

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
PERNYATAAN BEBAS PLAGIASI	iii
HALAMAN PERSEMBAHAN	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
INTISARI.....	xiv
<i>ABSTRACT</i>	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Pembatasan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Tinjauan Pustaka.....	3
1.5 Metode Penulisan.....	5
1.6 Sistematika Penulisan	6
BAB II LANDASAN TEORI	7
2.1 Matriks	7
2.1.1 Jenis-jenis matriks.....	7
2.1.2 Operasi matriks	8
2.2 Distribusi Variabel Random Bernoulli	9
2.3 Probabilitas	10
2.3.1 Probabilitas bersyarat.....	10
2.3.2 Kejadian bebas atau independen	11
2.4 Aturan Bayes.....	11

2.5	<i>Maximum a Posteriori</i>	12
2.6	Metode <i>Naïve Bayes</i>	13
2.7	Model probabilitas untuk <i>Naïve Bayes</i>	14
2.8	<i>Laplace correction</i>	16
2.9	<i>Chi-Square</i>	17
2.10	<i>Machine Learning</i>	17
2.11	Text mining	18
2.12	Klasifikasi	20
2.13	<i>Imbalanced Data</i>	20
2.14	<i>Web Scraping</i>	21
2.15	<i>E-commerce</i>	22
2.16	Tipe-Tipe Sampling	23
2.17	<i>Word Cloud</i>	26
2.18	Evaluasi Performa Model	26
BAB III ANALISIS SENTIMEN MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN SELEKSI FITUR CHI SQUARE DENGAN PENANGANAN DATA IMBALANCED		29
3.1	Analisis Sentimen	29
3.2	<i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	30
3.3	Model <i>Naïve Bayes</i> untuk Klasifikasi	35
3.4	Nilai <i>Likelihood</i> Distribusi Bernoulli	39
3.5	Simulasi dengan metode klasifikasi <i>Naïve Bayes</i>	42
3.6	Seleksi Fitur <i>Chi Square</i>	49
3.7	Penanganan Data <i>Imbalanced</i>	50
3.8	<i>Random Oversampling</i>	51
BAB IV STUDI KASUS		53
4.1	Deskripsi Kasus	53
4.2	Pengumpulan Data dengan <i>Web Scraping</i>	53
4.3	Analisis Statistika Deskriptif	56
4.4	Preprocessing Data	59
4.4.1	<i>Case folding</i>	60

4.4.2	<i>Tokenizing</i>	62
4.4.3	<i>Filtering</i>	62
4.4.4	<i>Stemming</i>	66
4.5	Pelabelan Kelas Sentimen	67
4.6	Ekstraksi Fitur	69
4.7	Pembentukan Data <i>Training</i> dan Data <i>Testing</i>	70
4.8	Metode <i>Naïve Bayes</i>	71
4.9	Evaluasi Performa Model Klasifikasi dengan Algoritma <i>Naïve Bayes</i> pada data <i>imbalanced</i>	73
4.10	Metode <i>Naïve Bayes</i> dan Seleksi Fitur <i>Chi Square</i> dengan Penanganan Data <i>Imbalanced</i>	74
4.10.1.	Penyeimbangan dataset dengan teknik <i>random oversampling</i>	74
4.10.2.	Seleksi Fitur <i>Chi Square</i>	75
4.11	Evaluasi Performa Metode Klasifikasi <i>Naïve Bayes</i> dan Seleksi Fitur <i>Chi Square</i> dengan Penanganan Data <i>Imbalanced</i>	76
4.12	Perbandingan Performa Klasifikasi	77
4.13	Visualisasi Data	78
BAB V	PENUTUP	81
5.1.	Kesimpulan	81
5.2.	Saran	82
DAFTAR	PUSTAKA	84
LAMPIRAN	87

DAFTAR TABEL

Tabel 2. 1 <i>Confusion Matrix</i>	27
Tabel 3. 1 Matriks Bobot $N \times M$	32
Tabel 3. 2 Perhitungan <i>term frequency</i>	33
Tabel 3. 3 Perhitungan <i>document frequency</i> dan <i>inverse document frequency</i>	33
Tabel 3. 4 Perhitungan bobot kata.....	34
Tabel 3. 5 Frekuensi kemunculan kata.....	42
Tabel 3. 6 Probabilitas kata kelas positif	44
Tabel 3. 7 Probabilitas kata kelas negatif.....	44
Tabel 3. 8 Nilai probabilitas <i>review</i> “jujur pakai suka cocok”	45
Tabel 3. 9 Nilai probabilitas <i>review</i> “pakai cekitcekit beruntus muncul”	45
Tabel 3. 10 Probabilitas kata kelas positif	47
Tabel 3. 11 Probabilitas kata kelas negatif.....	47
Tabel 3. 12 Nilai probabilitas <i>review</i> “jujur pakai suka cocok”	48
Tabel 3. 13 Nilai probabilitas <i>review</i> “pakai cekitcekit beruntus muncul”	48
Tabel 3. 14 Contoh hasil skor <i>chi-square</i> dan nilai <i>p-values</i>	50
Tabel 4. 1 Contoh <i>web scraping</i> pada Sociolla.....	54
Tabel 4. 2 Data <i>review</i> sebelum dan setelah melalui proses case folding.....	61
Tabel 4. 3 Data <i>review</i> setelah melalui proses tokenizing	62
Tabel 4. 4 <i>Output</i> Normalisasi	64
Tabel 4. 5 <i>Output Stopwords</i>	65
Tabel 4. 6 <i>Output</i> kata negasi.....	66
Tabel 4. 7 <i>Output Stemming</i>	67
Tabel 4. 8 Hasil pelabelan kelas sentimen	68
Tabel 4. 9 Perbandingan data <i>training</i> dan data <i>testing</i>	70
Tabel 4. 10 Probabilitas prior pada masing-masing kelas.....	71
Tabel 4. 11 Probabilitas pada masing-masing kata.....	71
Tabel 4. 12 Performa model klasifikasi dengan algoritma <i>Naïve Bayes</i>	73
Tabel 4. 13 <i>Confusion Matrix</i> klasifikasi <i>Naïve Bayes</i> pada perulangan ke-4.	74

Tabel 4. 14 Perbandingan data <i>training</i> sebelum dan sesudah <i>resampling</i> <i>random oversampling</i>	75
Tabel 4. 15 Peringkat hasil seleksi fitur <i>Chi Square</i> dengan alpha 0,05.....	76
Tabel 4. 16 <i>Confusion Matrix</i> klasifikasi <i>Naïve Bayes</i> dan seleksi fitur <i>Chi</i> <i>Square</i> ($\alpha=0,05$) dengan penanganan data <i>imbalanced</i>	76
Tabel 4. 17 Perbandingan performa klasifikasi	77
Tabel 4. 18 Frekuensi kata dari data <i>review</i> positif	78
Tabel 4. 19 Frekuensi kata dari data <i>review</i> negatif.....	79

DAFTAR GAMBAR

Gambar 2. 1 Tampilan awal pada situs Sociolla	23
Gambar 3. 1 <i>Resampling Random Oversampling</i>	52
Gambar 4. 1 Halaman <i>review</i> pada Sociolla.....	54
Gambar 4. 2 Barplot data <i>review</i> berdasarkan <i>rating</i>	56
Gambar 4. 3 Barplot berdasarkan <i>rating</i> terhadap <i>packaging</i>	56
Gambar 4. 4 Barplot berdasarkan <i>rating</i> terhadap <i>texture</i>	57
Gambar 4. 5 Barplot berdasarkan <i>rating</i> terhadap <i>effectiveness</i>	58
Gambar 4. 6 Barplot berdasarkan <i>rating</i> terhadap <i>value for money</i>	58
Gambar 4. 7 Barchart terhadap <i>repurchase</i>	59
Gambar 4. 8 <i>Word cloud review</i> positif	79
Gambar 4. 9. <i>Word cloud review</i> negatif	80

DAFTAR LAMPIRAN

Lampiran 1. <i>Data Review Hasil Web Scraping</i>	87
Lampiran 2. Hasil Pelabelan Kelas Sentimen	89
Lampiran 3. List <i>Stopwords</i> (819 kata)	89
Lampiran 4. Kamus Normalisasi (1493 kata)	91
Lampiran 5. Kamus Kata Negasi (218 kata)	92
Lampiran 6. Kamus Kata Positif (426 kata)	93
Lampiran 7. Kamus Kata Negatif (175 kata)	94
Lampiran 8. Visualisasi Data	95
Lampiran 9. <i>Syntax Preprocessing Data</i>	95
Lampiran 10. <i>Syntax</i> Pelabelan Kelas Sentimen	99
Lampiran 11. <i>Syntax</i> Visualisasi Data Pelabelan	100
Lampiran 12. <i>Syntax</i> Pembobotan <i>Term/Kata</i> (TF-IDF)	101
Lampiran 13. <i>Syntax</i> Klasifikasi <i>Naïve Bayes</i>	101
Lampiran 14. <i>Syntax Resampling Random Oversampling</i>	104
Lampiran 15. <i>Syntax</i> Klasifikasi <i>Naïve Bayes</i> dan Seleksi Fitur <i>Chi Square</i> .	104
Lampiran 16. <i>Syntax Wordcloud Data</i>	109
Lampiran 17. <i>Output</i> Klasifikasi <i>Naïve Bayes</i> pada Data <i>Imbalanced</i>	111
Lampiran 18. <i>Output</i> Klasifikasi <i>Naïve Bayes</i> dan Seleksi Fitur <i>Chi Square</i> dengan <i>Random Oversampling</i>	116