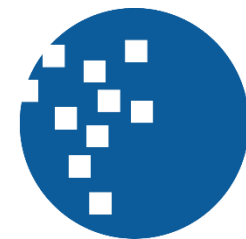


**ANALISIS SENTIMEN BERBASIS ASPEK TERHADAP
ULASAN APLIKASI DUOLINGO MENGGUNAKAN
ALGORITMA SVM DAN INDOBERT**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

Aura Lintang Pembayun Pinasti Aji

00000061031

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

**ANALISIS SENTIMEN BERBASIS ASPEK TERHADAP
ULASAN APLIKASI DUOLINGO MENGGUNAKAN
ALGORITMA SVM DAN INDOBERT**



Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Komputer (S.Kom.)

Aura Lintang Pembayun Pinasti Aji

00000061031

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Aura Lintang Pembayun Pinasti Aji

Nomor Induk Mahasiswa : 00000061031

Program Studi : Sistem Informasi

Skripsi dengan judul:

ANALISIS SENTIMEN BERBASIS ASPEK TERHADAP ULASAN APLIKASI
DUOLINGO MENGGUNAKAN ALGORITMA SVM DAN INDOBERT

Merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 26 Juni 2025



Aura Lintang Pembayun Pinasti Aji

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

Skripsi dengan judul

ANALISIS SENTIMEN BERBASIS ASPEK TERHADAP ULASAN APLIKASI
DUOLINGO MENGGUNAKAN ALGORITMA SVM DAN INDOBERT

Oleh

Nama : Aura Lintang Pembayun Pinasti Aji
NIM : 00000061031
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

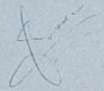
Telah diujikan pada hari Kamis, 12 Juni 2025

Pukul 13.00 s.d 15.00 dan dinyatakan

LULUS

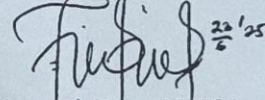
Dengan susunan penguji sebagai berikut.

Ketua Sidang



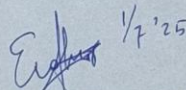
Prof. Friska Natalia, Ph.D.
0306128307

Penguji



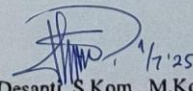
Dr. Irmawati, S.Kom., M.M.S.I.
0805097703

Pembimbing



Monika Evelin Johan, S.Kom., M.M.S.I.
0327059501/ 071381

Ketua Program Studi Sistem Informasi



Ririn Ikana Desanti, S.Kom., M.Kom.

iii

Analisis Sentimen Berbasis..., Aura Lintang Pembayun Pinasti Aji, Universitas Multimedia
Nusantara

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Aura Lintang Pembayun Pinasti Aji

NIM : 00000061031

Program Studi : Sistem Informasi

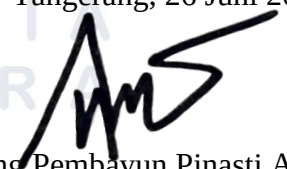
Jenjang : S1

Judul Karya Ilmiah : ANALISIS SENTIMEN BERBASIS
ASPEK TERHADAP ULASAN APLIKASI DUOLINGO
MENGUNAKAN ALGORITMA SVM DAN INDOBERT

Menyatakan dengan sesungguhnya bahwa saya bersedia* (**pilih salah satu**):

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
 - ☐ Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - ☐ Embargo publikasi karya ilmiah dalam kurun waktu 3 tahun.

Tangerang, 26 Juni 2025


Aura Lintang Pembayun Pinasti Aji

* Pilih salah satu

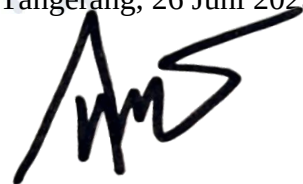
** Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

KATA PENGANTAR

Puji Syukur atas selesainya penulisan Tugas Akhir ini dengan judul: “Analisis Sentimen Berbasis Aspek Terhadap Ulasan Aplikasi Duolingo Menggunakan Algoritma SVM dan IndoBERT” dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Jurusan Sistem Informasi Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan tugas akhir ini, sangatlah sulit bagi saya untuk menyelesaikan tugas akhir ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Dr. Ir. Andrey Andoko, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Ibu Monika Evelin Johan, S.Kom., M.M.S.I., sebagai Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Ayah dan Bunda yang selalu memberikan dukungan, semangat, kasih sayang, dan cinta kepada penulis sehingga penulis dapat menyelesaikan studi hingga akhir. Terima kasih atas seluruh doa yang dipanjatkan tiada henti untuk penulis.
6. Adik penulis, Kharisma Bintang Nastiti Aji sebagai penghibur dan penyemangat penulis di rumah. Terima kasih telah bersedia untuk menemani penulis selama penyusunan Tugas Akhir.
7. Teman-teman yang selalu mendukung dan memberikan semangat kepada penulis sehingga penulis dapat menyelesaikan Tugas Akhir ini.

Tangerang, 26 Juni 2025



Aura Lintang Pembayun Pinasti Aji

ANALISIS SENTIMEN BERBASIS ASPEK TERHADAP ULASAN APLIKASI DUOLINGO MENGGUNAKAN ALGORITMA SVM DAN INDOBERT

Aura Lintang Pembayun Pinasti Aji

ABSTRAK

Pada tahun 2022, Indonesia berada di peringkat ke-81 dari 111 negara dalam hal kecakapan berbahasa Inggris, menunjukkan bahwa peningkatan kemampuan berbahasa asing masih menjadi kebutuhan. Salah satu solusi yang digunakan masyarakat adalah Duolingo, aplikasi pembelajaran bahasa asing populer yang telah diunduh lebih dari 100 juta kali di Google Play Store. Meskipun begitu, ulasan pengguna tetap menjadi pertimbangan penting bagi calon pengguna untuk menilai efektivitas aplikasi sebelum mengunduhnya.

Penelitian ini melakukan analisis sentimen berbasis aspek terhadap ulasan pengguna Duolingo, mengacu pada pedoman *Core App Quality* dari Android Developers. Proses analisis dengan framework CRISP-DM dengan menggunakan algoritma *Support Vector Machine* (SVM) untuk *training* klasifikasi dan model IndoBERT untuk ekstraksi fitur. Evaluasi performa dilakukan menggunakan metrik seperti *accuracy*, *precision*, *recall*, dan *f1-score* untuk menilai kemampuan model dalam mengklasifikasikan aspek dan sentimen dari ulasan pengguna.

Melalui analisis yang telah dilakukan dengan data ulasan Duolingo di Google Play Store antara tahun 2020 hingga 2024, diperoleh 10,561 data ulasan berbahasa Indonesia yang dikelompokkan ke dalam lima aspek berdasarkan Core App Quality. Hasil analisis menunjukkan bahwa kombinasi SVM dan IndoBERT mampu mencapai akurasi tinggi, yaitu 99% untuk klasifikasi sentimen dan 97% untuk klasifikasi aspek pada dua dataset (dataset 1 dan 2), serta 98% dan 96% pada dataset 3. Meskipun data tidak seimbang, penggunaan metode *oversampling* ADASYN berhasil meningkatkan performa model secara signifikan. Temuan ini memperlihatkan bahwa Duolingo mendapatkan tanggapan positif dari pengguna, dengan mayoritas ulasan bersentimen positif dan aspek fungsionalitas menjadi aspek yang paling sering dibahas. Penelitian ini membuktikan bahwa penggunaan algoritma SVM dan model IndoBERT pada analisis sentimen opini pengguna efektif dalam mengklasifikasikan sentimen dan aspek secara akurat, serta mampu memahami persepsi pengguna berdasarkan aspek-aspek aplikasi yang dapat digunakan sebagai dasar perbaikan pengembang.

Kata kunci: ADASYN, Aspek, Duolingo, IndoBERT, Sentimen.

ASPECT-BASED SENTIMENT ANALYSIS OF DUOLINGO APPLICATION REVIEWS USING SVM AND INDOBERT ALGORITHMS

Aura Lintang Pembayun Pinasti Aji

ABSTRACT (English)

In 2022, Indonesia ranked 81st out of 111 countries in terms of English proficiency, indicating that improving foreign language skills remains a significant need. One solution widely used by the public is Duolingo, a popular language learning application that has been downloaded more than 100 million times on the Google Play Store. Nevertheless, user reviews remain an important consideration for potential users to assess the app's effectiveness before downloading it.

This research conducts an aspect-based sentiment analysis on Duolingo user reviews, referring to the Core App Quality guidelines from Android Developers. The analysis follows the CRISP-DM framework, utilizing the Support Vector Machine (SVM) algorithm for sentiment classification training and the IndoBERT model for feature extraction. Performance evaluation is conducted using metrics such as accuracy, precision, recall, and F1-score to assess the model's ability to classify both aspects and sentiments of user reviews.

Through the analysis conducted on Duolingo reviews from the Google Play Store between 2020 and 2024, a total of 10,561 Indonesian-language reviews were obtained and categorized into five aspects based on the Core App Quality guidelines. The analysis results show that the combination of SVM and IndoBERT achieved high accuracy: 99% for sentiment classification and 97% for aspect classification on two datasets (dataset 1 and 2), as well as 98% and 96% on dataset 3. Although the data was imbalanced, the use of the ADASYN oversampling method significantly improved the model's performance. These findings indicate that Duolingo receives a generally positive response from users, with most reviews expressing positive sentiment and the functionality aspect being the most frequently discussed. This study demonstrates that using the SVM algorithm and the IndoBERT model for sentiment analysis is effective in accurately classifying sentiments and aspects, and capable of capturing user perceptions based on app-related aspects, which can serve as a basis for developer improvements.

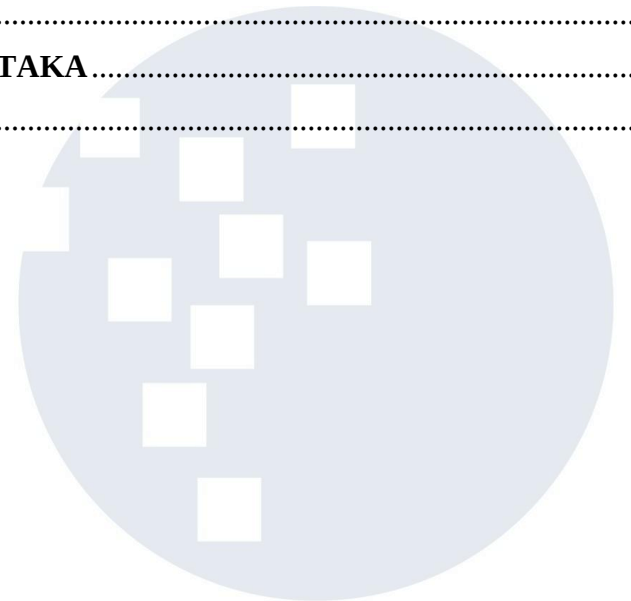
Keywords: ADASYN, Aspect, Duolingo, IndoBERT, Sentiment

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
KATA PENGANTAR	v
ABSTRAK	vi
ABSTRACT (English)	vii
DAFTAR ISI	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xv
DAFTAR RUMUS	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah	6
1.4 Tujuan dan Manfaat Penelitian	6
1.4.1 Tujuan Penelitian	6
1.4.2 Manfaat Penelitian	6
1.5 Sistematika Penulisan	7
BAB II LANDASAN TEORI	9
2.1 Penelitian Terkait	9
2.2 Teori Penelitian	14
2.2.1 Duolingo	14
2.2.2 Analisis Sentimen	16
2.2.3 Text Mining	18
2.3 Framework dan Algoritma Penelitian	19
2.3.1 Support Vector Machine (SVM)	19
2.3.2 Bidirectional Encoder Representations from Transformers (BERT)	20
2.3.3 IndoBERT	20

2.3.4	Confusion Matrix	21
2.3.5	Word2Vec	23
2.3.6	Stratified K-Fold Cross-Validation (SKCV)	24
2.3.7	Adaptive Synthetic Sampling (ADASYN).....	24
2.3.8	CRISP-DM.....	25
2.4	Tools dan Software Penelitian.....	27
2.4.1	Google Play Store	27
2.4.2	Python	28
2.4.3	Google Colaboratory.....	28
2.4.4	Microsoft Excel.....	28
BAB III METODOLOGI PENELITIAN		30
3.1	Gambaran Umum Objek Penelitian.....	30
3.2	Metode Penelitian	30
3.2.1	Alur Penelitian.....	30
3.2.2	Metode Data Mining	32
3.2.3	Metode Pengolahan Data.....	37
3.3	Teknik Pengumpulan Data.....	38
3.3.1	Populasi dan Sampel.....	38
3.3.2	Periode Pengambilan Data	39
3.4	Teknik Analisis Data	39
3.4.1	Variabel Penelitian.....	39
3.5	Teknik Pengujian atau Validasi Sistem.....	40
3.5.1	Tools Analisis Data.....	40
3.5.2	Accuracy, Precision, Recall, dan F-1 Score	41
BAB IV ANALISIS DAN HASIL PENELITIAN		42
4.1	Spesifikasi Sistem	42
4.2	Implementasi Sistem	42
4.2.1	Business Understanding	42
4.2.2	Data Understanding	43
4.2.3	Data Preparation.....	49
4.2.4	Data Modeling	51
4.2.5	Evaluasi Model	56

4.3 Uji Coba.....	57
4.3.1 Skenario Uji Coba.....	57
4.4 Evaluasi Hasil	60
BAB V SIMPULAN DAN SARAN	88
5.1 Simpulan.....	88
5.2 Saran.....	89
DAFTAR PUSTAKA.....	90
LAMPIRAN.....	97

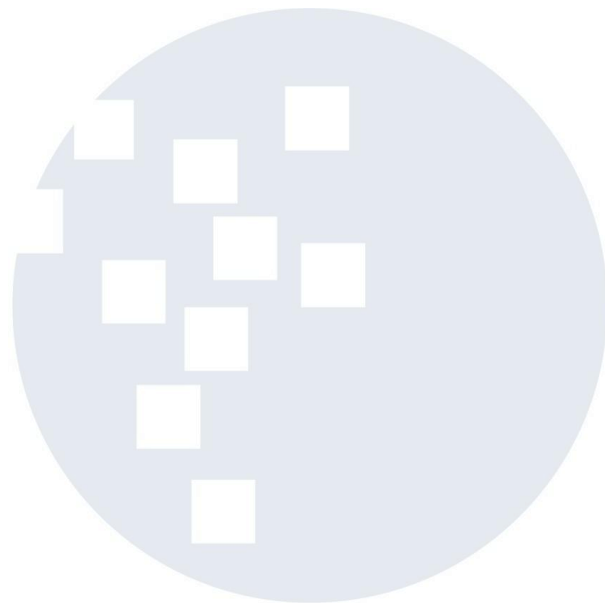


UMN
 UNIVERSITAS
 MULTIMEDIA
 NUSANTARA

DAFTAR TABEL

Tabel 1. 1 Tabel Perbandingan aplikasi Duolingo dan aplikasi belajar lainnya	3
Tabel 2. 1 Tabel Artikel Penelitian Terkait.....	9
Tabel 4. 1 Hasil data yang discraping	44
Tabel 4. 2 Keywords tiap Aspek	47
Tabel 4. 3 Total Ulasan berdasarkan Aspek.....	47
Tabel 4. 4 Total Data Ulasan.....	48
Tabel 4. 5 Hasil pembersihan teks	49
Tabel 4. 6 Hasil Stopwords Removal.....	50
Tabel 4. 7 Hasil Tokenisasi.....	50
Tabel 4. 8 Hasil Stemming.....	51
Tabel 4. 9 Hasil mapping label sentimen dan aspek	52
Tabel 4. 10 Tabel perbedaan jumlah data sentimen	59
Tabel 4. 11 Tabel perbedaan jumlah data aspek	59
Tabel 4. 12 Classification report sentimen data yang tidak seimbang.....	61
Tabel 4. 13 Classification report aspek data yang tidak seimbang	62
Tabel 4. 14 Classification report sentimen setelah cross-validation	62
Tabel 4. 15 Classification report aspek setelah cross-validation	63
Tabel 4. 16 Classification report sentimen setelah ADASYN.....	63
Tabel 4. 17 Classification report aspek setelah ADASYN	64
Tabel 4. 18 Classification report sentimen 2 data yang tidak seimbang.....	67
Tabel 4. 19 Classification report aspek 2 data yang tidak seimbang	67
Tabel 4. 20 Classification report sentimen 2 setelah cross-validation	67
Tabel 4. 21 Classification report aspek 2 setelah cross-validation	68
Tabel 4. 22 Classification report sentimen 2 setelah ADASYN.....	68
Tabel 4. 23 Classification report aspek 2 setelah ADASYN	68
Tabel 4. 24 Classification report sentimen 3 data tidak seimbang.....	72
Tabel 4. 25 Classification report aspek 3 data tidak seimbang.....	72
Tabel 4. 26 Classification report sentimen 3 setelah cross-validation	72
Tabel 4. 27 Classification report aspek 3 setelah cross-validation	73
Tabel 4. 28 Classification report sentimen 3 setelah ADASYN.....	73
Tabel 4. 29 Classification report aspek 3 setelah ADASYN	74
Tabel 4. 30 Classification report SVM sentimen dataset 1	76
Tabel 4. 31 Classification report SVM aspek dataset 1	76
Tabel 4. 32 Classification report SVM sentimen dataset 2.....	76
Tabel 4. 33 Classification report SVM aspek dataset 2	76
Tabel 4. 34 Classification report SVM sentimen dataset 3.....	77
Tabel 4. 35 Classification report SVM aspek dataset 3	77
Tabel 4. 36 Evaluasi hasil sentimen dataset 1.....	78
Tabel 4. 37 Evaluasi hasil sentimen dataset 2.....	79
Tabel 4. 38 Evaluasi hasil sentimen dataset 3.....	81
Tabel 4. 39 Evaluasi hasil aspek dataset 1	82

Tabel 4. 40 Evaluasi hasil aspek dataset 2	84
Tabel 4. 41 Evaluasi hasil aspek dataset 3	85
Tabel 4. 42 Tabel perbandingan hasil akurasi.....	87

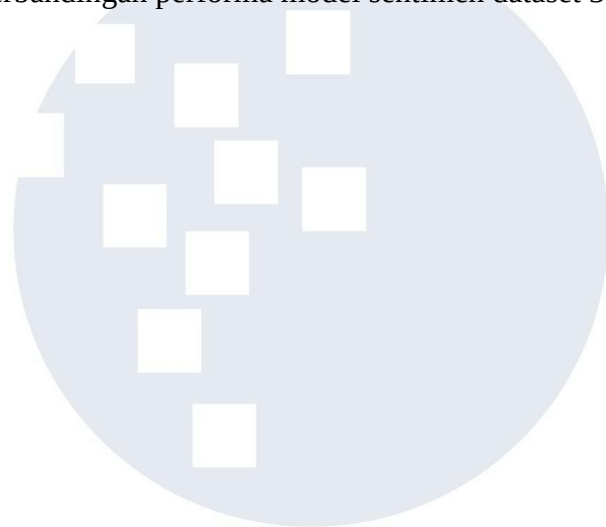


UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR GAMBAR

Gambar 1. 1 Diagram jumlah pengguna aktif Duolingo.....	2
Gambar 2. 1 Logo aplikasi Duolingo.....	14
Gambar 2. 2 Cara Kerja SVM.....	19
Gambar 2. 3 Tabel Confusion Matrix	21
Gambar 2. 4 Siklus pada CRISP-DM	25
Gambar 2. 5 Logo Google Play Store	27
Gambar 2. 6 Logo Python	28
Gambar 3. 1 Diagram Alur Penelitian.....	31
Gambar 3. 2 Flowchart Tahap Data Understanding.....	33
Gambar 3. 3 Flowchart Tahap Data Preparation.....	34
Gambar 3. 4 Flowchart Tahap Data Modeling.....	35
Gambar 3. 5 Tahap modeling tanpa IndoBERT.....	36
Gambar 4. 1 Kode Scraping Data Duolingo	43
Gambar 4. 2 Kode untuk Sortir Tahun dan Kolom.....	44
Gambar 4. 3 Kode untuk Labeling Sentimen.....	45
Gambar 4. 4 Kode untuk mendapatkan Keywords	46
Gambar 4. 5 Kode untuk melakukan labeling aspek	46
Gambar 4. 6 Hasil clustering menggunakan KMeans.....	48
Gambar 4. 7 Kode pembersihan teks	49
Gambar 4. 8 Potongan kode stopwords removal	50
Gambar 4. 9 Kode Tokenisasi.....	50
Gambar 4. 10 Kode Stemming.....	51
Gambar 4. 11 Library yang digunakan pada data modeling	52
Gambar 4. 12 Kode encoding label aspek dan sentimen	52
Gambar 4. 13 Potongan kode word embedding	53
Gambar 4. 14 Kode embedding dengan IndoBERT	55
Gambar 4. 15 Kode splitting data	55
Gambar 4. 16 Kode pelatihan data klasifikasi aspek dan sentimen	56
Gambar 4. 17 Kode evaluasi validasi.....	56
Gambar 4. 18 Kode evaluasi akhir	57
Gambar 4. 19 Kode implementasi stratified k-fold.....	58
Gambar 4. 20 Kode pelatihan data yang telah di cross-validation.....	58
Gambar 4. 21 Kode penerapan ADASYN	58
Gambar 4. 22 Kode pelatihan data pada data yang telah di oversampling	59
Gambar 4. 23 Distribusi kelas sentimen dataset 1	60
Gambar 4. 24 Distribusi kelas aspek dataset 1.....	61
Gambar 4. 25 Confusion matrix sentimen dataset 1	64
Gambar 4. 26 Confusion matrix aspek dataset 1.....	65
Gambar 4. 27 Distribusi kelas sentimen dataset 2	66
Gambar 4. 28 Distribusi kelas aspek dataset 2.....	66

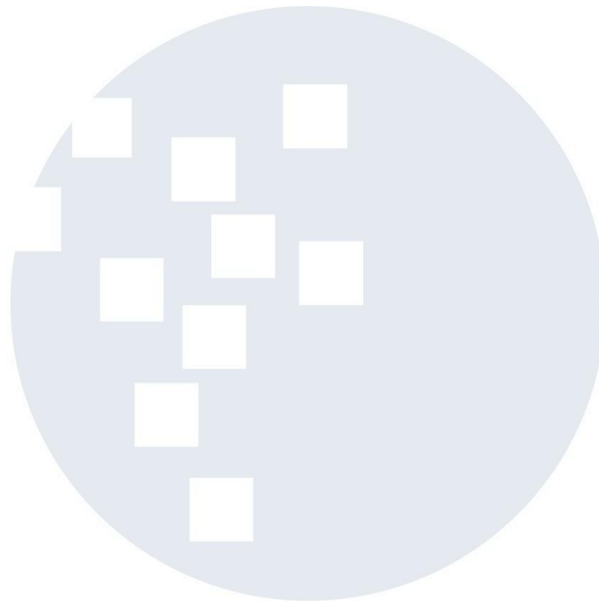
Gambar 4. 29 Confusion matrix sentimen dataset 2	69
Gambar 4. 30 Confusion matrix aspek dataset 2.....	70
Gambar 4. 31 Distribusi kelas sentimen dataset 3	71
Gambar 4. 32 Distribusi kelas aspek dataset 3.....	71
Gambar 4. 33 Confusion matrix sentimen dataset 3	74
Gambar 4. 34 Confusion matrix aspek dataset 3.....	75
Gambar 4. 35 Perbandingan performa model sentimen dataset 1.....	79
Gambar 4. 36 Perbandingan performa model sentimen dataset 2.....	81
Gambar 4. 37 Perbandingan performa model sentimen dataset 3.....	82



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

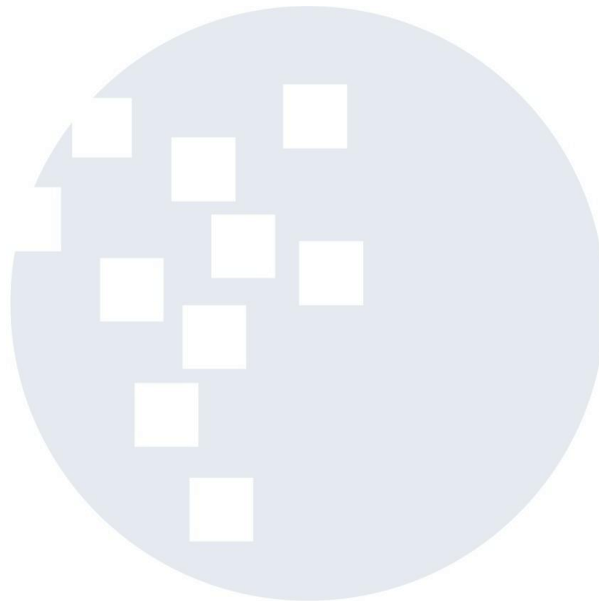
Lampiran A Turnitin Similarity Report	97
Lampiran B Form Konsultasi Bimbingan	104



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR RUMUS

Rumus 2. 1 Rumus Accuracy	22
Rumus 2. 2 Rumus F1-Score	22
Rumus 2. 3 Rumus Recall	23
Rumus 2. 4 Rumus Precision	23



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA