

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN PEMBIMBING	ii
HALAMAN PENGESAHAN PENGUJI.....	iii
SURAT PERNYATAAN KARYA ASLI TUGAS AKHIR.....	iv
PERNYATAAN BEBAS PLAGIAT.....	v
ABSTRAK	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	x
DAFTAR TABEL.....	xi
DAFTAR MODUL	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	2
1.5 Manfaat Penelitian.....	2
1.6 Metodologi Penelitian dan Pengembangan Sistem	3
1.6.1 Metodologi Penelitian	3
1.6.2 Metodologi Pengembangan Sistem	4
1.7 Sistematika Penulisan.....	4
BAB II TINJAUAN LITERATUR	6
2.1 Instagram.....	6
2.2 Spam.....	6
2.3 Python	6
2.4 Web Scraping.....	6
2.5 Data Mining	7
2.6 Preprocessing	7
2.7 Term Frequency-Inverse Document Frequency (TF-IDF).....	8
2.8 Support Vector Machine	8
2.9 Naive Bayes	10
2.10 Pengujian.....	11
2.10.1 Confusion Matrix	12

2.10.2	Akurasi	12
2.11	Studi Pustaka	12
BAB III METODOLOGI PENELITIAN DAN PENGEMBANGAN SISTEM		18
3.1	Metodologi Penelitian	18
3.1.1	Pengumpulan Data	18
3.1.2	Data Labelling.....	20
3.1.3	Preprocessing	20
3.1.4	Ekstraksi Fitur (TF-IDF)	23
3.1.5	Support Vector Machine	25
3.1.6	Naïve Bayes.....	28
3.1.7	Pengujian Model	29
3.2	Metodologi Pengembangan Sistem.....	29
3.2.1	Analisis Kebutuhan Sistem	30
3.2.2	Perancangan Sistem.....	31
3.2.3	Perancangan Proses	32
3.2.4	Perancangan Antarmuka	42
BAB IV HASIL DAN PEMBAHASAN		44
4.1	Hasil Penelitian	44
4.1.1	Implementasi Pengumpulan Dataset	44
4.1.2	Implementasi Labelling.....	45
4.1.3	Implementasi Preprocessing.....	45
4.1.4	Implementasi Seleksi Fitur.....	49
4.1.5	Implementasi Algoritma SVM	49
4.1.6	Implementasi Algoritma NB	50
4.1.7	Implementasi Sistem	51
4.2	Pengujian.....	52
4.2.1	Pengujian Balanced Dataset	52
4.2.2	Pengujian Unbalanced Dataset.....	54
4.3	Pembahasan.....	55
BAB V PENUTUP.....		57
5.1	Kesimpulan	57
5.2	Saran.....	57
DAFTAR PUSTAKA		xiii

DAFTAR GAMBAR

Gambar 2.1	<i>hyperplane</i>	9
Gambar 3.1	metodologi penelitian	18
Gambar 3.2	Hasil Observasi Akun Instagram	19
Gambar 3.3	Komentar spam.....	19
Gambar 3.4	Metode Modified Waterfall	30
Gambar 3.5	Arsitektur Sistem	31
Gambar 3.6	Flowchart Keseluruhan Sistem.....	32
Gambar 3.7	Flowchart Text Preprocessing	33
Gambar 3.8	Flowchart Casefolding.....	34
Gambar 3.9	Flowchart Cleaning.....	35
Gambar 3.10	Flowchart Normalisasi Bahasa	36
Gambar 3.11	Flowchart Slangword.....	37
Gambar 3.12	Flowchart Stemming.....	38
Gambar 3.13	Flowchart Tokenizing dan Stopwords Removal.....	39
Gambar 3.14	Flowchart Pembobotan TF-IDF.....	40
Gambar 3.15	Flowchart SVM	41
Gambar 3.16	Flowchart Naïve Bayes.....	41
Gambar 3.17	Rancangan Halaman Preprocessing.....	42
Gambar 3.18	Rancangan Halaman Klasifikasi.....	43
Gambar 3.19	Rancangan Halaman Deteksi Komentar Spam.....	43
Gambar 4.1	Hasil Scraping Data	45
Gambar 4.2	Dataset hasil Labelisasi.....	45
Gambar 4.3	Halaman Preprocessing	51
Gambar 4.4	Halaman Pelatihan dan Pengujian	51
Gambar 4.5	Halaman Deteksi Komentar Spam.....	52

DAFTAR TABEL

Tabel 2.1	<i>Confusion Matrix</i>	12
Tabel 2.2	<i>State of the Art</i>	14-16
Tabel 3.1	Contoh Data Hasil Web Scraping	20
Tabel 3.2	Contoh Data Hasil Labelling.....	20
Tabel 3.3	Contoh Data Hasil Casefolding.....	21
Tabel 3.4	Contoh Data Hasil Cleaning.....	21
Tabel 3.5	Contoh Data Hasil Normalisasi Bahasa	22
Tabel 3.6	Contoh Data Hasil Slangword.....	22
Tabel 3.7	Contoh Data Hasil Stemming.....	22
Tabel 3.8	Contoh Data Hasil Tokenizing dan Stopwords Removal.....	23
Tabel 3.9	Contoh Data Training Setelah Preprocessing.....	23
Tabel 3.10	Contoh Perhitungan TF-IDF.....	23
Tabel 3.11	TF-IDF Data Training ternormalisasi.....	24
Tabel 3.12	Contoh Data Testing Setelah Preprocessing.....	25
Tabel 3.13	TF-IDF Data Testing ternormalisasi	25
Tabel 3.14	Contoh hitung nilai w	26
Tabel 3.15	Hasil contoh hitung nilai w	27
Tabel 3.16	Contoh hitung nilai b.....	27
Tabel 3.17	Hasil contoh hitung nilai b	27
Tabel 3.18	Contoh hitung kernel data testing.....	27
Tabel 3.19	Hasil contoh hitung kernel data testing	28
Tabel 3.20	Hasil kernelisasi data testing	28
Tabel 3.21	Confussion Matrix Contoh Data Testing.....	29
Tabel 3.22	Tabel Spesifikasi Perangkat Keras	31
Tabel 3.23	Tabel Spesifikasi Perangkat Lunak	31
Tabel 4.1	Confussion Matriks SVM kernel Linear dengan Balanced Dataset.....	52
Tabel 4.2	SVM Linear Balanced Dataset, pengujian nilai C yang berbeda	53
Tabel 4.3	Confussion Matriks Multinomial NB dengan Balanced Dataset	53
Tabel 4.4	Multinomial NB Balanced Dataset, Lidstone Smoothing.....	53
Tabel 4.5	Confussion Matriks SVM kernel Linear dengan Unbalanced Dataset.....	54
Tabel 4.6	SVM Linear Unbalanced Dataset, pengujian nilai C yang berbeda.....	54
Tabel 4.7	Confussion Matriks Multinomial NB dengan Unbalanced Dataset	55
Tabel 4.8	Multinomial NB Unbalanced Dataset, Lidstone Smoothing.....	55

DAFTAR MODUL

Modul Program 4.1	Pengumpulan Dataset.....	44
Modul Program 4.2	Inisiasi Dataset	45
Modul Program 4.3	Inisiasi Dataset	45
Modul Program 4.4	Cleansing.....	46
Modul Program 4.5	Normalisasi Bahasa.....	46-47
Modul Program 4.6	Slangword	47
Modul Program 4.7	Hapus duplikasi data	48
Modul Program 4.8	Stemming	48
Modul Program 4.9	Tokenizing dan Stopwords Removal	48
Modul Program 4.10	Penyetaraan Dataset	48
Modul Program 4.11	Pembagian Dataset.....	49
Modul Program 4.12	Pembobotan TF-IDF	49
Modul Program 4.13	Pelatihan dan Pengujian SVM	50
Modul Program 4.14	Akurasi Pelatihan dan Pengujian SVM.....	50
Modul Program 4.15	Pelatihan dan Pengujian NB	50
Modul Program 4.16	Akurasi Pelatihan dan Pengujian NB.....	51