testing</i> dan evaluasi.....	54
5.5.1	Implementasi testing <i>K-Nearest Neighbors with Decision Rules</i> .....	55
5.5.2	Implementasi testing <i>Multinomial Naïve Bayes</i> .....	58
5.6.	Implementasi <i>k-fold Cross Validation</i> .....	60
5.6.1	Implementasi <i>k-fold Cross Validation K-Nearest Neighbors Decision Rules</i> .....	61
5.6.2	Implementasi <i>k-fold Cross Validation Multinomial Naïve Bayes</i> .....	62

<b>BAB VI HASIL DAN PEMBAHASAN .....</b>	<b>64</b>
6.1. Hasil Preprocessing .....	64
6.2. Hasil Seleksi Fitur .....	65
6.3. Hasil Pembentukan <i>Term Document Matrix</i> .....	65
6.4. Hasil Proses <i>Training</i> .....	67
6.4.1 Hasil Proses <i>Training Multinomial Naïve Bayes</i> .....	67
6.4.2 Hasil Proses <i>Training K-Nearest Neighbors Decision Rules</i> .....	67
6.5. Hasil Proses <i>Testing</i> .....	67
6.6. Hasil Perbandingan Evaluasi Nilai Performa .....	68
6.6.1 Hasil Perbandingan Evaluasi Nilai Performa Akurasi .....	68
6.6.2 Hasil Perbandingan Evaluasi Nilai Performa Presisi .....	70
6.6.3 Hasil Perbandingan Evaluasi Nilai Performa <i>Recall</i> .....	71
6.6.4 Hasil Perbandingan Evaluasi Nilai Performa <i>F-Measure</i> ...	72
6.7. Hasil Perbandingan <i>Running Time</i> .....	73
6.8. Pengaruh Metode Seleksi Fitur terhadap Hasil Performa Klasifikasi .....	74
<b>BAB VII KESIMPULAN .....</b>	<b>76</b>
7.1. Kesimpulan .....	76
7.2. Saran .....	77
<b>DAFTAR PUSTAKA .....</b>	<b>78</b>
<b>LAMPIRAN .....</b>	<b>81</b>

## DAFTAR TABEL

Tabel 2. 1 Perbandingan dengan penelitian terdahulu .....	10
Tabel 3. 1 Contoh kegagalan algoritma ECS .....	16
Tabel 3. 2 Contoh <i>confusion matrix</i> .....	25
Tabel 6. 1 Perbandingan Akurasi <i>K-Nearest Neighbors Decision Rules</i> menggunakan <i>chisquare &amp; TFIDF</i> .....	69
Tabel 6. 2 Perbandingan Presisi <i>K-Nearest Neighbors Decision Rules</i> menggunakan <i>chisquare &amp; TFIDF</i> .....	70
Tabel 6. 3 Perbandingan Recall <i>K-Nearest Neighbors Decision Rules</i> menggunakan <i>chisquare &amp; TFIDF</i> .....	71
Tabel 6. 4 Perbandingan <i>F-Measure K-Nearest Neighbors Decision Rules</i> menggunakan <i>chisquare &amp; TFIDF</i> .....	73

## DAFTAR GAMBAR

Gambar 3.1 Proses text mining dalam mengolah teks dokumen .....	14
Gambar 3.2 Ilustrasi term document matrix .....	20
Gambar 3.3 Algoritma Improvisasi <i>KNN with Decision Rules</i> .....	22
Gambar 3.4 Algoritma <i>Multinomial Naïve Bayes Classifier</i> .....	23
Gambar 3.5 Ilustrasi proses <i>k-fold cross validation</i> .....	277
Gambar 4. 1 Rancangan Umum Sistem .....	30
Gambar 4. 1 Proses <i>training</i> .....	31
Gambar 4. 2 Proses <i>testing</i> .....	31
Gambar 4. 3 Diagram Alir proses <i>preprocessing</i> .....	33
Gambar 4. 4 Proses <i>Transform Case</i> .....	33
Gambar 4. 5 Proses <i>Tokenisasi</i> .....	34
Gambar 4. 6 Proses Seleksi Fitur <i>Chi-square</i> .....	36
Gambar 4. 7 Proses seleksi fitur <i>TF-IDF</i> .....	38
Gambar 4. 8 Proses Training <i>K-Nearest Neighbors with Decision Rules</i> .....	39
Gambar 4. 9 Proses <i>Testing K-Nearest Neighbors with Decision Rules</i> .....	40
Gambar 4. 10 Ilustrasi Perhitungan <i>Cosine Similarity</i> .....	41
Gambar 4. 11 Proses <i>training Multinomial Naive Bayes</i> .....	43
Gambar 4. 12 Proses testing Algoritma <i>Multinomial Naive Bayes</i> .....	44
Gambar 4. 13 Proses Evaluasi Nilai Performa.....	46
Gambar 5. 1 Kode implementasi <i>transform case</i> .....	47
Gambar 5. 2 Kode implementasi <i>tokenisasi</i> .....	48
Gambar 5. 3 Kode Penghapusan <i>Stopwords</i> .....	48
Gambar 5. 4 Kode Implementasi <i>stemming</i> .....	49
Gambar 5. 5 Kode Seleksi Fitur <i>Chi-square</i> .....	50
Gambar 5. 6 Kode seleksi fitur <i>TFIDF</i> .....	51
Gambar 5. 7 Kode <i>Training K-Nearest Neighbors Decision Rules</i> .....	52
Gambar 5. 8 Kode Training <i>Multinomial Naive Bayes</i> .....	53
Gambar 5. 9 Kode perhitungan <i>class_log_prior &amp; feature_log_prior</i> .....	54
Gambar 5. 10 Kode <i>Testing K-Nearest Neighbors Decision Rules</i> .....	55
Gambar 5. 11 Kode Fungsi <i>KNN</i> .....	56
Gambar 5. 12 Kode perhitungan jarak .....	58
Gambar 5. 13 Kode testing <i>Multinomial Naive Bayes</i> .....	59
Gambar 5. 14 Kode fungsi <i>predict_mnb()</i> .....	60
Gambar 5. 15 Kode <i>K-Fold K-Nearest Neighbors Decision Rules</i> .....	61
Gambar 5. 16 Kode <i>K-Fold Multinomial Naive Bayes</i> .....	62
Gambar 6. 1 Data berita sebelum preprocessing.....	64
Gambar 6. 2 Data berita pasca preprocessing .....	64
Gambar 6. 3 Hasil seleksi fitur TF-IDF .....	65
Gambar 6. 4 Hasil seleksi fitur <i>Chi-square</i> .....	65
Gambar 6. 5 Term Document Matrix dengan model term frequency.....	66
Gambar 6. 6 Term Document Matrix dengan model normalized term frequency.....	66
Gambar 6. 7 Array <i>class_log_prior</i> (kiri) & <i>feature_log_prob</i> (kanan).....	67
Gambar 6. 8 Contoh output prediksi klasifikasi.....	68

Gambar 6. 9 Perbandingan running time antar algoritma klasifikasi berdasarkan jumlah fitur..... 74