

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN DOSEN PEMBIMBING.....	Error! Bookmark not defined.
HALAMAN PENGESAHAN DOSEN PENGUJI	Error! Bookmark not defined.
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR.....	Error! Bookmark not defined.
HALAMAN PERSEMBAHAN	v
HALAMAN MOTO	vi
KATA PENGANTAR.....	vii
SARI	ix
GLOSARIUM	x
DAFTAR ISI	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR.....	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Metodologi Penelitian	4
1.7 Sistematika Penulisan	6
BAB II LANDASAN TEORI	8
2.1 Penelitian Sebelumnya.....	8
2.2 Tinjauan Pustaka	13
2.2.1 Klasifikasi Dokumen Teks	13
2.2.2 Machine Learning untuk Klasifikasi Dokumen Teks.....	15
2.2.3 Teknik <i>Semi-Supervised Learning</i>	16
2.2.4 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	17
2.2.5 <i>Multinomial Naïve Bayes</i>	18
2.2.6 <i>Pseudo Labeling</i>	19
BAB III METODE PENELITIAN	21
3.1 Langkah-langkah Pengerjaan Tugas Akhir.....	21
3.2 Uraian Metodologi	22
3.2.1 Identifikasi Masalah	22
3.2.2 Studi Literatur.....	22
3.2.3 Pengumpulan Data	22
3.2.4 Preprocessing.....	24
3.2.5 <i>Feature Extraction</i>	31
3.2.6 <i>Training</i>	33
3.2.7 Analisis dan Evaluasi	35
3.2.8 Implementasi Aplikasi.....	38
BAB IV HASIL DAN PEMBAHASAN	39
4.1 Pengumpulan Data	39
4.2 <i>Preprocessing</i>	40
4.3 Training menggunakan <i>Semi-Supervised Learning</i>	47
4.3.1 Membaca Data.....	49
4.3.2 <i>Training Data</i>	50

4.3.3	Prediksi Data Tidak Berlabel	51
4.3.4	Menggabungkan Data Berlabel & Data Tidak Berlabel	55
4.4	Analisis dan evaluasi.....	57
4.4.1	Membaca data.....	57
4.4.2	Analisis model <i>classifier</i> data berlabel.....	58
4.4.3	Analisis model <i>classifier</i> data tidak berlabel.....	61
4.4.4	Analisis model <i>classifier</i> data kombinasi.....	65
4.5	Implementasi Aplikasi	68
4.6	Hasil pengujian <i>website</i> untuk klasifikasi dokumen medis	76
BAB V KESIMPULAN DAN SARAN		78
5.1	Kesimpulan	78
5.2	Saran.....	79
DAFTAR PUSTAKA.....		80
LAMPIRAN.....		81



UNIVERSITAS ISLAM INDONESIA

DAFTAR TABEL

Tabel 2. 1 Kesimpulan dari penelitian sebelumnya.....	10
Tabel 3. 1 Jumlah dokumen klasifikasi.....	24
Tabel 3. 2 Menghilangkan <i>punctuation</i>	26
Tabel 3. 3 Menghilangkan <i>digit</i>	27
Tabel 3. 4 Menghilangkan <i>URL</i>	28
Tabel 3. 5 Penerapan <i>casefolding</i>	29
Tabel 3. 6 <i>Stop words</i>	29
Tabel 3. 7 Menghapus <i>stopword</i>	30
Tabel 3. 8 Penerapan <i>stemming</i>	31
Tabel 3. 9 Ilustrasi data yang digunakan	32
Tabel 3. 10 <i>Term frequency</i>	32
Tabel 3. 11 <i>Idf value</i>	32
Tabel 3. 12 Hasil <i>tf-idf</i>	33
Tabel 3. 13 Penerapan algoritma <i>multinomial naïve bayes</i>	34
Tabel 3. 14 Model probabilistik.....	34
Tabel 3. 15 Dokumen <i>testing</i>	36
Tabel 4. 1 Pengujian yang akan dilakukan.....	48
Tabel 4. 2 Akurasi model 4:3 data berlabel	59
Tabel 4. 3 Akurasi model 3:4 data berlabel	59
Tabel 4. 4 Akurasi model 2:5 data berlabel	60
Tabel 4. 5 Akurasi model 1:6 data berlabel	60
Tabel 4. 6 Hasil akurasi data <i>testing</i> dari model <i>classifier</i> data berlabel.....	61
Tabel 4. 7 Akurasi model 4:3 data tidak berlabel	62
Tabel 4. 8 Akurasi model 3:4 data tidak berlabel	63
Tabel 4. 9 Akurasi model 2:5 data tidak berlabel	63
Tabel 4. 10 Akurasi model 1:6 data tidak berlabel	64
Tabel 4. 11 Hasil akurasi data <i>testing</i> dari model <i>classifier</i> data tidak berlabel.....	64
Tabel 4. 12 Akurasi model 4:3 data kombinasi	65
Tabel 4. 13 Akurasi model 3:4 data kombinasi	66
Tabel 4. 14 Akurasi model 2:5 data kombinasi	66
Tabel 4. 15 Akurasi model 1:6 data kombinasi	67
Tabel 4. 16 Hasil akurasi data <i>testing</i> dari model <i>classifier</i> data kombinasi.....	67

Tabel 4. 18 Pengujian fungsionalitas sistem.....77



DAFTAR GAMBAR

Gambar 2. 1 Teknik <i>semi-supervised learning</i>	17
Gambar 2. 2 Probabilitas bersyarat (Basuki & Ahmad 2006)	18
Gambar 2. 3 Langkah <i>pseudo labeling</i>	20
Gambar 3. 1 Langkah pengerjaan tugas akhir	21
Gambar 3. 2 Skema preprocessing.....	25
Gambar 3. 3 Alur proses menghilangkan <i>punctuation</i>	25
Gambar 3. 4 Alur proses menghilangkan <i>digit</i>	26
Gambar 3. 5 Alur proses menghilangkan <i>URL</i>	27
Gambar 3. 6 Alur proses menghilangkan <i>whitespace</i>	28
Gambar 4. 1 <i>Import</i> modul <i>glob</i> dan <i>os</i>	40
Gambar 4. 2 Kode program membaca folder dokumen training berlabel dan tidak berlabel..	40
Gambar 4. 3 Kode program membaca folder dokumen testing	41
Gambar 4. 4 <i>Import</i> modul <i>re</i>	41
Gambar 4. 5 Kode program <i>cleaning</i>	42
Gambar 4. 6 Kode program melakukan <i>casefolding</i>	43
Gambar 4. 7 <i>Import</i> modul <i>stemming</i> sastrawi	43
Gambar 4. 8 Kode program <i>stemming</i> bahasa indonesia.....	43
Gambar 4. 9 <i>Import nltk</i> untuk <i>stopword</i>	44
Gambar 4. 10 Kode program untuk menghilangkan <i>stopwords</i>	44
Gambar 4. 11 Kode program untuk <i>preprocessing</i> dan menyimpan dokumen berlabel ke dalam dokumen baru.....	44
Gambar 4. 12 Kode program untuk <i>preprocessing</i> dan menyimpan dokumen tidak berlabel ke dalam dokumen baru.....	45
Gambar 4. 13 Kode program untuk <i>preprocessing</i> dan menyimpan dokumen <i>testing</i> ke dalam dokumen baru.....	45
Gambar 4. 14 <i>Converting</i> dokumen <i>training</i> berlabel	46
Gambar 4. 15 <i>Converting</i> dokumen <i>training</i> tidak berlabel	46
Gambar 4. 16 <i>Converting</i> dokumen <i>testing</i>	47
Gambar 4. 17 Alur proses klasifikasi.....	49
Gambar 4. 18 <i>Import</i> modul <i>pandas</i>	49
Gambar 4. 19 Membaca data latih berlabel	50
Gambar 4. 20 <i>Import</i> modul <i>tf-idf</i>	50

Gambar 4. 21 Transformasi <i>tf-idf</i>	50
Gambar 4. 22 <i>Import</i> modul <i>classifier</i>	51
Gambar 4. 23 Model <i>Multinomial Naive Bayes</i>	51
Gambar 4. 24 Membaca data latih tidak berlabel	51
Gambar 4. 25 Kode program untuk melakukan prediksi.....	52
Gambar 4. 26 Hasil prediksi percobaan 1 pada data tidak berlabel.....	52
Gambar 4. 27 Hasil prediksi percobaan 2 pada data tidak berlabel.....	52
Gambar 4. 28 Hasil prediksi percobaan 3 pada data tidak berlabel.....	53
Gambar 4. 29 Hasil prediksi percobaan 4 pada tidak berlabel	53
Gambar 4. 30 Menyimpan kategori hasil prediksi.....	54
Gambar 4. 31 Menggabungkan label dan teks data	54
Gambar 4. 32 Kode program untuk membaca dokumen hasil klasifikasi.....	54
Gambar 4. 33 <i>tf-idf</i> pada data tidak berlabel.....	54
Gambar 4. 34 Model <i>classifier</i> data tidak berlabel	55
Gambar 4. 35 Membaca data kombinasi.....	55
Gambar 4. 36 Menggabungkan dokumen kombinasi	56
Gambar 4. 37 Membaca dokumen gabungan	56
Gambar 4. 38 <i>tf-idf</i> pada dokumen data kombinasi	56
Gambar 4. 39 Model <i>classifier</i> dari data kombinasi	56
Gambar 4. 40 Membaca data testing.....	57
Gambar 4. 41 Menyimpan data ke dalam <i>list</i>	58
Gambar 4. 42 Prediksi data testing menggunakan model <i>classifier</i> berlabel	58
Gambar 4. 43 <i>Import</i> modul <i>numpy</i>	58
Gambar 4. 44 Menghitung akurasi dari model <i>classifier</i> data berlabel	58
Gambar 4. 45 Prediksi data <i>testing</i> menggunakan model <i>classifier</i> tidak berlabel	61
Gambar 4. 46 Menghitung akurasi dari model <i>classifier</i> data tidak berlabel	62
Gambar 4. 47 Prediksi data <i>testing</i> menggunakan model <i>classifier</i> kombinasi data.....	65
Gambar 4. 48 Menghitung akurasi dari model <i>classifier</i> data kombinasi	65
Gambar 4. 49 Tabel hasil akurasi klasifikasi dokumen medis.....	68
Gambar 4. 50 Grafik hasil akurasi klasifikasi dokumen medis	69
Gambar 4. 51 Grafik hasil <i>training</i> data berlabel	69
Gambar 4. 52 Grafik hasil <i>training</i> data tidak berlabel	70
Gambar 4. 53 Grafik hasil <i>training</i> data kombinasi	70
Gambar 4. 54 Menu <i>dropdown</i> pilihan model <i>classifier</i>	71

Gambar 4. 55 <i>Text input</i> menggunakan model 4:3	72
Gambar 4. 56 <i>Text input</i> menggunakan model 3:4	72
Gambar 4. 57 <i>Text input</i> menggunakan model 2:5	73
Gambar 4. 58 <i>Text input</i> menggunakan model 1:6	73
Gambar 4. 59 Hasil prediksi menggunakan model 4:3	74
Gambar 4. 60 Hasil prediksi menggunakan model 3:4	75
Gambar 4. 61 Hasil prediksi menggunakan model 2:5	75
Gambar 4. 62 Hasil prediksi menggunakan model 1:6	76

