

**ANALISIS SENTIMEN KOMENTAR *YOUTUBE*
TERHADAP KEBIJAKAN POLITIK DAYA ANAGATA
NUSANTARA MENGGUNAKAN *LONG SHORT-TERM*
MEMORY DAN *SKIP-GRAM***

TUGAS AKHIR

Oleh :
Aldira Ryan Pratama
202110715223



**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BHAYANGKARA JAKARTA RAYA
2025**

LEMBAR PERSETUJUAN PEMBIMBING

Judul Tugas Akhir : Analisis Sentimen Komentar YouTube terhadap Kebijakan Politik Daya Anagata Nusantara menggunakan Long Short-Term Memory dan Skip-gram

Nama Mahasiswa : Aldira Ryan Pratama

Nomor Pokok Mahasiswa : 202110715223

Program Studi/Fakultas : Informatika/Ilmu Komputer

Jakarta, 11 Juli 2025

MENYETUJUI,

Pembimbing I



Wowon Priatna, S.T., M.TI

NIDN : 0429118007

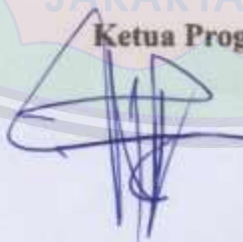
Pembimbing II



Siti Setyawati, S.Pd., M.Pd.

NIDN : 0313107904

Ketua Program Studi



Ahmad Fathurozi, S.E., M.M.S.I

NIP. 2012486

Program Studi Informatika

Fakultas Ilmu Komputer

Universitas Bhayangkara Jakarta Raya

2025

LEMBAR PENGESAHAN

Judul Tugas Akhir : Analisis Sentimen Komentar YouTube terhadap Kebijakan Politik Daya Anagata Nusantara menggunakan Long Short-Term Memory dan Skip-gram

Nama Mahasiswa : Aldira Ryan Pratama

Nomor Pokok Mahasiswa : 202110715223

Program Studi/Fakultas : Informatika/Ilmu Komputer

Tanggal Lulus Ujian : 24 Juli 2025

Sidang Akhir

Jakarta, 31 Juli 2025

MENGESAHKAN,

Penguji 1 : Prima Dina Atika, S.Kom., M.Kom

NIDN : 0311037107

Penguji 2 : Ir. Muhammad Khaerudin, M.Kom.

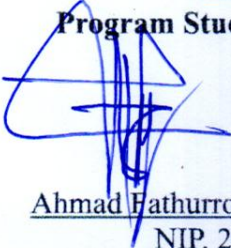
NIDN : 0413066604

Penguji 3 : Wowon Priatna, S.T., M.Ti

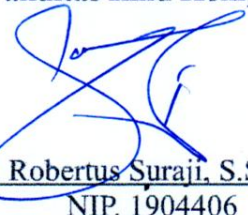
NIDN : 0313107904

MENGETAHUI,

Ketua
Program Studi Informatika


Ahmad Fathurrozi, S.E., M.M.S.I
NIP. 2012486

Dekan
Fakultas Ilmu Komputer


Dr. Robertus Suraji, S.S., M.A
NIP. 1904406



UNIVERSITAS BHAYANGKARA JAKARTA RAYA
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA

LEMBAR PERNYATAAN BUKAN PLAGIASI

Yang bertanda tangan dibawah ini :

Nama : Aldira Ryan Pratama
NPM : 202110715223
Program Studi : Informatika
Fakultas : Ilmu Komputer
Judul Tugas Akhir : Analisis Sentimen Komentar *YouTube* terhadap Kebijakan Politik Daya Anagata Nusantara menggunakan Long Short-Term Memory dan *Skip-gram*

Dengan ini menyatakan bahwa hasil penulisan skripsi yang telah saya buat ini merupakan **hasil karya saya sendiri dan benar keasliannya**. Apabila dikemudian hari penulisan skripsi ini merupakan plagiat atau penjiplakan terhadap karya orang lain, maka saya bersedia mempertanggungjawabkan sekaligus bersedia menerima sanksi berdasarkan tata tertib di Universitas Bhayangkara Jakarta Raya.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan dari pihak manapun.

Jakarta, 31 Juli 2025

Penulis



Aldira Ryan Pratama

ABSTRAK

Aldira Ryan Pratama. 202110715223. Analisis Sentimen Komentar *YouTube* terhadap Kebijakan Politik Daya Anagata Nusantara menggunakan *Long Short-Term Memory* dan *Skip-gram*. Bekasi: Fakultas Ilmu Komputer. Universitas Bhayangkara Jakarta Raya. 2025.

Media sosial, khususnya *YouTube*, telah menjadi ruang diskusi publik terhadap berbagai isu kebijakan, termasuk pembentukan Daya Anagata Nusantara (Danantara). Di tengah meningkatnya opini publik yang beragam, belum tersedia sistem otomatis yang mampu menganalisis kecenderungan sentimen masyarakat secara efisien. Model LSTM *standalone* sering kali menghasilkan akurasi yang fluktuatif pada data teks berbahasa Indonesia, sementara kombinasi LSTM dan *Skip-gram* masih jarang diterapkan dalam konteks kebijakan politik domestik. Penelitian ini bertujuan untuk mengidentifikasi pola sentimen publik terhadap kebijakan Danantara melalui analisis komentar *YouTube*, serta mengevaluasi performa model kombinasi LSTM dan *Skip-gram* dibandingkan dengan LSTM *standalone*. Metode yang digunakan mengikuti pendekatan SEMMA (*Sample, Explore, Modify, Model, Assess*). Sebanyak 1.000 komentar digunakan untuk analisis sentimen, sementara 2.480 data digunakan sebagai korpus pelatihan *embedding Skip-gram*. Pelabelan dilakukan dengan pendekatan leksikon (InSet) dan klasifikasi bersifat biner (positif-negatif). Evaluasi performa dilakukan menggunakan *confusion matrix* dan metrik klasifikasi (*accuracy, precision, recall, F1-score*) pada tiga skema data: 70:15:15, 80:10:10, dan 90:5:5. Hasil menunjukkan bahwa sentimen publik terhadap Danantara cenderung negatif, dengan banyak komentar yang menyoroti kekhawatiran atas potensi penyalahgunaan kebijakan. Model LSTM *standalone* menghasilkan performa tertinggi pada skema 90:5:5 dengan akurasi 92% dan *F1-score* 91,66%. Sebaliknya, kombinasi LSTM dan *Skip-gram* menunjukkan validasi yang lebih stabil (84%-86%) namun mencatat *F1-score* terendah 78,91% pada data uji akibat rendahnya *recall*. Dengan demikian, meskipun model kombinasi lebih stabil selama pelatihan dan validasi, LSTM *standalone* lebih unggul dalam menggeneralisasi terhadap data nyata.

Kata Kunci: Analisis Sentimen, *YouTube*, Danantara, *Long Short-Term Memory*, *Skip-gram*.

ABSTRACT

Aldira Ryan Pratama. 202110715223. *Sentiment Analysis of YouTube Comments on the Political Policy of Daya Anagata Nusantara Using Long Short-Term Memory and Skip-gram.* Bekasi: Faculty of Computer Science. Bhayangkara University Jakarta Raya. 2025.

Social media, particularly YouTube, has become a public forum for discussing various policy issues, including the formation of Daya Anagata Nusantara (Danantara). Amid the growing diversity of public opinion, there is still no automated system capable of efficiently analyzing sentiment trends in society. Standalone LSTM models often produce fluctuating accuracy when applied to Indonesian-language text data, while the combination of LSTM and Skip-gram remains rarely explored within the context of domestic political policies. This study aims to identify public sentiment patterns toward the Danantara policy through YouTube comment analysis and to evaluate the performance of a combined LSTM and Skip-gram model compared to a standalone LSTM model. The methodology follows the SEMMA approach (Sample, Explore, Modify, Model, Assess). A total of 1,000 comments were used for sentiment analysis, while 2,480 data points served as the training corpus for Skip-gram embedding. Sentiment labeling was conducted using a lexicon-based method (InSet), and classification was binary (positive-negative). Model performance was evaluated using a confusion matrix and classification metrics (accuracy, precision, recall, F1-score) across three data splits: 70:15:15, 80:10:10, and 90:5:5. Results indicate that public sentiment toward Danantara tends to be negative, with many comments expressing concerns about potential policy misuse. The standalone LSTM model achieved the highest performance under the 90:5:5 split, with an accuracy of 92% and an F1-score of 91.66%. In contrast, the LSTM and Skip-gram combination demonstrated more stable validation scores (ranging from 84% to 86%) but recorded the lowest F1-score of 78.91% on test data due to low recall. Therefore, while the combined model is more stable during training and validation, the standalone LSTM model is more effective at generalizing to real-world data.

Keywords: *Sentiment Analysis, YouTube, Danantara, Long Short-Term Memory, Skip-gram*

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIK

Sebagai sivitas akademik Universitas Bhayangkara Jakarta Raya, saya yang bertanda tangan di bawah ini :

Nama : Aldira Ryan Pratama
NPM : 202110715223
Program Studi : Informatika
Fakultas : Ilmu Komputer
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bhayangkara Jakarta Raya **Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalty-Free Right*)**, atas karya ilmiah saya yang berjudul :

Analisis Sentimen Komentar *YouTube* terhadap Kebijakan Politik Daya Anagata Nusantara menggunakan Long Short-Term Memory dan *Skip-gram*

beserta perangkat yang ada (bila diperlukan). Dengan hak bebas royalti non-eksklusif ini, Universitas Bhayangkara Jakarta Raya berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya dan mempublikasikannya di Internet atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik hak cipta.

Segala bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini menjadi tanggung jawab saya pribadi

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada tanggal : 31 Juli 2025
Yang Menyatakan



Aldira Ryan Pratama

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah Swt. atas limpahan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan tugas akhir skripsi yang berjudul “**Analisis Sentimen Komentar YouTube terhadap Kebijakan Politik Daya Anagata Nusantara menggunakan Long Short-Term Memory dan Skip-gram**”. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Fakultas Ilmu Komputer, Universitas Bhayangkara Jakarta Raya.

Dalam proses penyusunan skripsi ini, penulis menerima banyak dukungan dan bantuan dari berbagai pihak. Ucapan terima kasih secara khusus penulis sampaikan kepada kedua orang tua, Yanto Haryanto dan Romlah, atas doa, kasih sayang, dan dukungan yang tidak pernah terputus selama penulis menempuh pendidikan. Dengan penuh rasa hormat dan terima kasih, penulis juga menyampaikan apresiasi kepada:

1. Bapak Irjen Polisi (Purn) Prof. Dr. Drs. Bambang Karsono, S.H., M.M., Ph.D., D.Crim (HC) selaku Rektor Universitas Bhayangkara Jakarta Raya;
2. Bapak Dr. Robertus Suraji, M.A. selaku Dekan Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya;
3. Bapak Ahmad Fathurrozi, S.E., M.M.S.I. selaku Ketua Program Studi S1 Informatika Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya;
4. Bapak Wowon Priatna, S.T, M.TI. selaku Dosen Pembimbing Akademik dan Dosen Pembimbing I dalam penelitian skripsi Program Studi

Informatika Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya;

5. Ibu Siti Setyawati, S.Pd, M.Pd. selaku Dosen Pembimbing II dalam penelitian skripsi Program Studi Informatika Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis membuka diri terhadap kritik dan saran yang membangun sebagai bahan evaluasi dan perbaikan ke depan. Akhir kata, besar harapan penulis agar karya ini dapat memberikan manfaat, baik bagi penulis sendiri maupun bagi pembaca yang memiliki minat dalam bidanin g analisis sentimen dan pengolahan bahasa alami.

Jakarta, 31 Juli 2025

Hormat Saya



Aldira Ryan Pratama

DAFTAR ISI

LEMBAR PERSETUJUAN PEMBIMBING	ii
LEMBAR PENGESAHAN	iii
LEMBAR PERNYATAAN BUKAN PLAGIASI	iv
ABSTRAK	v
ABSTRACT	vi
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL	xiv
DAFTAR GAMBAR.....	xvi
DAFTAR LAMPIRAN	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah	6
1.3 Rumusan Masalah	7
1.4 Tujuan Penelitian.....	7
1.5 Manfaat Penelitian.....	8
1.6 Batasan Masalah.....	8
1.7 Sistematika Tugas Akhir.....	9
BAB II TINJAUAN PUSTAKA	11
2.1 State of The Art	11
2.2 Daya Anagata Nusantara (Danantara)	12
2.3 YouTube	13

2.4	Analisis Sentimen.....	14
2.5	Machine Learning.....	15
2.6	Deep Learning	16
2.7	Natural Language Processing (NLP).....	17
2.8	Word Embedding.....	18
2.8.1	Word2Vec.....	18
2.8.2	Skip-gram.....	19
2.9	Recurrent Neural Network (RNN)	20
2.9.1	Long Short-Term Memory (LSTM).....	21
2.10	SEMMA	27
2.11	Text Preprocessing.....	28
2.12	Confusion Matrix	30
2.13	Evaluation Metrics.....	31
BAB III METODOLOGI PENELITIAN		34
3.1	Objek Penelitian	34
3.2	Metode Pengumpulan Data	34
3.3	Kerangka Penelitian.....	36
3.4	SEMMA	39
3.4.1	Sample.....	40
3.4.2	Explore	42
3.4.3	Modify.....	46
3.4.4	Model	51
3.4.5	Assess.....	57
3.5	Perangkat Penelitian	62
BAB IV HASIL DAN PEMBAHASAN.....		64

4.1	Persiapan	64
4.2	Sample	65
4.2.1	Scraping Data	66
4.2.2	Labeling Data	71
4.2.3	Sampling Data	74
4.3	Explore	75
4.3.1	Preliminary Text Cleaning	76
4.3.2	Eksplorasi Distribusi Label Sentimen	81
4.3.3	Eksplorasi Frekuensi Kata	83
4.3.4	Eksplorasi Panjang Komentar	85
4.3.5	Analisis Outlier Data	88
4.4	Modify	89
4.4.1	Text Preprocessing	90
4.4.2	Train-valid-test Splitting	98
4.4.3	Text Vectorization	100
4.4.4	Padding	103
4.5	Model	104
4.5.1	Kombinasi LSTM dan Skip-gram	104
4.5.2	LSTM Standalone	122
4.6	Assess	128
4.6.1	Confusion Matrix Kombinasi LSTM dan Skip-gram	131
4.6.2	Confusion Matrix LSTM Standalone	134
4.6.3	Evaluation Metrics Kombinasi LSTM dan Skip-gram	137
4.6.4	Evaluation Metrics LSTM Standalone	140
4.7	Pembahasan	144

BAB V PENUTUP	151
5.1 Kesimpulan.....	151
5.2 Saran	152
DAFTAR PUSTAKA.....	154
LAMPIRAN.....	161



DAFTAR TABEL

Tabel 2.1 State of The Art.....	11
Tabel 3.1 Daftar Konten YouTube pada Data Primer	35
Tabel 3.2 Perangkat Penelitian	62
Tabel 4.1 Daftar Objek Penelitian	66
Tabel 4.2 Sebagian Skor Polaritas pada InSet Lexicon	72
Tabel 4.3 Perbandingan Sebelum dan Sesudah Preliminary Text Cleaning	80
Tabel 4.4 Perbandingan Sebelum dan Sesudah Proses Cleansing	91
Tabel 4.5 Perbandingan Sebelum dan Sesudah Proses Tokenization	92
Tabel 4.6 Perbandingan Sebelum dan Sesudah Proses Standardization.....	94
Tabel 4.7 Perbandingan Sebelum dan Sesudah Proses Stemming	95
Tabel 4.8 Perbandingan Sebelum dan Sesudah Proses Removing Stopwords	97
Tabel 4.9 Proporsi Sentimen tiap Skema Pembagian Data	100
Tabel 4.10 Perbandingan Sebelum dan Sesudah Proses Text Vectorization.....	102
Tabel 4.11 Perbandingan Sebelum dan Sesudah Proses Padding.....	104
Tabel 4.12 Contoh Representasi Vektor Kata Hasil dari Pelatihan Skip-gram ..	106
Tabel 4.13 Data untuk Contoh Perhitungan Manual	111
Tabel 4.14 Rata-rata Akurasi dari Hasil Pelatihan dan Hyperparameter Tuning Kombinasi LSTM dan Skip-gram.....	120
Tabel 4.15 Daftar Model Terbaik dari Pelatihan dan Hyperparameter Tuning Kombinasi LSTM dan Skip-gram.....	121
Tabel 4.16 Rata-rata Akurasi dari Hasil Pelatihan dan Hyperparameter Tuning LSTM Standalone	126
Tabel 4.17 Daftar Model Terbaik dari Pelatihan dan Hyperparameter Tuning LSTM Standalone	127
Tabel 4.18 Nilai Confusion Matrix Kombinasi LSTM dan Skip-gram.....	134
Tabel 4.19 Nilai Confusion Matrix LSTM Standalone	136
Tabel 4.20 Perbandingan Performa Rata-rata Akurasi Pelatihan dan Validasi Model dari 50 kali Percobaan	145

Tabel 4.21 Perbandingan Performa Model Terbaik pada Tahap Pelatihan dan Validasi.....	146
Tabel 4.22 Perbandingan Performa Model Terbaik pada Data Uji.....	148



DAFTAR GAMBAR

Gambar 1.1 Daftar 7 Negara dengan Pengguna YouTube Terbesar (Per Januari 2025)	1
Gambar 2.1 Arsitektur LSTM	22
Gambar 3.1 Kerangka Penelitian.....	38
Gambar 3.2 Kerangka Tahapan SEMMA.....	39
Gambar 3.3 Kerangka Tahapan Sample	40
Gambar 4.1 Kode Menginstall Pustaka yang Dibutuhkan	64
Gambar 4.2 Kode Mengimpor Pustaka	65
Gambar 4.3 Variable API_KEY di Google Colab	67
Gambar 4.4 Kode Proses Scraping Data	69
Gambar 4.5 Proses Penggabungan dan Penyalinan Data	70
Gambar 4.6 Tampilan Seluruh Data Mentah	71
Gambar 4.7 Persiapan Labeling Data	71
Gambar 4.8 Kode Proses Labeling Data	73
Gambar 4.9 Penyalinan Seluruh Data dengan Label.....	73
Gambar 4.10 Tampilan Seluruh Data dengan Label.....	74
Gambar 4.11 Kode Proses Sampling Data.....	75
Gambar 4.12 Tampilan Data untuk Klasifikasi	75
Gambar 4.13 Kode Proses Pemisahan Subset untuk Eksplorasi	76
Gambar 4.14 Kode Pengunduhan Daftar Kosakata Slang.....	77
Gambar 4.15 Tampilan Sebagian Isi Kamus Slang	77
Gambar 4.16 Kode Pembentukan Kamus Typo.....	78
Gambar 4.17 Kode Pembentukan Stopwords Manual.....	78
Gambar 4.18 Kode Penggabungan Stopwords dari Pustaka NLTK dan Manual Stopwords.....	79
Gambar 4.19 Kode Proses Preliminary Text Cleaning	80
Gambar 4.20 Diagram Batang Jumlah Data per Sentimen.....	82
Gambar 4.21 Tampilan Diagram Batang Distribusi Sentimen dari Setiap Kanal	83
Gambar 4.22 Tampilan Word Cloud dan Tabel Frekuensi Kata Tertinggi	84

Gambar 4.23 Tampilan Word Cloud dan Tabel Frekuensi Kata Tertinggi pada Data Eksplorasi Negatif	85
Gambar 4.24 Histogram dan Tabel Statistik Deskriptif dari Panjang Komentar pada Data Eksplorasi Positif	86
Gambar 4.25 Histogram dan Tabel Statistik Deskriptif dari Panjang Komentar pada Data Eksplorasi Negatif.....	87
Gambar 4.26 Histogram dan Tabel Statistik Deskriptif dari Panjang Komentar pada Data untuk Klasifikasi	87
Gambar 4.27 Boxplot dan Tabel Statistik Deskriptif pada Data untuk Klasifikasi	89
Gambar 4.28 Kode Proses Cleansing	90
Gambar 4.29 Kode Proses Tokenization	92
Gambar 4.30 Kode Proses Standardization	93
Gambar 4.31 Kode Proses Stemming.....	95
Gambar 4.32 Kode Proses Removing Stopwords	97
Gambar 4.33 Kode Pembentukan DataFrame Hasil Proses Text Preprocessing	98
Gambar 4.34 Kode Proses Train-Valid-Test Splitting	99
Gambar 4.35 Kode Proses Pelatihan Tokenizer	101
Gambar 4.36 Kode Proses Transformasi Data dengan Tokenizer	102
Gambar 4.37 Kode Proses Padding	103
Gambar 4.38 Kode Proses Pelatihan Skip-gram	105
Gambar 4.39 Kode Proses Pembentukan Embedding Matrix	109
Gambar 4.40 Kode Pembentukan Ilustrasi Arsitektur Kombinasi LSTM dan Skip-gram.....	110
Gambar 4.41 Kode Proses Pelatihan dan Hyperparameter Tuning Kombinasi LSTM dan Skip-gram	119
Gambar 4.42 Kode Pembentukan Ilustasi Arsitektur LSTM Standalone	123
Gambar 4.43 Kode Proses Pelatihan dan Hyperparameter Tuning LSTM Standalone	125
Gambar 4.44 Kode Proses Persiapan Sebelum Evaluasi Model	129
Gambar 4.45 Kode Proses Analisis Confusion Matrix.....	130

Gambar 4.46 Kode Proses Analisis Classification Report	131
Gambar 4.47 Confusion Matrix Kombinasi LSTM dan Skip-gram Skema 70:15:15	132
Gambar 4.48 Confusion Matrix Kombinasi LSTM dan Skip-gram Skema 80:10:10	133
Gambar 4.49 Confusion Matrix Kombinasi LSTM dan Skip-gram Skema 90:5:5	133
Gambar 4.50 Confusion Matrix LSTM Standalone Skema 70:15:15	135
Gambar 4.51 Confusion Matrix LSTM Standalone Skema 80:10:10	135
Gambar 4.52 Confusion Matrix LSTM Standalone Skema 90:5:5	136
Gambar 4.53 Classification Report Kombinasi LSTM dan Skip-gram Skema 70:15:15	138
Gambar 4.54 Classification Report Kombinasi LSTM dan Skip-gram Skema 80:10:10	139
Gambar 4.55 Classification Report Kombinasi LSTM dan Skip-gram Skema 90:5:5	140
Gambar 4.56 Classification Report LSTM Standalone Skema 70:15:15	142
Gambar 4.57 Classification Report LSTM Standalone Skema 80:10:10	143
Gambar 4.58 Classification Report LSTM Standalone Skema 90:5:5	144

DAFTAR LAMPIRAN

Lampiran 1 : Hasil Cek Plagiarisme	161
Lampiran 2 : Biodata Mahasiswa.....	170
Lampiran 3 : Kartu Konsultasi Bimbingan Pembimbing I	171
Lampiran 4 : Kartu Konsultasi Bimbingan Pembimbing II	172
Lampiran 5 : Rekomendasi Pembimbing I.....	173
Lampiran 6 : Rekomendasi Pembimbing II	174

