

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
PERNYATAAN BEBAS PLAGIASI	iii
PRAKATA	iv
DAFTAR ISI	vi
DAFTAR GAMBAR	ix
DAFTAR TABEL	xi
DAFTAR SIMBOL	xii
DAFTAR SINGKATAN	xiv
INTISARI	xv
ABSTRACT	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	7
1.3 Batasan Masalah	7
1.4 Tujuan Penelitian	7
1.5 Kontribusi Penelitian	8
1.6 Sistematika Penulisan	8
BAB II TINJAUAN PUSTAKA	10
2.1 Seleksi Fitur Menggunakan Swarm Intelligence	10
2.2 Hybrid Swarm Intelligence	14
2.3 Seleksi Fitur Filter-Wrapper	16
BAB III LANDASAN TEORI	25
3.1 Teks Mining	25
3.2 Dokumen teks	26
3.3 Klasifikasi Teks	26
3.3.1 Klasifikasi Single Label	27
3.3.2 Klasifikasi Multi-class	28
3.3.3 Klasifikasi Multi-label	28
3.4 Prapemrosesan Teks	29
3.5 Ekstraksi Fitur Teks	30
3.5.1 TF-IDF	31

3.5.2 Bag of words	33
3.6 Seleksi Fitur Teks	33
3.6.1 Pendekatan Filter	35
3.6.1.1 Chi Square	36
3.6.2 Pendekatan Wrapper	39
3.7 Swarm Intelligence	41
3.7.1 Swarm Intelligence dalam Pendekatan Wrapper	47
3.7.2 Ant Colony Optimization	48
3.6.2.1 Ant Colony Optimization untuk Seleksi Fitur	50
3.7.3 Grey Wolf Optimizer	54
3.7.3.1 Grey Wolf Optimizer untuk Seleksi Fitur	56
3.8 Multilayer Perceptron	61
3.9 Naïve Bayes Classifier	63
3.9.1 Mutinomial Naïve Bayes	63
3.9.2 Complement Naïve Bayes Classifier	64
3.10 Support Vector Machine	64
3.11 Confusion Matrix	66
3.12 Hamming Loss	68
BAB IV PERANCANGAN MODEL	70
4.1 Pengumpulan Data	73
4.2 Arsitektur Model Seleksi Fitur Teks	74
4.3 Prapemrosesan Teks	80
4.3.1 Ekstraksi Fitur Teks	83
4.4 Rancangan Seleksi Fitur Teks	86
4.5 Rancangan Algoritma Ant Colony Grey Wolf Optimization	91
4.6 Klasifikasi Teks	94
4.6.1 Rancangan Arsitektur MLP	95
4.7 Skenario Pengujian	96
4.7.1 Pengujian dengan Micro F1 Score	98
4.7.2 Pengujian dengan Hamming Loss	101
BAB V IMPLEMENTASI MODEL	102
5.1 Identifikasi Data	102
5.2 Analisis Data	103
5.3 Implementasi Model	105

5.3.1	Prapemrosesan Teks.....	105
5.3.2	Ekstraksi Fitur	108
5.3.3	Seleksi Fitur Filter.....	110
5.3.4	Seleksi Fitur Wrapper	111
5.3.5	Algoritma ACOGWO	116
5.3.6	Klasifikasi	120
BAB VI HASIL DAN PEMBAHASAN		122
6.1	Pemodelan.....	122
6.2	Evaluasi Model	124
6.2.1	Evaluasi Model dengan Ekstraksi Fitur BOW	126
6.2.2	Evaluasi Model dengan Ekstraksi Fitur TF-IDF.....	131
6.3	Pembahasan Evaluasi Model	137
6.4	Prototipe Sistem	143
BAB VII KESIMPULAN DAN SARAN.....		147
7.1	Kesimpulan	147
7.2	Saran	147
DAFTAR PUSTAKA.....		148

DAFTAR GAMBAR

Gambar 3.1 Framework Seleksi Fitur Wrapper	40
Gambar 3.2 Flowchart Swarm Intelligence System.....	45
Gambar 3.3 Ilustrasi Algoritma Ant Colony.....	49
Gambar 3.4 Representasi ACO dalam Seleksi Fitur.....	51
Gambar 3.5 Struktur Hirarki Sosial Grey Wolf.....	56
Gambar 3.6 Mekanisme GWO.....	58
Gambar 3.7 Struktur MLP	62
Gambar 3.8 Hyperplane SVM	65
Gambar 4.1 Tahapan Penelitian.....	71
Gambar 4.2 Sampel data	74
Gambar 4.3 Arsitektur Model.....	77
Gambar 4.4 Tahapan preprocessing.....	80
Gambar 4.5 Teks Sebelum Preprocessing.....	81
Gambar 4.6 Teks Setelah Case Folding.....	81
Gambar 4.7 Teks Setelah Stopwords Removal.....	82
Gambar 4.8 Teks Setelah Lemmatization.....	82
Gambar 4.9 Flowchart Seleksi Fitur Usulan.....	89
Gambar 4.10 Alur Klasifikasi Teks	94
Gambar 4.11 Rancangan Arsitektur MLP.....	95
Gambar 4.12 Model Multi-label Confusion Matrix.....	99
Gambar 5.1 Sampel Dataset Korpus	102
Gambar 5.2 Distribusi label pada dataset.....	103
Gambar 5.3 Distribusi Normal Jumlah Kata.....	104
Gambar 5.4 Potongan Instruksi untuk jumlah fitur	105
Gambar 5.5 Potongan Instruksi untuk Prapemrosesan	107
Gambar 5.6 Potongan Instruksi untuk Ekstraksi Fitur.....	109
Gambar 5.7 Potongan Instruksi untuk Chi-square.....	110
Gambar 5.8 Potongan Instruksi Klasifikasi dalam Wrapper.....	112
Gambar 5.9 Potongan Instruksi untuk Inisiasi pada Wrapper	114

Gambar 5.10 Potongan Instruksi Algoritma ACOGWO	118
Gambar 5.11 Potongan Instruksi Inisiasi MLP	120
Gambar 6.1 Grafik Seleksi Fitur pada Model BOW	130
Gambar 6.2 Grafik MicroF1 Score	131
Gambar 6.3 Grafik Seleksi Fitur Pada Model TF-IDF	136
Gambar 6.4 Grafik Perbandingan Micro F1 Score	136
Gambar 6.5 Grafik Perbandingan Waktu Eksekusi	140
Gambar 6.6 Grafik Persentase Seleksi Fitur	141
Gambar 6.7 Grafik Perbandingan Micro F1 score	141
Gambar 6.8 Grafik Perbandingan Hamming Loss score	142
Gambar 6.9 Tampilan Prototipe.....	144
Gambar 6.10 Tampilan Hasil Ekstrak Dokumen Teks	144
Gambar 6.11 Tampilan Klasifikasi Dokumen Teks.....	145
Gambar 6.12 Tampilan Hasil Prediksi Label.....	145
Gambar 6.13 Tampilan Fitur Terbaik	146

DAFTAR TABEL

Tabel 2.1 Kajian Seleksi Fitur Menggunakan Swarm Intelligence	17
Tabel 2.2 Kajian Hybrid Swarm Intelligence	22
Tabel 2.3 Kajian Seleksi Fitur Filter-Wrapper	24
Tabel 3.1 Contoh Klasifikasi Single-Label.....	28
Tabel 3.2 Contoh Klasifikasi Multi-Class	28
Tabel 3.3 Contoh Klasifikasi Multi-Label	29
Tabel 3.4 Prinsip Kerja Swarm Intelligence	43
Tabel 3.5 Confusion Matrix	66
Tabel 4.1 Sampel Dataset.....	74
Tabel 4.2 Sampel Korpus Dokumen Teks	83
Tabel 4.3 Representasi Matriks TF	84
Tabel 4.4 Representasi Matriks IDF	84
Tabel 4.5 Representasi Matriks TF-IDF	86
Tabel 4.6 Representasi Matriks Bag of Words	86
Tabel 6.1 Pengaturan Parameter	124
Tabel 6.2 Tanpa Seleksi Fitur	126
Tabel 6.3 Evaluasi Pendekatan Filter.....	127
Tabel 6.4 Evaluasi Pendekatan Wrapper	127
Tabel 6.5 Evaluasi Pendekatan Filter-Wrapper	129
Tabel 6.6 Evaluasi Tanpa Seleksi Fitur	132
Tabel 6.7 Evaluasi Pendekatan Filter.....	132
Tabel 6.8 Evaluasi Pendekatan Wrapper	132
Tabel 6.9 Evaluasi Pendekatan Filter-Wrapper	134

DAFTAR SIMBOL

Simbol	Arti
$TF(t_i, d_j)$	Frekuensi kemunculan kata t_i dalam dokumen d_j
$IDF(t_i, D)$	Nilai fitur penting dalam dokumen
$f(t_i, d_j)$	Kemunculan fitur t_i dalam dokumen d_j
N	Jumlah total kata dalam dokumen
χ^2	Fungsi chi square
T	Term atau fitur
k	Jumlah fitur
c	Kelas
A	Jumlah fitur dan kelas muncul bersamaan
B	Jumlah fitur muncul tanpa kelas
C	Jumlah kelas muncul tanpa fitur
D	Jumlah kelas dan fitur tidak muncul sama sekali
N	Total baris dari dokumen training
DF	Degree of freedom
R	Jumlah level pada satu variabel kelas
C	Jumlah level pada variabel yang lain
$p_{ij}^k(t)$	Probabilitas ant- k memilih fitur i
J^k	Jumlah fitur yang belum dikunjungi oleh ant k
η_i	Nilai heuristik pada fitur i
$\tau_i(t)$	Nilai pheromone pada fitur i dalam iterasi t
$\tau_j(t)$	Nilai pheromone pada fitur j dalam iterasi t
A	Pembobotan pheromone
B	Pembobotan heuristik
ρ	Koefisien penguapan pheromone
$\Delta\tau_i$	Variabel perubahan pheromone
$P(c d)$	Probabilitas dokumen d berada di kelas c
$P(c)$	Probabilitas prior suatu dokumen berada di kelas c
t_n	Token dalam dokumen d
$P(t_k c)$	Probabilitas bersyarat fitur t_k berada di kelas c
$\Delta\tau_i$	Perubahan pheromone
τ_i	Pheromone di fitur i
ρ	Penguapan pheromone
m	Jumlah ant keseluruhan
$\Delta\phi_{ij}^k$	Jumlah pheromone yang dihasilkan ant k pada vertex i
Q	jumlah pheromone yang disimpan pada agen ant
$f(x)$	Probabilitas pheromone
σ	Faktor intensitas ACO
i	Jumlah populasi ant
$\vec{D}$	Update posisi setiap grey wolf

Simbol	Arti
$\vec{X}$	Posisi grey wolf
$\vec{X}_p$	Posisi grey wolf
$\vec{X}_\alpha$	Posisi alfa wolf
$\vec{X}_\beta$	Posisi beta wolf
$\vec{X}_\delta$	Posisi delta wolf
$\vec{D}_\alpha$	Update posisi alpa wolf
$\vec{D}_\beta$	Update posisi beta wolf
$\vec{D}_\delta$	Update posisi delta wolf
$\vec{A}$	Koefisien vektor wolf
$\vec{C}$	Koefisien vektor wolf
$\vec{r}_1$	Vektor acak dalam GWO
$\vec{r}_2$	Vektor acak dalam GWO
W	Vektor bobot hyperplane
x_i	Himpunan data training
y_i	Label kelas dari X_i
b	Parameter bias

DAFTAR SINGKATAN

ABC	: Artificial Bee Colony
ACO	: Ant Colony Optimization
BOW	: Bag of Words
CS	: Chi Square
CNB	: Complement Naïve Bayes
FA	: Firefly Search
GOA	: Grasshopper Optimization Algorithm
GWO	: Grey Wolf Optimizer
HHO	: Harris Hawks Optimizer
MLP	: Multi-Layer Perceptron
MNB	: Multinomial Naïve Bayes
NB	: Naïve Bayes
RW-GWO	: Random Walk-Grey Wolf Optimizer
SVM	: Support Vector Machine
TF-IDF	: Term Frequency-Inverse Document Frequency
WOA	: Whale Optimization Algorithm