

**SKRIPSI**

**KOMPARASI METODE LABEL POWERSET K-NN  
DAN ML-KNN DALAM KLASIFIKASI *MULTI-  
LABEL CYBERBULLYING* PADA KOMENTAR  
INSTAGRAM**

**IMAMAH NUR FADLILAH**

NPM. 21082010067

**DOSEN PEMBIMBING**

Eka Dyar Wahyuni, S.Kom., M.Kom.

Reisa Permatasari, S.T., M.Kom.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI  
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR  
FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI SISTEM INFORMASI  
SURABAYA  
2025**

*Halaman ini sengaja dikosongkan*


## LEMBAR PENGESAHAN SKRIPSI

### KOMPARASI METODE LABEL POWERSET K-NN DAN ML-KNN DALAM KLASIFIKASI MULTI-LABEL CYBERBULLYING PADA KOMENTAR INSTAGRAM

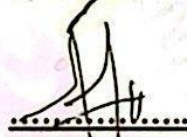
Oleh:  
IMAMAH NUR FADLILAH  
NPM. 21082010067

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Sistem Informasi Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 07 Maret 2025

Eka Dyar Wahyuni, S. Kom., M. Kom.  
NIP. 198412012021212005

  
..... (Pembimbing I)

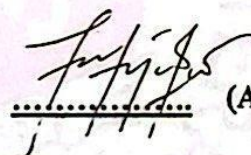
Reisa Permatasari, S.T., M.Kom.  
NIP. 199205142022032007

  
..... (Pembimbing II)

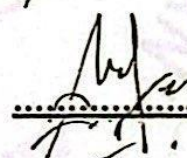
Nur Cahyo Wibowo, S. Kom., M. Kom.  
NIP. 197903172021211002

  
..... (Ketua Penguji)


Seftin Fitri Ana Wati, S. Kom., M. Kom.  
NPT. 21219910320267

  
..... (Anggota Penguji II)

Anindo Saka Fitri, S. Kom., M. Kom.  
NIP. 199303252024062001

  
..... (Anggota Penguji III)

Mengetahui,  
Dekan Fakultas Ilmu Komputer

  
Prof. Dr. Ir. Novirina Hendrasarie, MT.  
NIP. 196811261994032001

*Halaman ini sengaja dikosongkan*


**LEMBAR PERSETUJUAN SKRIPSI**

**KOMPARASI METODE LABEL POWERSET K-NN DAN ML-KNN  
DALAM KLASIFIKASI MULTI-LABEL CYBERBULLYING PADA  
KOMENTAR INSTAGRAM**

Oleh:  
IMAMAH NUR FADLILAH  
NPM. 21082010107

Menyetujui,

Koordinator Program Studi Sistem Informasi  
Fakultas Ilmu Komputer

  
Agung Brastama Putra, S.Kom., M.Kom.  
NIP. 198511242021211003

*Halaman ini sengaja dikosongkan*

## SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan dibawah ini :

Nama : Imamah Nur Fadlilah  
NPM : Sistem Informasi  
Program : Sarjana (S1)  
Program Studi : Sistem Informasi  
Fakultas : Ilmu Komputer


Menyatakan bahwa dalam dokumen Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila dikemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.



Surabaya, 12 Maret 2025  
Yang Membuat Pernyataan,

  
MEPERA  
TEMPE  
14AMX196171044  
Imamah Nur Fadlilah  
NPM 21082010067

*Halaman ini sengaja dikosongkan*



## ABSTRAK

Nama Mahasiswa / NPM : Imamah Nur Fadlilah / 21082010067  
Judul Skripsi : Komparasi Metode Label Powerset K-NN dan ML-KNN dalam Klasifikasi Multi-label Cyberbullyng pada Komentar Instagram  
Dosen Pembimbing : 1. Eka Dyar Wahyuni, S.Kom., M.Kom.  
2. Reisa Permatasari, S.T., M.Kom.

Cyberbullying merupakan bentuk perundungan yang dilakukan melalui media digital, termasuk platform media sosial seperti Instagram. Dengan tingginya jumlah pengguna dan fitur komentar yang interaktif, Instagram menjadi salah satu platform dengan tingkat insiden cyberbullying tertinggi. Data dalam penelitian ini diperoleh dari komentar pada akun public figure. Penelitian ini bertujuan untuk menerapkan klasifikasi multi-label dalam mendeteksi cyberbullying pada komentar Instagram. Dua pendekatan utama dibandingkan, yaitu Problem Transformation dengan Label Powerset KNN dan Algorithm Adaptation dengan ML-KNN. Data penelitian dikonversi menggunakan teknik ekstraksi fitur TF-IDF dengan kombinasi n-grams (1-3) untuk meningkatkan akurasi model. Hasil penelitian menunjukkan bahwa ML-KNN memiliki performa lebih baik dibandingkan Label Powerset KNN. ML-KNN menunjukkan F1-score lebih tinggi sebesar 0.849 dibandingkan LP-KNN sebesar 0.828, serta hamming loss lebih rendah sebesar 0.110 dibandingkan 0.124 pada LP-KNN. Dengan demikian, ML-KNN lebih akurat dalam menangani data multi-label. Selain itu, sistem yang dikembangkan dapat mengklasifikasikan komentar ke dalam berbagai kategori cyberbullying secara simultan, mendukung input dalam bentuk teks tunggal maupun file CSV, sehingga memungkinkan analisis dalam skala besar. Dengan temuan ini, penelitian diharapkan dapat berkontribusi dalam pengembangan sistem deteksi cyberbullying untuk meningkatkan keamanan digital di media sosial.

**Kata kunci :** Cyberbullying, Instagram, Klasifikasi Multi-label, Label Powerset KNN, ML-KNN, TF-IDF.

*Halaman ini sengaja dikosongkan*

## ABSTRACT

Student Name / NPM : Imamah Nur Fadlilah / 21082010067  
Thesis Title : Komparasi Metode Label Powerset K-NN dan ML-KNN dalam Klasifikasi Multi-label Cyberbullyng pada Komentar Instagram  
Advisors : 1. Eka Dyar Wahyuni, S.Kom., M.Kom.  
2. Reisa Permatasari, S.T., M.Kom.

## ABSTRACT

Cyberbullying is a form of bullying conducted through digital media, including social media platforms such as Instagram. With its high number of users and interactive comment features, Instagram is one of the platforms with the highest incidence of cyberbullying. The data in this study was obtained from comments on public figure accounts. This research aims to apply multi-label classification in detecting cyberbullying on Instagram comments. Two main approaches were compared, namely Problem Transformation with Label Powerset KNN and Algorithm Adaptation with ML-KNN. The research data was converted using TF-IDF feature extraction technique with a combination of n-grams (1-3) to improve the model accuracy. The results showed that ML-KNN performed better than Label Powerset KNN. ML-KNN shows a higher F1-score of 0.849 compared to LP-KNN of 0.828, and a lower hamming loss of 0.110 compared to 0.124 in LP-KNN. Thus, ML-KNN is more accurate in handling multi-label data. In addition, the developed system can classify comments into various cyberbullying categories simultaneously, supports input in the form of single text as well as CSV files, thus enabling large-scale analysis. With these findings, the research is expected to contribute to the development of cyberbullying detection systems to enhance digital safety on social media.

**Keywords:** Cyberbullying, Instagram, Multi-label Classification, Label Powerset KNN, ML-KNN, TF-IDF

*Halaman ini sengaja dikosongkan*

## KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini yang berjudul "Komparasi Metode Label Powerset KNN dan ML-KNN dalam Klasifikasi Multi-label Cyberbullying pada Komentar Instagram" dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi.

Dalam penyusunan skripsi ini, penulis mendapatkan banyak dukungan, bimbingan, serta motivasi dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua tercinta, yang doa-doanya selalu menjadi penguat di setiap langkah. Terima kasih atas segala kasih sayang, dukungan tanpa henti, dan kesabaran yang tiada batasnya.
2. Ibu Eka Dyar Wahyuni, S.Kom., M.Kom., dan Ibu Reisa Permatasari, S.T., M.Kom., selaku dosen pembimbing I dan II, yang dengan sabar membimbing, memberikan arahan dan masukan. Terima kasih atas kesediaannya meluangkan waktu, tenaga, dan perhatian di setiap tahap penyelesaian skripsi ini.
3. Seluruh bapak dan ibu dosen Program Studi Sistem Informasi, yang telah menjadi bagian dari perjalanan akademik ini. Ilmu dan pengalaman yang telah diberikan akan selalu menjadi bekal berharga di masa depan.
4. Muhammad Adam Rachman, yang tidak hanya membantu, tetapi juga selalu ada di setiap proses, memberikan motivasi, menemani di saat sulit, dan mengingatkan untuk tetap semangat. Terima kasih sudah selalu ada dan sabar menemani proses panjang ini.
5. Dianita, Aidah, Dinda, Eka, dan Pipit, yang dengan tulus membantu. Tanpa kalian, mungkin proses ini akan terasa jauh lebih berat. Terima kasih atas waktu dan tenaga yang telah diberikan.
6. Rekan-rekan seperjuangan dari grup Mencuri Abang Saleh dan Penghuni Bayangan Kos Parlin, yang selalu berbagi suka dan duka dalam proses panjang ini. Terima kasih atas obrolan random, atas keluhan bersama yang akhirnya berubah menjadi semangat baru, dan atas semua dukungan yang membuat perjalanan ini lebih bermakna.

7. Diri sendiri, yang telah berusaha dan bertahan sampai sejauh ini. Terima kasih telah berjuang, melewati proses panjang dengan segala tantangan dan rintangan, serta tidak menyerah meskipun banyak hambatan yang harus dilalui.
8. Seluruh pihak terkait yang tidak bisa disebutkan satu per satu, tetapi telah memberikan bantuan, dorongan, dan semangat dalam berbagai bentuk. Terima kasih telah menjadi bagian dari perjalanan ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan untuk perbaikan di masa mendatang. Semoga skripsi ini dapat memberikan manfaat bagi pembaca serta menjadi referensi bagi penelitian selanjutnya.

Surabaya, 05 Maret 2025

Penulis

## DAFTAR ISI

<b>LEMBAR PENGESAHAN SKRIPSI .....</b>	<b>iii</b>
<b>LEMBAR PERSUTUJUAN SKRIPSI.....</b>	<b>v</b>
<b>ABSTRAK .....</b>	<b>ix</b>
<b>KATA PENGANTAR.....</b>	<b>xiii</b>
<b>DAFTAR ISI.....</b>	<b>xv</b>
<b>DAFTAR GAMBAR.....</b>	<b>xix</b>
<b>DAFTAR TABEL .....</b>	<b>xxiii</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	5
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian .....	5
1.5 Manfaat Penelitian .....	6
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>7</b>
2.1 Penelitian Terdahulu .....	7
2.2 Landasan Teori.....	8
2.2.1 Instagram.....	9
2.2.2 Text Mining.....	9
2.2.3 Klasifikasi <i>Multi-label</i> .....	10
2.2.4 <i>Cyberbullying</i> .....	10
2.2.5 <i>Preprocessing Text</i> .....	11
2.2.6 <i>Holdout</i> .....	12
2.2.7 TF-IDF .....	12
2.2.8 <i>K-Nearest Neighbors (KNN)</i> .....	12

2.2.9	<i>Multi-label K-Nearest Neighbors (ML-KNN)</i> .....	13
2.2.10	<i>Label Powerset</i> .....	13
2.2.11	<i>Confusion Matrix</i> .....	14
2.2.12	Sistem Prediksi Klasifikasi .....	15
2.2.13	Support Vector Machine .....	15
2.2.14	Hamming Loss .....	16
<b>BAB III METODOLOGI PENELITIAN</b> .....		<b>17</b>
3.1	Studi Literatur .....	17
3.2	Analisis Kebutuhan .....	18
3.2.1	Kebutuhan Data .....	18
3.2.2	Kebutuhan <i>Software</i> dan <i>Hardware</i> .....	18
3.3	Pengumpulan Data .....	18
3.4	Filtering Data .....	19
3.5	Pelabelan Data .....	19
3.6	Pembagian Data .....	19
3.7	Eksplorasi Data .....	20
3.8	Augmentasi Data .....	20
3.9	Praproses Data .....	20
3.9.1	<i>Cleaning</i> .....	21
3.9.2	<i>Case Folding</i> .....	21
3.9.3	<i>Normalization</i> .....	22
3.9.4	<i>Stopwords Removal</i> .....	22
3.9.5	<i>Stemming</i> .....	22
3.9.6	<i>Tokenization</i> .....	22
3.10	Perancangan Model .....	23
3.10.1	Pembagian Dataset untuk Training dan Testing .....	24



3.10.2	Pembobotan Kata dengan TF-IDF .....	24
3.10.3	Klasifikasi Cyberbullying Skenario A .....	24
3.10.4	Klasifikasi Cyberbullying Skenario B .....	25
3.10.5	Evaluasi Model.....	25
3.10.6	Penyimpanan Model.....	26
3.11	Penerapan Model.....	26
3.12	Validasi Model .....	26
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>		<b>27</b>
4.1	Hasil .....	27
4.1.1	Studi Literatur .....	27
4.1.2	Implementasi Kebutuhan .....	27
4.1.3	Pengumpulan Data .....	28
4.1.4	Filtering Data .....	30
4.1.5	Pelabelan Data.....	32
4.1.6	Pembagian Data .....	36
4.1.7	Eksplorasi Data .....	36
4.1.8	Augmentasi Data.....	38
4.1.9	Praproses Data.....	40
4.1.10	Perancangan Model.....	44
4.1.11	Penerapan Model.....	79
4.1.12	Validasi Model .....	80
4.2	Pembahasan.....	82
<b>BAB V PENUTUP.....</b>		<b>85</b>
5.1	Kesimpulan .....	85
5.2	Saran.....	85
<b>DAFTAR PUSTAKA .....</b>		<b>87</b>

*Halaman ini sengaja dikosongkan*

## DAFTAR GAMBAR

Gambar 1. 1 Peringkat Cyberbullying di Setiap Platform .....	1
Gambar 2. 1 Ilustrasi Label Powerset .....	14
Gambar 3.1 Diagram Alir Penelitian .....	17
Gambar 3. 2 Diagram Praproses Data.....	21
Gambar 3. 3 Perancangan Model .....	23
Gambar 4. 1 Potongan Kode Pengumpulan Data.....	28
Gambar 4. 2 Contoh Hasil Scraping Data.....	29
Gambar 4. 3 Kode Menggabungkan Data.....	30
Gambar 4. 4 Kode Menghapus Data Tahap Filtering Data.....	30
Gambar 4. 5 Kode Menghapus Spam Tahap Filtering Data .....	31
Gambar 4. 6 Widget Pelabelan Data .....	32
Gambar 4. 7 Kode Perhitungan Fleiss' Kappa.....	34
Gambar 4. 8 Hasil Perhitungan Fleiss' Kappa .....	34
Gambar 4. 9 Hasil Distribusi Label.....	35
Gambar 4. 10 Kode untuk Pembagian Data Modelling dan Validation .....	36
Gambar 4. 11 Grafik Distribusi Panjang Komentar.....	37
Gambar 4. 12 Grafik Top 20 Kata Sering Muncul.....	37
Gambar 4. 13 WordCloud per-Label .....	38
Gambar 4. 14 Kode Back Translation.....	39
Gambar 4. 15 Kode Penggabungan Data Augmentasi.....	40
Gambar 4. 16 Distribusi Label Setelah Augmentasi.....	40
Gambar 4. 17 Kode Proses Cleaning .....	41
Gambar 4. 18 Kode Proses Case Folding .....	41
Gambar 4. 19 Kode Proses Normalisasi .....	42
Gambar 4. 20 Kode Proses Stopwords Removal .....	42
Gambar 4. 21 Kode Proses Stemming .....	43
Gambar 4. 22 Kode Proses Tokenisasi .....	43
Gambar 4. 23 Wordcloud Setelah Praproses.....	44
Gambar 4. 24 Kode Pembagian Data Latih dan Data Uji.....	46
Gambar 4. 25 Kode Pembobotan Kata dengan TF-IDF.....	47

Gambar 4. 26 <i>Import Pustaka Model</i> .....	48
Gambar 4. 27 Kode <i>Fitting Model 1 80:20 Skenario A</i> .....	49
Gambar 4. 28 <i>Classification Report Model 1 80:20 Skenario A</i> .....	49
Gambar 4. 29 Kode <i>Fitting Model 1 70:30 Skenario A</i> .....	50
Gambar 4. 30 <i>Classification Report Model 1 70:30 Skenario A</i> .....	50
Gambar 4. 31 Kode <i>Fitting Model 2 80:20 Skenario A</i> .....	50
Gambar 4. 32 <i>Classification Report Model 2 80:20 Skenario A</i> .....	51
Gambar 4. 33 Kode <i>Fitting Model 2 70:30 Skenario A</i> .....	51
Gambar 4. 34 <i>Classification Report Model 2 70:30 Skenario A</i> .....	52
Gambar 4. 35 Kode <i>Fitting Model 3 80:20 Skenario A</i> .....	52
Gambar 4. 36 <i>Classification Report Model 3 80:20 Skenario A</i> .....	53
Gambar 4. 37 Kode <i>Fitting Model 3 70:30 Skenario A</i> .....	53
Gambar 4. 38 <i>Classification Report Model 3 70:30 Skenario A</i> .....	54
Gambar 4. 39 Kode <i>Fitting Model 4 80:20 Skenario A</i> .....	54
Gambar 4. 40 <i>Classification Report Model 4 80:20 Skenario A</i> .....	55
Gambar 4. 41 Kode <i>Fitting Model 4 70:30 Skenario A</i> .....	55
Gambar 4. 42 <i>Classification Report Model 4 70:30 Skenario A</i> .....	56
Gambar 4. 43 Kode <i>Fitting Model 5 80:20 Skenario A</i> .....	56
Gambar 4. 44 <i>Classification Report Model 5 80:20 Skenario A</i> .....	57
Gambar 4. 45 Kode <i>Fitting Model 5 70:30 Skenario A</i> .....	57
Gambar 4. 46 <i>Classification Report Model 5 70:30 Skenario A</i> .....	58
Gambar 4.47 Kode <i>Fitting Model 6 80:20 Skenario A</i> .....	58
Gambar 4.48 <i>Classification Report Model 6 80:20 Skenario A</i> .....	59
Gambar 4. 49 Kode <i>Fitting Model 6 70:30 Skenario A</i> .....	59
Gambar 4. 50 <i>Classification Report Model 6 70:30 Skenario A</i> .....	60
Gambar 4. 51 Inisialisasi Model SVM .....	60
Gambar 4. 52 <i>Classification Report Model SVM</i> .....	61
Gambar 4. 53 Kode <i>Fitting Model 1 80:20 Skenario B</i> .....	61
Gambar 4. 54 <i>Classification Report Model 1 80:20 Skenario B</i> .....	62
Gambar 4. 55 Kode <i>Fitting Model 1 70:30 Skenario B</i> .....	62
Gambar 4. 56 <i>Classification Report Model 1 70:30 Skenario B</i> .....	63
Gambar 4. 57 Kode <i>Fitting Model 2 80:20 Skenario B</i> .....	63

Gambar 4. 58 <i>Classification Report</i> Model 2 80:20 Skenario B .....	64
Gambar 4. 59 Kode <i>Fitting</i> Model 2 70:30 Skenario B .....	64
Gambar 4. 60 <i>Classification Report</i> Model 2 70:30 Skenario B .....	65
Gambar 4. 61 Kode <i>Fitting</i> Model 3 80:20 Skenario B .....	65
Gambar 4. 62 <i>Classification Report</i> Model 3 80:20 Skenario B .....	66
Gambar 4. 63 Kode <i>Fitting</i> Model 3 70:30 Skenario B .....	66
Gambar 4. 64 <i>Classification Report</i> Model 3 70:30 Skenario B .....	67
Gambar 4. 65 Kode <i>Fitting</i> Model 4 80:20 Skenario B .....	67
Gambar 4. 66 <i>Classification Report</i> Model 4 80:20 Skenario B .....	68
Gambar 4. 67 Kode <i>Fitting</i> Model 4 70:30 Skenario B .....	68
Gambar 4. 68 <i>Classification Report</i> Model 4 70:30 Skenario B .....	69
Gambar 4. 69 Kode <i>Fitting</i> Model 5 80:20 Skenario B .....	69
Gambar 4. 70 <i>Classification Report</i> Model 5 80:20 Skenario B .....	70
Gambar 4. 71 Kode <i>Fitting</i> Model 5 70:30 Skenario B .....	70
Gambar 4. 72 <i>Classification Report</i> Model 5 70:30 Skenario B .....	71
Gambar 4. 73 Kode <i>Fitting</i> Model 6 80:20 Skenario B .....	71
Gambar 4. 74 <i>Classification Report</i> Model 6 80:20 Skenario B .....	72
Gambar 4. 75 Kode <i>Fitting</i> Model 6 70:30 Skenario B .....	72
Gambar 4. 76 <i>Classification Report</i> Model 6 70:30 Skenario B .....	73
Gambar 4. 77 Kode Penyimpanan Model Terbaik .....	78
Gambar 4. 78 Potongan Kode Integrasi Flask .....	79
Gambar 4. 79 Fitur Prediksi Cyberbullying Berbasis Teks .....	79
Gambar 4. 80 Fitur Prediksi Cyberbullying Berbasis <i>File Input</i> .....	80
Gambar 4. 81 Contoh Hasil Prediksi Berbasis Teks .....	81
Gambar 4. 82 Tampilan Awal Setelah Mengunggah Data CSV .....	81
Gambar 4. 83 Tampilan Setelah Melakukan Prediksi File CSV .....	81
Gambar 4. 84 Tampilan Visualisasi Bar Chart .....	82
Gambar 4. 85 Visualiasi Wordcloud Setiap Label .....	82
Gambar 4. 86 Confusion Matrix Setiap Label .....	83
Gambar 4. 87 Confusion Matrix Kombinasi Label .....	84

*Halaman ini sengaja dikosongkan*

## DAFTAR TABEL

Tabel 4. 1 Contoh Hasil Pelabelan Data .....	32
Tabel 4. 2 Hasil Label Dominan .....	35
Tabel 4. 3 Contoh Hasil Praproses Data .....	43
Tabel 4. 4 Contoh Hasil Konversi Label.....	45
Tabel 4. 5 Proporsi Pembagian Data Latih dan Data Uji.....	46
Tabel 4. 6 Jenis Skenario Pemodelan.....	48
Tabel 4.7 Hasil Evaluasi Model Skenario A.....	73
Tabel 4. 8 Hasil Evaluasi Model Skenario B .....	75
Tabel 4. 9 Hasil Lima Model Terbaik.....	77

*Halaman ini sengaja dikosongkan*