

**OPTIMASI PARAMETER KERNEL RADIAL BASIS
FUNCTION PADA SUPPORT VECTOR MACHINE
DENGAN GENETIC ALGORITHM UNTUK
MELAKUKAN SENTIMENT ANALYSIS (STUDI
KASUS: KOMENTAR MASYARAKAT PADA
INSTAGRAM UNIVERSITAS PENDIDIKAN
GANESHA)**



**OLEH
I PUTU DEDY WIRA DARMAWAN
NIM 1715051017**

**PROGRAM STUDI PENDIDIKAN TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KEJURUAN
UNIVERSITAS PENDIDIKAN GANESHA
SINGARAJA
2022**

**OPTIMASI PARAMETER KERNEL RADIAL BASIS
FUNCTION PADA SUPPORT VECTOR MACHINE
DENGAN GENETIC ALGORITHM UNTUK
MELAKUKAN SENTIMENT ANALYSIS (STUDI
KASUS: KOMENTAR MASYARAKAT PADA
INSTAGRAM UNIVERSITAS PENDIDIKAN
GANESHA)**

SKRIPSI

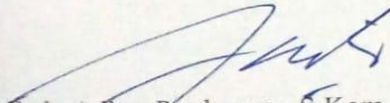


**PROGRAM STUDI PENDIDIKAN TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KEJURUAN
UNIVERSITAS PENDIDIKAN GANESHA
SINGARAJA
2022**

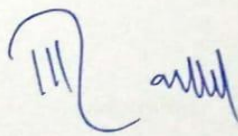
SKRIPSI
DIAJUKAN UNTUK MELENGKAPI TUGAS DAN
MEMENUHI SYARAT - SYARAT UNTUK MENCAPAI
GELAR SARJANA PENDIDIKAN

Menyetujui,

Pembimbing I


Gede Aditra Pradnyana, S.Kom., M.Kom.
NIP. 198901192015041004

Pembimbing II

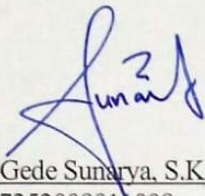

Ida Bagus Nyoman Pascima, S.Pd., M.Cs.
NIP. 198907132019031017

Skripsi Oleh I Putu Dedy Wira Darmawan

Telah dipertahankan di depan penguji

Pada Tanggal: 07 Oktober 2022

Dewan Penguji,



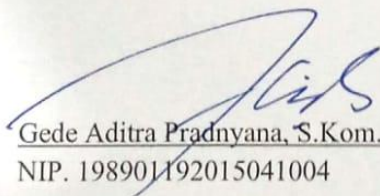
Dr. I Made Gede Sunarya, S.Kom., M.Cs.
NIP. 198307252008011008

(Ketua)



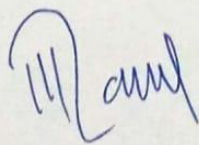
I Gede Mahendra Darmawiguna, S.Kom., M.Sc.
NIP. 198501042010121004

(Anggota)



Gede Aditra Pradnyana, S.Kom., M.Kom.
NIP. 198901192015041004

(Anggota)



Ida Bagus Nyoman Pascima, S.Pd., M.Cs.
NIP. 198907132019031017

(Anggota)

Diterima oleh Panitia Ujian Fakultas Teknik dan Kejuruan
Universitas Pendidikan Ganesha Singaraja
Guna memenuhi syarat-syarat untuk mencapai Gelar Sarjana Pendidikan

Pada:

Hari : *Senin*

Tanggal : *05 Desember 2022*

Mengetahui,

Ketua Ujian,



Dr. Ketut Agustini, S.Si., M.Si
NIP. 19740801 200003 2 001

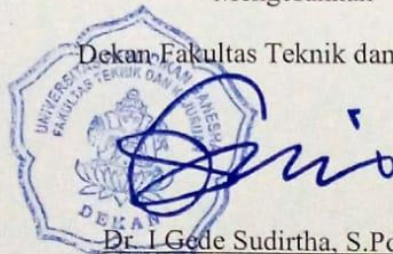
Sekretari Ujian,



Dr. phil. Dessy Seri Wahyuni, S.Kom., M.Eng.
NIP. 19850215 200812 2 007

Mengesahkan

Dekan Fakultas Teknik dan Kejuruan



Dr. I Gede Sudirtha, S.Pd., M.Pd
NIP. 19710616 199602 1 001

LEMBAR PERNYATAAN

Dengan ini, saya menyatakan bahwa karya tulis yang berjudul “**OPTIMASI PARAMETER KERNEL *RADIAL BASIS FUNCTION* PADA *SUPPORT VECTOR MACHINE* DENGAN *GENETIC ALGORITHM* UNTUK MELAKUKAN *SENTIMENT ANALYSIS* (STUDI KASUS: KOMENTAR MASYARAKAT PADA *INSTAGRAM* UNIVERSITAS PENDIDIKAN GANESHA)**” beserta seluruh isinya adalah benar karya sendiri dan saya tidak melakukan penjiplakan dan pengutipan dengan cara - cara yang tidak sesuai dengan etika yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini, saya siap menanggung resiko/sanksi yang dijatuhkan kepada saya apabila dikemudian hari ditemukan adanya pelanggaran atas etika keilmuan dalam karya saya ini atau ada klaim terhadap keaslian karya saya ini.

Singaraja, 07 Oktober 2022
Yang Membuat Pernyataan,



I Putu Dedy Wira Darmawan
NIM. 1715051017

KATA PERSEMBAHAN

TERIMAKASIH YANG TERAMAT DALAM SAYA UCAPKAN KEPADA:

IDA SANG HYANG WIDHI WASA

Atas Berkat dan Rahmat Beliau, Penulis Bisa Menyelesaikan Skripsi Ini. Atas Segala Kemudahan dan Kelancaran yang Diberikan.

SKRIPSI INI SAYA DEDIKASIKAN KEPADA:

ORANG TUA

Yang telah membesarkan, mendidik dengan kesabaran dan keikhlasan dan selalu memberikan semangat, dukungan, motivasi, serta doa dalam setiap langkah saya dalam menempuh kehidupan.

KELUARGA

Yang selalu menghibur, memotivasi, dan menyemangati dalam mengerjakan skripsi. Terimakasih atas doa yang telah diberikan

SELURUH STAF DOSEN PRODI PENDIDIKAN TEKNIK INFORMATIKA

Yang telah membimbing dalam menyelesaikan skripsi ini khususnya pembimbing saya, Bapak Gede Aditra Pradnyana dan Bapak Ida Bagus Nyoman Pascima.

TEMAN-TEMAN PRODI PENDIDIKAN TEKNIK INFORMATIKA

Seluruh rekan mahasiswa Prodi Pendidikan Teknik Informatika Angkatan 2017.

MOTTO



“Great Minds Discuss Idea”

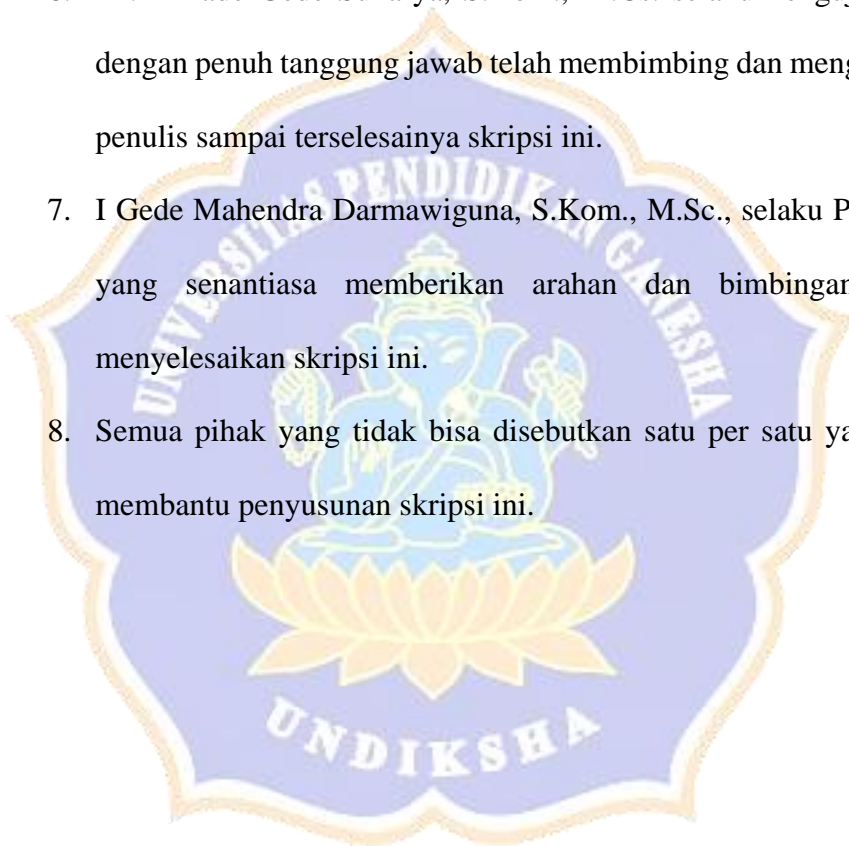
PRAKATA

Puji syukur penulis panjatkan kehadirat Ida Sang Hyang Widi Wasa, Tuhan Yang Maha Esa, karena berkat dan rahmatNya, saya selaku penulis dapat menyelesaikan skripsi yang berjudul **“OPTIMASI PARAMETER KERNEL RADIAL BASIS FUNCTION PADA SUPPORT VECTOR MACHINE DENGAN GENETIC ALGORITHM UNTUK MELAKUKAN SENTIMENT ANALYSIS (STUDI KASUS: KOMENTAR MASYARAKAT PADA INSTAGRAM UNIVERSITAS PENDIDIKAN GANESHA)”**. Tujuan dari penulisan skripsi ini adalah untuk memenuhi syarat yang harus dipenuhi dalam menyelesaikan program sarjana Pendidikan Teknik Informatika, Universitas Pendidikan Ganesha. Dalam proses pengerjaan skripsi ini tidak terlepas dari dukungan, motivasi, semangat, bantuan dan kerjasama dari berbagai pihak. Maka pada kesempatan ini, penulis ingin menyampaikan ucapan terimakasih serta penghargaan sebagai rasa syukur dan hormat penulis kepada:

1. Prof. Dr. I Nyoman Jampel, M.Pd., selaku Rektor Universitas Pendidikan Ganesha.
2. Dr. I Gede Sudirtha, S.Pd., M.Pd., selaku Dekan Fakultas Teknik dan Kejuruan Universitas Pendidikan Ganesha.
3. Gede Saindra Santyadiputra, S.T., M.Cs., selaku Koordinator Program Studi Pendidikan Teknik Informatika yang telah memberikan pengarahan baik selama perkuliahan maupun dalam penyusunan skripsi ini.
4. Gede Aditra Pradnyana, S.Kom., M.Kom. selaku Pembimbing I yang telah banyak meluangkan waktu, tenaga, dan pikiran serta

penuh kesabaran dalam memberikan bimbingan kepada penulis sehingga skripsi ini dapat diselesaikan.

5. Ida Bagus Nyoman Pascima, S.Pd., M.Cs., selaku Pembimbing II yang juga telah memberikan motivasi, bimbingan, dan waktu yang telah diluangkan ditengah-tengah kesibukkan beliau kepada penulis sehingga skripsi ini dapat terselesaikan.
6. Dr. I Made Gede Sunarya, S.Kom., M.Cs. selaku Penguji I yang dengan penuh tanggung jawab telah membimbing dan mengarahkan penulis sampai terselesainya skripsi ini.
7. I Gede Mahendra Darmawiguna, S.Kom., M.Sc., selaku Penguji II yang senantiasa memberikan arahan dan bimbingan dalam menyelesaikan skripsi ini.
8. Semua pihak yang tidak bisa disebutkan satu per satu yang telah membantu penyusunan skripsi ini.



**OPTIMASI PARAMETER KERNEL *RADIAL BASIS FUNCTION* PADA
SUPPORT VECTOR MACHINE DENGAN *GENETIC ALGORITHM*
UNTUK MELAKUKAN *SENTIMENT ANALYSIS* (STUDI KASUS:
KOMENTAR MASYARAKAT PADA *INSTAGRAM* UNIVERSITAS
PENDIDIKAN GANESHA)**

Oleh

I Putu Dedy Wira Darmawan, NIM 1715051017

Program Studi Pendidikan Teknik Informatika

Jurusan Teknik Informatika

Fakultas Teknik dan Kejuruan

Universitas Pendidikan Ganesha

Singaraja

Email: iputudedywiradarmawan28@undiksha.ac.id

ABSTRAK

Universitas Pendidikan Ganesha memanfaatkan media sosial khususnya Instagram sebagai media informasi dan *branding*. Media sosial ini membuat pengguna lain dapat berinteraksi, beropini, dan memberikan penilaian secara terbuka kepada Universitas Pendidikan Ganesha dengan cara memberikan komentar secara langsung. Kumpulan komentar ini dapat dimanfaatkan menjadi sebuah informasi yang berguna menggunakan *sentiment analysis*. *Sentiment analysis* adalah bidang studi yang menganalisis opini, sentimen, emosi, penilaian, dan sikap seseorang terhadap suatu produk atau sebuah kebijakan. Metode yang umum digunakan untuk melakukan analisis sentimen adalah klasifikasi menggunakan *machine learning*. Salah satu *machine learning* yang sering digunakan adalah *Support Vector Machine (SVM)*. Namun, pada masalah yang bersifat non-linear seperti analisis sentimen ini, *SVM* memerlukan kernel untuk memetakan vektor ke dalam ruang berdimensi tinggi agar dapat menyelesaikan permasalahan non-linear. Tantangan yang dihadapi dalam menggunakan kernel adalah memilih parameter yang optimal untuk model klasifikasi agar dapat menghasilkan model klasifikasi yang baik. Tujuan dari penelitian ini adalah mencari parameter yang paling optimal untuk *SVM* menggunakan *Genetic Algorithm*. Penelitian ini menjelaskan bagaimana merancang sebuah model klasifikasi *SVM-GA* mulai dari tahap pengumpulan data, pengolahan data, klasifikasi, hingga evaluasi. Hasil penelitian menunjukkan bahwa akurasi terbaik yang dihasilkan dengan parameter yang dioptimasi dengan algoritma genetika adalah 81,6%. Hasil ini lebih baik dari menggunakan parameter default *SVM-RBF* yang hanya mendapatkan akurasi 74,4% dan pemilihan parameter acak secara manual yang menghasilkan 79,2%.

Kata Kunci: Analisis Sentimen, *Support Vector Machine*, Algoritma Genetika

**PARAMETERS OPTIMIZATION OF RADIAL BASIS FUNCTION
KERNEL ON SUPPORT VECTOR MACHINE USING GENETIC
ALGORITHM TO PERFORM SENTIMENT ANALYSIS (CASE STUDY:
PUBLIC COMMENTS ON INSTAGRAM UNIVERSITY OF EDUCATION
GANESHA)**

By

I Putu Dedy Wira Darmawan, NIM 1715051017

Study Program of Informatic Education

Departement of Informatics

Faculty of Engineering and Vocational

Ganesha Education University

Email: iputudedywiradarmawan28@undiksha.ac.id

ABSTRACT

Ganesha Education University uses social media, especially Instagram as a medium of information and branding. This social media allows other users to interact, have opinions, and give an open assessment to Ganesha Education University by providing direct comments. This collection of comments can be used as useful information using sentiment analysis. Sentiment analysis is a field of study that analyzes a person's opinions, sentiments, emotions, judgments, and attitudes towards a product or a policy. The method commonly used to perform sentiment analysis is classification using machine learning. One machine learning that is often used is the Support Vector Machine (SVM). However, in non-linear problems such as sentiment analysis, SVM requires a kernel to map vectors into high-dimensional space in order to solve non-linear problems. The challenge faced in using the kernel is to choose the optimal parameters for the classification model in order to produce a good classification model. The purpose of this research is to find the most optimal parameters for SVM using Genetic Algorithm. This study explains how to design an SVM-GA classification model starting from the stages of data collection, data processing, classification, to evaluation. The results showed that the best accuracy produced with the parameters optimized by the genetic algorithm was 81.6%. This result is better than using the default parameter SVM-RBF which only gets 74.4% accuracy and manual random parameter selection which produces 79.2%.

Keywords: Sentiment Analysis, Support Vector Machine, Genetic Algorithm

DAFTAR ISI

	HALAMAN
SKRIPSI	ii
LEMBAR PERNYATAAN	vi
KATA PERSEMBAHAN	vii
MOTTO	viii
PRAKATA	ix
ABSTRAK	xi
ABSTRACT	xii
DAFTAR ISI	xiii
DAFTAR TABEL	xvii
DAFTAR GAMBAR	xix
DAFTAR LAMPIRAN	xxi
BAB I PENDAHULUAN	1
1.1 LATAR BELAKANG	1
1.2 RUMUSAN MASALAH.....	6
1.3 TUJUAN PENELITIAN.....	7
1.4 BATASAN PENELITIAN	7
1.5 MANFAAT PENELITIAN.....	7
BAB II TINJAUAN PUSTAKA	9
2.1 KAJIAN PUSTAKA.....	9

2.2 LANDASAN TEORI.....	22
2.2.1 Data Mining	22
2.2.2 Sentiment Analysis	23
2.2.3 Support Vector Machine	24
2.2.4 Genetic Algorithm.....	26
2.2.5 TF-IDF (Term Frequency-Inverse Document Frequency)	32
2.2.6 K-Folds Cross Validation.....	33
2.2.7 Confusion Matrix	34
BAB III METODE PENELITIAN	38
3.1 JENIS PENELITIAN	38
3.2 TAHAPAN PENELITIAN	38
3.2.1 Pengumpulan Data (<i>Selection</i>).....	43
3.2.2 Pre-processing Data	43
3.2.2.1 Case Folding	44
3.2.2.2 Tokenization.....	44
3.2.2.3 Stopword Removal.....	44
3.2.2.4 Stemming	45
3.2.3 Pembobotan Kata (<i>Transformation</i>)	45
3.2.4 Klasifikasi (<i>Data Mining</i>)	47
3.2.4.1 Proses Training	48
3.2.4.2 Proses Testing	54

3.2.5 Evaluasi (<i>Evaluation</i>).....	54
3.2.5.1 K-fold Cross Validation	54
3.2.5.2 Confusion Matrix	57
BAB IV HASIL DAN PEMBAHASAN	59
4.1 HASIL PENELITIAN.....	59
4.1.1 Hasil Pengumpulan Data (<i>Selection</i>)	59
4.1.2 Hasil Pre-Processing Data.....	67
4.1.2.1 Case Folding	67
4.1.2.2 Remove Number	68
4.1.2.3 Tokenization.....	69
4.1.2.3 Stopword Removal.....	70
4.1.2.4 Stemming	71
4.1.3 Hasil Pembobotan Kata (<i>Transformation</i>).....	73
4.1.4 Hasil Implementasi Klasifikasi SVM-GA	80
4.1.5 Hasil Evaluasi (<i>Evaluation</i>)	94
4.2 PEMBAHASAN	106
BAB V PENUTUP	117
5.1 SIMPULAN	117
5.2 SARAN	119
DAFTAR PUSTAKA	120
DAFTAR RIWAYAT HIDUP	124



DAFTAR TABEL

Tabel	Halaman
Tabel 2. 1 Penelitian Terkait	12
Tabel 2. 2 Confusion Matrix	35
Tabel 2. 3 Istilah dalam Confusion Matrix	35
Tabel 3. 1 Sampel Data Sebelum Pre-Processing	46
Tabel 3. 2 Sampel Data Setelah Pre-processing	46
Tabel 3. 3 Contoh Variabel Individu	49
Tabel 3. 4 Populasi Generasi Kedua Hasil dari Proses GA	53
Tabel 3. 5 Skenario Pengujian 10-fold Cross Validation.....	55
Tabel 3. 6 Contoh Hasil Klasifikasi dari Data Testing	57
Tabel 4. 1 Sampel Dataset.....	65
Tabel 4. 2 Contoh Case Folding.....	68
Tabel 4. 3 Contoh Remove Number	69
Tabel 4. 4 Contoh Tokenization.....	70
Tabel 4. 5 Contoh Stopwords Removal	71
Tabel 4. 6 Contoh Stemming	72
Tabel 4. 7 Contoh Sampel TF (Term Frequency)	75
Tabel 4. 8 Contoh Sampel IDF (Inverse Document Frequency)	76
Tabel 4. 9 Contoh Sampel TF-IDF	77
Tabel 4. 10 Referensi Penelitian Terdahulu dalam Menentukan Parameter GA (Hassanat et al., 2019).....	84
Tabel 4. 11 Testing Scenario Pencarian Parameter Terbaik Algoritma Genetika	85
Tabel 4. 12 Percobaan SVM-GA dengan Beberapa Kombinasi Parameter GA ...	87
Tabel 4. 13 Perbandingan Hasil SVM-GA Berdasarkan Jumlah Population Size	91
Tabel 4. 14 Percobaan SVM-GA dengan Beberapa Nilai Population Size	93

Tabel 4. 15 Model Confusion Matrix 3x3.....	94
Tabel 4. 16 Percobaan SVM-RBF Menggunakan Parameter Acak Secara Manual	101
Tabel 4. 17 Perbandingan Performa SVM-RBF dengan optimasi GA dan SVM- RBF tanpa Optimasi.....	105



DAFTAR GAMBAR

Gambar	Halaman
Gambar 2. 1 Ilustrasi Support Vector Machine	25
Gambar 2. 2 Ilustrasi Kernel Trick	26
Gambar 2. 3 Ilustrasi Gen, Allele, Kromosom, Individu, dan Populasi dalam Algoritma Genetika.....	28
Gambar 2. 4 Ilustrasi K-Folds Cross Validation.....	34
Gambar 3. 1 Siklus Knowledge Discovery in Database (KDD).....	40
Gambar 3. 2 Flowchart Metode Penelitian	41
Gambar 3. 3 Ilustrasi Seleksi Individu pada Tournament Selection.....	50
Gambar 3. 4 Whole Arithmetic Crossover dengan $\alpha = 0,4$	52
Gambar 3. 5 Proses Mutasi	53
Gambar 4. 1 Kode Program untuk Scrapping Data	63
Gambar 4. 2 Akun Instagram Universitas Pendidikan Ganesha	64
Gambar 4. 3 Kode Program Pembacaan Dataset	66
Gambar 4. 4 Kode Program Menampilkan Label Dataset ke Bentuk Grafik	66
Gambar 4. 5 Perbandingan jumlah dataset kelas positif (2), kelas netral (1), dan kelas negatif (0).....	66
Gambar 4. 6 Kode Program Case Folding	67
Gambar 4. 7 Kode Program Remove Number.....	68
Gambar 4. 8 Kode Program Tokenization	69
Gambar 4. 9 Kode Program Stopword Removal	71
Gambar 4. 10 Kode Implementasi Stemming	72
Gambar 4. 11 Sampel Hasil Pre-Processing	73
Gambar 4. 12 Visualisasi Frekuensi Kata yang Sering dengan Wordcloud	74
Gambar 4. 13 Kode Program Pembobotan Kata dengan TF-IDF.....	78
Gambar 4. 14 Sampel Hasil Pembobotan Kata dengan TF-IDF.....	79
Gambar 4. 15 Nilai TF-IDF pada Setiap Token (Kata)	79
Gambar 4. 16 Kode Program Split Data Training dan Data Testing	80

Gambar 4. 17 Kode Program 10-Folds Cross Validation	80
Gambar 4. 18 Kode Program Implementasi SVM-GA.....	83
Gambar 4. 19 Kode Program Confusion Matrix.....	95
Gambar 4. 20 Kode Program Pencarian Accuracy, Precission, Recall, dan F1-Score.....	96
Gambar 4. 21 Confusion Matrix dari Parameter Terbaik SVM-RBF + GA.....	97
Gambar 4. 22 Performa dari Parameter Terbaik SVM-RBF + GA	97
Gambar 4. 23 Grafik Evolusi Algoritma Genetika	98
Gambar 4. 24 Confusion Matrix dari SVM-RBF dengan Nilai Parameter Default	99
Gambar 4. 25 Performa SVM-RBF dengan Nilai Parameter Default.....	100
Gambar 4. 26 Confusion Matrix dari SVM-RBF dengan Input Parameter Manual (C=10, $\gamma=0.1$).....	103
Gambar 4. 27 Performa SVM-RBF dengan Input Parameter Manual (C=10, $\gamma=0.1$)	103
Gambar 4. 28 Bentuk Data Tidak Linear. Kelas (0) Negatif, (1) Netral, dan (2) Positif	109
Gambar 4. 29 Ilustrasi Pengaruh Nilai Parameter C. Nilai C yang Digunakan pada Gambar Sebelah Kiri Lebih Besar dari Gambar Sebelah Kanan	110
Gambar 4. 30 Pengaruh Nilai Parameter γ . Nilai γ pada Gambar Sebelah Kanan Memiliki Nilai Lebih Besar dari Gambar Sebelah Kiri	111

DAFTAR LAMPIRAN

Lampiran 1 Hasil Scrapping Data.....	126
Lampiran 2 Dataset yang Telah Diberi Label	130
Lampiran 3 Daftar Stopwords.....	132
Lampiran 4 Hasil Pre-Processing.....	137
Lampiran 5 Kuesioner Wawancara Pelabelan Dataset	143
Lampiran 6 Dokumentasi.....	183
Lampiran 7 Hasil TF-IDF	185

