

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>HALAMAN PENGESAHAN.....</b>	<b>ii</b>
<b>PERNYATAAN BEBAS PLAGIASI .....</b>	<b>iii</b>
<b>PRAKATA .....</b>	<b>iv</b>
<b>DAFTAR ISI.....</b>	<b>vi</b>
<b>DAFTAR GAMBAR .....</b>	<b>viii</b>
<b>DAFTAR TABEL .....</b>	<b>ix</b>
<b>INTISARI .....</b>	<b>xi</b>
<b>ABSTRACT .....</b>	<b>xii</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah .....	3
1.4 Tujuan.....	4
1.5 Manfaat.....	4
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>5</b>
<b>BAB III LANDASAN TEORI.....</b>	<b>14</b>
3.1 Natural Language Processing .....	14
3.2 Text Mining .....	14
3.3 Text Preprocessing.....	14
3.4 <i>Bag of Words</i> (BoW).....	15
3.5 Klasifikasi Multilabel .....	17
3.6 <i>Binary Relevance</i> .....	17
3.7 <i>Conditional Probability</i> .....	18
3.8 Metode .....	18
3.8.1 Multinomial Naïve Bayes .....	18
3.8.2 Logistic Regression.....	20
3.8.3 Support Vector Machine.....	22
3.9 Evaluasi .....	24
3.9.1 Hamming Loss .....	24
3.9.2 Confusion Matrix .....	25
<b>BAB IV METODOLOGI PENELITIAN.....</b>	<b>27</b>
4.1 Deskripsi Umum Penelitian.....	27
4.2 Pengumpulan Data.....	27
4.3 Rancangan Model.....	29
4.3.1 <i>Text Preprocessing</i> .....	30
4.3.2 Pembagian <i>Data Train</i> dan <i>Data Test</i> .....	35
4.3.3 <i>Bag of Words</i> .....	36
4.3.4 MultilabelBinarizer .....	40
4.3.5 Multinomial Naïve Bayes untuk Multi Label .....	40
4.3.6 Logistic Regression.....	46
4.3.7 SVM.....	47
4.4 Evaluasi Model.....	48



4.4.1	<i>Hamming Loss</i> .....	48
4.4.2	<i>Confusion Matrix</i> .....	51
<b>BAB V</b>	<b>IMPLEMENTASI</b> .....	55
5.1	Spesifikasi Perangkat Keras dan Lunak .....	55
5.2	Data Implementasi.....	56
5.3	<i>Text Preprocessing</i> .....	56
5.3.1	<i>Case Folding</i> .....	56
5.3.2	<i>Cleaning</i> .....	57
5.3.3	<i>Stopword Removal</i> .....	58
5.3.4	<i>Lemmatization</i> .....	59
5.4	Pembagian Dataset .....	60
5.5	<i>Feature Extraction</i> .....	61
5.6	Pemodelan Multinomial Naïve Bayes .....	61
5.7	Metode .....	62
5.7.1	Pemodelan Multinomial Naïve Bayes.....	62
5.7.2	Pemodelan Logistic Regression .....	63
5.7.3	Pemodelan Support Vector Machine (SVM).....	64
5.8	Evaluasi .....	64
<b>BAB VI</b>	<b>HASIL DAN PEMBAHASAN</b> .....	67
6.1	Hasil Pengumpulan Data .....	67
6.2	Hasil <i>Text Preprocessing</i> .....	70
6.3	Hasil Pembagian Dataset.....	74
6.4	Hasil <i>Feature Extraction Bag of Words</i> .....	74
6.5	Hasil <i>MultiLabelBinarizer</i> .....	75
6.6	Hasil Klasifikasi Metode <i>Multinomial Naïve Bayes</i> .....	76
6.6.1	Evaluasi menggunakan <i>Confusion Matrix</i> .....	76
6.6.2	Evaluasi menggunakan <i>Hamming Loss</i> .....	84
6.7	Perbandingan Metode .....	85
<b>BAB VII</b>	<b>KESIMPULAN DAN SARAN</b> .....	87
7.1	Kesimpulan.....	87
7.2	Saran .....	87
<b>DAFTAR PUSTAKA</b>	.....	89

## DAFTAR GAMBAR

Gambar 3. 1 Binary Relevance (BR) .....	17
Gambar 3. 2 Fungsi Sigmoid .....	21
Gambar 3. 3 Himpunan data satu dimensi, hyperplane pemisah berupa sebuah titik (kotak) .....	22
Gambar 3. 4 Himpunan data dua dimensi, hyperplane pemisah berupa sebuah garis lurus.....	22
Gambar 3. 5 Himpunan data tiga dimensi, hyperplane pemisah berupa sebuah bidang datar.....	23
Gambar 3. 6 Hyperplane dengan margin maksimum .....	23
 Gambar 4. 1 Alur Rancangan Penelitian.....	27
Gambar 4. 2 Rancangan Model.....	30
Gambar 4. 3 Text Preprocessing .....	31
Gambar 4. 4 Multinomial Naive Bayes untuk Multi Label .....	41
Gambar 4. 5 Logistic Regression.....	46
Gambar 4. 6 Support Vector Machine.....	47
 Gambar 5. 1 Source Code dari Case Folding.....	56
Gambar 5. 2 Source Code dari Cleaning .....	57
Gambar 5. 3 Source Code Proses Stopword Removal.....	59
Gambar 5. 4 Source Code Proses Lemmatization .....	60
Gambar 5. 5 Source Code Proses Pembagian Dataset.....	60
Gambar 5. 6 Source Code Proses Bag of Word .....	61
Gambar 5. 7 Source Code Pengonversian Label menggunakan MultiLabelBinarizer.....	62
Gambar 5. 8 Source Code Pemodelan Multinomial Naive Bayes .....	62
Gambar 5. 9 Source Code Pemodelan Logistic Regression .....	63
Gambar 5. 10 Source Code Pemodelan SVM.....	64
Gambar 5. 11 Source Code Evaluasi menggunakan Confusion Matrix.....	65
Gambar 5. 12 Source Code Evaluasi menggunakan Hamming Loss .....	66
 Gambar 6. 1 Distribusi Jumlah Kategori Label .....	68
Gambar 6. 2 Distribusi Label MultiLabel.....	69
Gambar 6. 3 Hasil Pembagian Dataset.....	74
Gambar 6. 4 Hasil dari Bag of Words .....	75
Gambar 6. 5 Confusion Matrix Multilabel .....	83
Gambar 6. 6 Hamming Loss .....	85

## DAFTAR TABEL

Tabel 2. 1 Tinjauan Pustaka .....	9
Tabel 3. 1 Contoh data teks yang akan direpresentasikan.....	16
Tabel 3. 2 Kalimat dijadikan sebuah list kata .....	16
Tabel 3. 3 Menghitung frekuensi setiap kata .....	16
Tabel 3. 4 Tabel Confusion Matrix .....	26
Tabel 4. 1 Dataset yang dikumpulkan.....	28
Tabel 4. 2 Case Folding.....	31
Tabel 4. 3 Cleaning .....	32
Tabel 4. 4 Stopwords Removal .....	33
Tabel 4. 5 Lemmatization.....	34
Tabel 4. 6 Data Train.....	35
Tabel 4. 7 Data Test.....	36
Tabel 4. 8 Bag of Words.....	36
Tabel 4. 9 Jumlah Term dengan Kategori Label .....	40
Tabel 4. 10 MultiLabelBinarizer .....	40
Tabel 4. 11 Data Actual .....	49
Tabel 4. 12 Data Predict .....	49
Tabel 4. 13 Membandingkan Data Actual dan Data Predict .....	50
Tabel 4. 14 Confusion Matrix Artificial Intelligence .....	51
Tabel 4. 15 Confusion Matrix Electrical & Electric .....	52
Tabel 4. 16 Confusion Matrix Engineering.....	52
Tabel 4. 17 Confusion Matrix Information System .....	52
Tabel 4. 18 Confusion Matrix Robotic.....	53
Tabel 4. 19 Confusion Matrix .....	53
Tabel 5. 1 Spesifikasi Perangkat Keras (Hardware).....	55
Tabel 5. 2 Spesifikasi Perangkat Lunak (Software).....	55
Tabel 6. 1 Pengumpulan Dataset Penelitian.....	67
Tabel 6. 2 Presentase Kategori Label.....	69
Tabel 6. 3 Presentase Distribusi MultiLabel .....	70
Tabel 6. 4 Sebelum Case Folding dan Setelah Case Folding.....	70
Tabel 6. 5 Sebelum Cleaning dan Setelah Cleaning .....	71
Tabel 6. 6 Sebelum Stopwords Removal dan Setelah Stopwords Removal .....	72
Tabel 6. 7 Sebelum Lemmatization dan Setelah Lemmatization.....	72
Tabel 6. 8 Hasil Dataset yang telah di Text Preprocessing .....	73
Tabel 6. 9 Hasil Konversi Label Multilabel menjadi Biner .....	75



Tabel 6. 10 Hasil Confusion Matrix Artificial Intelligence .....	76
Tabel 6. 11 Hasil Confusion Matrix Electrical & Electric .....	77
Tabel 6. 12 Hasil Confusion Matrix Engineering .....	77
Tabel 6. 13 Hasil Confusion Matrix Information System .....	78
Tabel 6. 14 Hasil Confusion Matrix Robotic .....	78
Tabel 6. 15 Hasil Perhitungan menggunakan Macro-Average.....	81
Tabel 6. 16 Perbandingan Threshold pada MNB .....	86
Tabel 6. 17 Perbandingan Methode.....	86