

**PERBAIKAN FEATURE SELECTION PADA SUPPORT VECTOR MACHINE
MENGGUNAKAN PARTICLE SWARM OPTIMIZATION UNTUK
SENTIMENT ANALYSIS TWITTER
PARIWISATA INDONESIA**

SKRIPSI



disusun oleh :

Wahyu Setiawan

15.21.0845

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2017**

**PERBAIKAN FEATURE SELECTION PADA SUPPORT VECTOR MACHINE
MENGGUNAKAN PARTICLE SWARM OPTIMIZATION UNTUK
SENTIMENT ANALYSIS TWITTER
PARIWISATA INDONESIA**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh :

Wahyu Setiawan

15.21.0845

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2017**

Persetujuan

SKRIPSI

**PERBAIKAN FEATURE SELECTION PADA SUPPORT VECTOR MACHINE
MENGGUNAKAN PARTICLE SWARM OPTIMIZATION UNTUK
SENTIMENT ANALYSIS TWITTER
PARIWISATA INDONESIA**

yang dipersiapkan dan disusun oleh

**Wahyu Setiawan
15.21.0845**

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 5 Agustus 2017

Dosen Pembimbing,

Kusrini S.Kom, M.Kom, Dr
NIK. 190302106

PENGESAHAN

SKRIPSI

PERBAIKAN FEATURE SELECTION PADA SUPPORT VECTOR MACHINE
MENGGUNAKAN PARTICLE SWARM OPTIMIZATION UNTUK
SENTIMENT ANALYSIS TWITTER
PARIWISATA INDONESIA

yang dipersiapkan dan disusun oleh

Wahyu Setiawan

15.21.0845

telah dipertahankan di depan Dewan Pengaji
Pada tanggal 22 Agustus 2017

Susunan Dewan Pengaji

Nama Pengaji

Kusrini S.Kom, M.Kom, Dr.
NIK. 190302106

Tanda Tangan

Melwin Syafrizal, S.Kom, M.Eng.
NIK. 190302105

Mei Parwanto Kurniawan, M.Kom.
NIK. 190302187

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 Agustus 2017



PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggung jawab saya pribadi.

Yogyakarta, 30 Agustus 2017



Wahyu Setiawan

NIM. 15.21.0845

MOTTO

Man jadda wa jadda

Sesungguhnya bersama kesulitan itu ada kemudahan

QS. Al Insyirah : 6

*Saat sebuah urusan telah terselesaikan mulailah untuk
mengerjakan urusan yang lain dengan sungguh-sungguh*

QS. Al Insyirah : 8

*Sebaik-baik kalian adalah yang paling bermanfaat bagi orang
lain*

HR. Tirmidzi

*“...sesungguhnya yang benar-benar takut kepada Allah,
adalah hamba-hamba-Nya yang mempunyai ilmu..”*

QS. Fathir : 28

“Percayalah pada kemauan, perasaan dan harapan anda sendiri”

*“Berdo’alah seperti segala-galanya bergantung kepada Tuhan dan
berkerjalah seperti segala-galanya bergantung kepada anda ”*

PERSEMBAHAN

Alhamdulillah segala puji syukur atas kehadirat Allah SWT yang telah memberikan rahmat dan karunia-Nya sehingga karya ini dapat terselesaikan dengan sebaik – baiknya, tidak lepas dari bantuan dan dukungan berbagai pihak.

Skripsi ini saya persembahkan dengan rasa syukur kepada saksi dan penolong seumur hidupku, Allah SWT

Untuk Bapak, Ibu, Adikku *jazakumullah khairan katsiir* atas pengertian dan harapan kalian, semoga Allah kuatkan kami untuk selalu berbakti..

Untuk Muhammad Jaga *jazakillah* atas bimbingan, ilmu, alur berfikir, kerja keras dan pengorbanannya. Semua pihak yang telah banyak membantu dalam penyusunan skripsi ini yang tidak dapat disebutkan satu persatu.

KATA PENGANTAR

Puji syukur penulis persembahkan atas kehadiran Allah SWT yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul perbaikan feature selection pada *Support Vector Machine* menggunakan *Particle Swarm Optimization* untuk *Sentiment Analysis* Twitter pariwisata Indonesia dengan sebaik – baiknya. Tidak lupa sholawat serta salam penulis haturkan kepada junjungan umat Nabi Muhammad SAW, yang telah membawa umat Islam dari jaman jahiliyah ke jaman yang penuh ilmu pengetahuan.

Dengan selesainya skripsi ini, maka penulis mengucapkan terima kasih kepada :

1. Ibu Kusrini, Dr., M.Kom selaku dosen pembimbing yang telah sabar membimbing dan banyak memberikan pengarahan bagi penulis dalam pembuatan skripsi ini.
2. Bapak Hastari Utama, M.Cs selaku dosen yang membantu dalam penelitian skripsi ini.
3. Ibu Hartatik, ST, M.Cs selaku dosen yang selalu membantu dalam saran, membantu pembentukan tema dalam penelitian dan motivasi dalam skripsi ini.
4. Fregy Damara, Ganang Yoga Kusuma, dan teman – teman yang selalu menemani dan turut berpartisipasi dalam pengembangan penelitian ini.

5. Para Dosen dan Staff STMIK AMIKOM Yogyakarta yang telah banyak memberikan ilmu pengetahuan, pengalaman, dan bantuannya selama penulis kuliah hingga terselesaikannya skripsi ini.
6. Ibu, Bapak, dan Saudara – saudara penulis yang telah memberikan dukungan baik moril ataupun materil.
7. Teman-teman kuliah di kelas S1 Transfer 2015, terutama Yusuf, Via, Fitri, Ega, Dea, Ronggo, Aga dan Ardi.
8. Semua pihak yang telah banyak membantu dalam penyusunan skripsi ini yang tidak dapat disebutkan satu persatu.

Penulis tentunya menyadari bahwa pembuatan skripsi ini masih banyak sekali kekurangan – kekurangan dan kelemahan – kelemahannya. Oleh karena itu penulis berharap kepada semua pihak agar dapat menyampaikan kritik dan saran yang membangun untuk menambah kesempurnaan skripsi ini. Semoga skripsi ini dapat bermanfaat bagi pihak terkait dan pembaca pada umumnya.

Tanggal, 30 Agustus 2017

Wahyu Setiawan

15.21.0845

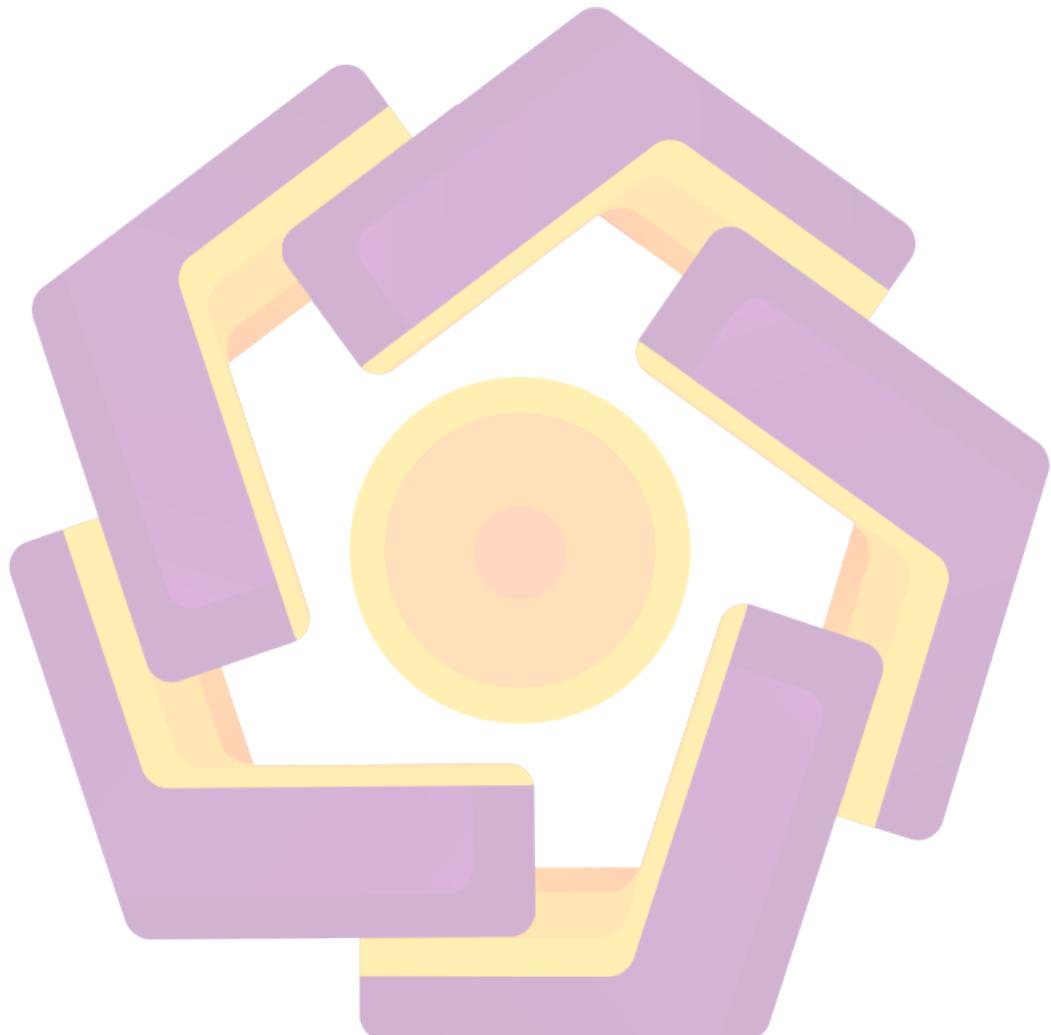
DAFTARI ISI

HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN	iv
MOTTO	v
PERSEMAHAN	vi
KATA PENGANTAR	vii
DAFTARI ISI	ix
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xvi
INTISARI	xviii
ABSTRACT	xix
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penulisan	3
1.5 Manfaat Penulisan	4
1.6 Metode Pengumpulan Data	4
1.6.1 Studi pustaka	4
1.6.2 Interview	4
1.6.3 Survey	5
1.6.4 Ekperimen (Percobaan)	5
1.7 Sistematika Penulisan.....	5
BAB II LANDASAN TEORI	7
2.1 Tinjauan Pustaka	7
2.2 Pengertian Data Mining.....	10
2.3 Pengertian Sentiment Analysis.....	10

2.4	Pengertian Pre-Processing	12
2.5	Term Frequency-Inverse Document Frequency (TF-IDF)	12
2.6	Pengertian Support Vector Machine	13
2.7	Multi Class Support Vector Machine	16
2.7.1	Metode “one-against-all”	16
2.7.2	Metode “one-against-one”	18
2.8	Pengertian Feature Selection	19
2.9	Pengertian Particle Swarm Optimization	20
2.10	Cosine Similarities	22
2.11	K-Fold Cross Validation	23
2.12	Confusion Matix	24
2.13	UML (Unified Modeling Language)	24
2.13.1	Use Case Diagram	25
2.13.2	Class Diagram	25
2.13.3	Activity Diagram	27
2.13.4	Squence Diagram	28
2.14	Langkah – langkah Pengembangan Aplikasi	28
BAB III	PERANCANGAN	34
3.1	Analisis Masalah	34
3.2	Solusi Yang Dipilih	35
3.3	Analisis Kebutuhan	36
3.3.1	Analisis Kebutuhan Fungsional	36
3.3.2	Analisis Kebutuhan Non-Fungsional	37
3.4	Analisa Algoritma	38
3.4.1	Cleaning Data	38
3.4.2	Tokenization	40
3.4.3	Stopwords	41
3.4.4	Steamming	42
3.4.5	TF-IDF	43
3.4.6	Consine Similarity	51

3.4.7	Support Vector Machine	59
3.4.8	Particle Swarm Optimization	63
3.4.9	K-Fold Cross Validation	66
3.5	Perancangan Aplikasi	70
3.5.1	Perancangan Use Case Diagram	70
3.5.2	Perancangan Activity Diagram	78
3.5.3	Perancangan Class Diagram.....	82
3.5.4	Perancangan Sequence Diagram	83
3.6	Perancangan Basis Data	88
3.6.1	Perancangan Database.....	88
3.6.2	Perancangan File CSV	89
3.6.3	Perancangan File Json	90
3.7	Perancangan Antarmuka Pengguna.....	92
3.7.1	Perancangan pada Aplikasi Download Data Tweet	92
3.7.2	Perancangan Pada Aplikasi Training Data.....	93
3.7.3	Perancangan Pada Aplikasi Web Training.....	93
BAB IV IMPLEMENTASI	96
4.1	Implementasi Sistem	96
4.1.1	Pembuatan Database	96
4.1.2	Pembuatan Data Training.....	98
4.1.3	Pembuatan Interface.....	99
4.2	Black-box Testing	109
4.3	White-box Testing.....	113
4.4	Manual Program	124
4.4.1	Aplikasi Download Data Twitter	124
4.4.2	Aplikasi Training Data.....	126
4.4.3	Aplikasi Klasifikasi Data Twitter.....	131
4.5	Hasil Implementasi	136
4.5.1	Evaluasi Optimasi Particle Swarm Optimization.....	138
4.5.2	Evaluasi Model Terhadap Data Testing	145

BAB V KESIMPULAN	149
5.1 Kesimpulan.....	149
5.2 Saran	150
DAFTAR PUSTAKA	151



DAFTAR GAMBAR

Gambar 2.1 Sample Penyebaran Pemasalahan Dua Dimensi	14
Gambar 2.2 Hyperplane pada support vector machine	14
Gambar 2.3 Bagan metode “one-against-all”.....	17
Gambar 2.4 Bagan metode “one-against-one”.....	19
Gambar 2.5 Use Case Diagram.....	25
Gambar 2.6 Class diagram	26
Gambar 2.7 Activity Diagram.....	27
Gambar 2.8 Squence Diagram	28
Gambar 2.9 Alur atau Langkah - Langkah Analisa Sentimen	29
Gambar 2.10 Langkah – Langkah Pre-Processing.....	29
Gambar 2.11 Langkah – Langkah Training Data.....	31
Gambar 2.12 Langkah – Langkah Testing Data	33
Gambar 3.1 Use Case Diagram.....	70
Gambar 3.2 Activiry Download Data Tweet	78
Gambar 3.3 Activity Training Data	79
Gambar 3.4 Activity Model Presistence	79
Gambar 3.5 Activity Import Data Tweet	80
Gambar 3.6 Acitivy Prediksi Tweet.....	80
Gambar 3.7 Activity Upload Model Presistence.....	81
Gambar 3.8 Activity Upload Data Tweet	81
Gambar 3.9 Class Diagram Training Data.....	82
Gambar 3.10 Class Diagram Prediksi Data	83
Gambar 3.11 Sequence Diagram Predik Data Tweet.....	84
Gambar 3.12 Sequence Diagram Upload File Tweet.....	85
Gambar 3.13 Sequence Import File Stopwords	86
Gambar 3.14 Sequence Export Training Data	86
Gambar 3.15 Sequence Training Data	87

Gambar 3.16 Isi yang Terdapat dalam File CSV Training	89
Gambar 3.17 Isi yang terdapat dalam File CSV Prediksi	90
Gambar 3.18 File JSON Kofigurasi	91
Gambar 3.19 File JOSN Twitter	91
Gambar 3.20 User Inteface Download Data Tweet	92
Gambar 3.21 User Interface Training Data.....	93
Gambar 3.22 User Inteface Aplikasi Web Training.....	94
Gambar 3.23 User Interface Upload File Model Presistence.....	95
Gambar 3.24 User Interface Hasil Prediksi.....	95
Gambar 4.1 Script Class ModelTraining	97
Gambar 4.2 Script Class Setting	97
Gambar 4.3 Script Class TestingData	97
Gambar 4.4 Script Class FeatureList	98
Gambar 4.5 Script Class Stopwords	98
Gambar 4.6 Data Training yang Telah Dilabel.....	98
Gambar 4.7 Data Testing	99
Gambar 4.8 Interface Download Data Twitter.....	100
Gambar 4.9 Interface Download Data Twitter Setelah Proses Download.....	101
Gambar 4.10 Interface Training Data	102
Gambar 4.11 Interface Training Data Insert Dataset	102
Gambar 4.12 Interface Training Data Insert Stopword.....	103
Gambar 4.13 User Interface Sistem Prediksi Data Twitter Halaman Utama.....	104
Gambar 4.14 User Interface Sistem Prediksi Data Twitter Halaman Model Training	105
Gambar 4.15 User Interface Sistem Prediksi Data Twitter Halaman Feature List	106
Gambar 4.16 User Interface Sistem Prediksi Data Twitter Halaman Stopwords	107
Gambar 4.17 User Interface Sistem Prediksi Data Twitter Daftar Testing.....	108
Gambar 4.18 User Interface Sistem Prediksi Data Twitter Halaman Daftar Testing	109

Gambar 4.19 Script Cleaning Data	114
Gambar 4.20 Script Tokenization	115
Gambar 4.21 Script Stopwords	115
Gambar 4.22 Script Steamming	116
Gambar 4.23 Script Feature List	117
Gambar 4.24 Hasil Feature List	118
Gambar 4.25 Script TF-IDF.....	119
Gambar 4.26 Script Support Vector Machine.....	120
Gambar 4.27 Script <i>Particle Swarm Optimization</i> 1	121
Gambar 4.28 Script <i>Particle Swarm Optimization</i> 2	122
Gambar 4.29 Script <i>Particle Swarm Optimization</i> 3	122
Gambar 4.30 Script <i>Particle Swarm Optimization</i> 4	123
Gambar 4.31 Isi Dalam File Conf.json	124
Gambar 4.32 Tampilan Download Data Twitter Pertama Dijalankan	125
Gambar 4.33 Tampilan Download Data Twitter Setelah Proses	126
Gambar 4.34 Tampilan Pada Dataset Training	127
Gambar 4.35 Tampilan Memilih File Dataset Training.....	128
Gambar 4.36 Tampilan Memilih Data Stopwords	128
Gambar 4.37 Tampilan Bila Semua File Telah Dipilih	129
Gambar 4.38 Tampilan Bila Proses Training Data Telah Selesai.....	130
Gambar 4.39 Tampilan Hasil Run Server Django	131
Gambar 4.40 Tampilan Halaman Model Testing.....	132
Gambar 4.41 Tampilan Halaman Feature List	133
Gambar 4.42 Tampilan Halaman Stopwords	134
Gambar 4.43 Tampilan Halaman Selesai Proses Klasifikasi	135
Gambar 4.44 Tampilan Hasil yang Tersimpan Dalam Halaman Daftar Testing	136
Gambar 4.45 Hasil Data Training 300 data Twitter yang telah didapat	137
Gambar 4.46 Perbandingan SVM Sebelum dan Sesudah PSO.....	138
Gambar 4.47 Grafik Evaluasi Perbandingan	144
Gambar 4.48 Perbandingan Hasil Akurasi Training dan Akurasi Testing.....	145

DAFTAR TABEL

Tabel 2.1 Perbandingan Penelitian Terkait Dengan Sistem Yang Dibangun	9
Tabel 2.2 Contoh Metode “one-against-all”	17
Tabel 2.3 Contoh Metode “one-against-one”.....	18
Tabel 2.4 <i>Confusion Matrix</i>	24
Tabel 3.1 Data Tweet Sebelum Dilakukan Cleaning Data	38
Tabel 3.2 Data Setelah Dilakukan Cleaning Data.....	39
Tabel 3.3 Hasil Tokenization Dari Cleaning Data	40
Tabel 3.4 Hasil Stopword Dari Tokenization	41
Tabel 3.5 Hasil Stemming Dari Stopwords	42
Tabel 3.6 Feature List	43
Tabel 3.7 <i>Term Frequency</i> Positive	45
Tabel 3.8 <i>TF-TDF</i> Positive	47
Tabel 3.9 <i>TF-TDF</i> Negative.....	49
Tabel 3.10 <i>TF-TDF</i> Neutral	50
Tabel 3.11 <i>Consine Similarity</i> Feature Positive.....	53
Tabel 3.12 <i>Consine Similarity</i> Feature Negative	55
Tabel 3.13 <i>Consine Similarity</i> Feature Netral.....	57
Tabel 3.14 <i>Consine Similarity</i>	58
Tabel 3.15 Perhitungan Persamaan Positif Terhadap Training Data Positif.....	61
Tabel 3.16 Persamaan w_1 , w_2 w_3 dan b Pada Setiap Label	61
Tabel 3.17 Perhitungan Persamaan Positif Terhadap Training Data Negatif	62
Tabel 3.18 Perhitungan Persamaan Positif Terhadap Training Data Netral	62
Tabel 3.19 Predik Pada Data Training	63
Tabel 3.20 Perubahan Bobot Parameter SVM Dengan PSO	65
Tabel 3.21 Presik Pada Data Dengan Parameter PSO	66
Tabel 3.22 Fold Pertama	66
Tabel 3.23 Fold Kedua.....	67

Tabel 3.24 Fold Ketiga.....	67
Tabel 3.25 <i>Confusion Matrix</i> Pada Fold Pertama	68
Tabel 3.26 <i>Confusion Matrix</i> Pada Fold Kedua.....	69
Tabel 3.27 <i>Confusion Matrix</i> Pada Fold Ketiga.....	69
Tabel 3.28 <i>Use Case Description</i> Download Data Tweet	71
Tabel 3.29 <i>Use Case Description</i> Training Data	72
Tabel 3.30 <i>Use Case Description</i> Predik Tweet	73
Tabel 3.31 <i>Use Case Description</i> Import Data Tweet	74
Tabel 3.32 <i>Use Case Description</i> Export Model Presistence	75
Tabel 3.33 <i>Use Case Description</i> Upload Data Tweet	76
Tabel 3.34 <i>Use Case Description</i> Upload Model Presistence	77
Tabel 3.35 Detail Tabel modelData	88
Tabel 3.36 Detail Tabel Setting	88
Tabel 3.37 Detail Tabel TestingData	89
Tabel 4.1 Pengujian <i>blackbox</i> sistem download data Twitter.....	109
Tabel 4.2 Pengujian <i>blackbox</i> sistem training data Twitter	110
Tabel 4.3 Pengujian <i>blackbox</i> sistem prediksi halaman utama.....	110
Tabel 4.4 Pengujian <i>blackbox</i> sistem prediksi halaman model.....	111
Tabel 4.5 Pengujian <i>blackbox</i> prediksi halaman feature.....	112
Tabel 4.6 Pengujian <i>blackbox</i> prediksi halaman daftar testing.....	112
Tabel 4.7 Confusion Matrix SVM	139
Tabel 4.8 Confusion Matrix SVM Dengan PSO.....	139
Tabel 4.9 Table Hasil Cross Validation SVM Dengan PSO.....	140
Tabel 4.10 Akurasi Mean Cross Validation SVM dengan PSO	144
Tabel 4.11 <i>Confusion Matrix</i> Prediksi Testing Model SVM	146
Tabel 4.12 <i>Confusion Matrix</i> Prediksi Testing Model SVM + PSO	147
Tabel 4.13 Laporan Klasifikasi Data Testing Prediksi	147
Tabel 4.14 Perbandingan Feature List Pada Training Tweet dan Testing	148

INTISARI

Dengan perkembangan teknologi yang cepat pada saat ini, terdapat banyak informasi yang tersaji melalui media sosial. Salah satunya merupakan twitter dalam menyampaikan berita terhadap teman, kerabat, dan media dengan cepat dan mudah. Melalui text mining maka menambah atau mengumpulkan statement yang terdapat dalam media twitter yang dijadikan bahan utama untuk sentiment analysis.

Pendekatan support vector machine merupakan suatu pendekatan dengan machine learning yang memprediksi kelas berdasarkan training. Pendekatan Particle Swarm Optimization merupakan pendekatan yang kelompokan data berdasarkan sifat dari setiap particle tersebut yang dapat mendukung dari Feature Selection.

Feature Selection adalah suatu kegiatan yang umumnya bisa dilakukan secara preprocessing dan bertujuan untuk memilih feature yang berpengaruh dan mengesampingkan Feature yang tidak berpengaruh dalam suatu kegiatan pemodelan atau penganalisaan data. Sehingga pendapat itu dapat digolongkan pada bisa digolongkan dalam negatif, positif, dan netral berdasarkan sifat dari masing-masing particle.

Kata Kunci : Support Vector Machine, Particle Swarm Optimization, Data Mining, Kecerdasan Buatan, Sistem, Optimalisasi, Sentiment Analysis.

ABSTRACT

With the rapid technological developments at this time, there is a lot of information presented through social media. The social media is allow us to send a news to friends, relatives, media with ease of use. Examples such as the use of Twitter. Through text mining, we can collect statements contained in Twitter as the main material for sentiment analysis.

Therefore, Support vector machine is one of many approaches in machine learning to predict sentiment classification task. Particle Swarm Optimization is an approach to classify data based on properties of each particle which can support selection task in the Feature Selection Process.

Feature Selection is an activity that generally can be done by preprocessing. Main task of this process is to choose the features that effect and override a feature where the features does not affect in any activity modeling or analyzing data. So that the opinion can be identified can be classified in the negative, positive and neutral based on the properties of each particle.

Keywords : Support Vector Machine, Particle Swarm Optimization, Data Mining, Kecerdasan Buatan, Sistem, Optimalisasi, Sentiment Analysis.

