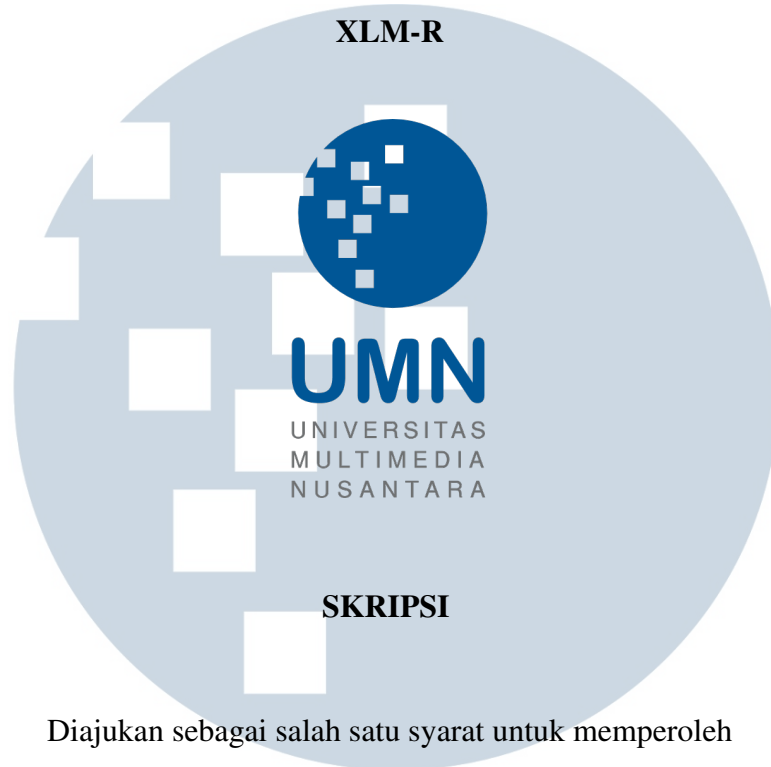


**IMPLEMENTASI KLASIFIKASI SENTIMEN ULASAN PRODUK  
E-COMMERCE LAZADA BAHASA INDONESIA MENGGUNAKAN**

**XLM-R**



**SKRIPSI**

Diajukan sebagai salah satu syarat untuk memperoleh  
Gelar Sarjana Komputer (S.Kom.)

**Evan**

**00000035947**

**UMN**

**UNIVERSITAS**

**MULTIMEDIA**

**NUSANTARA**

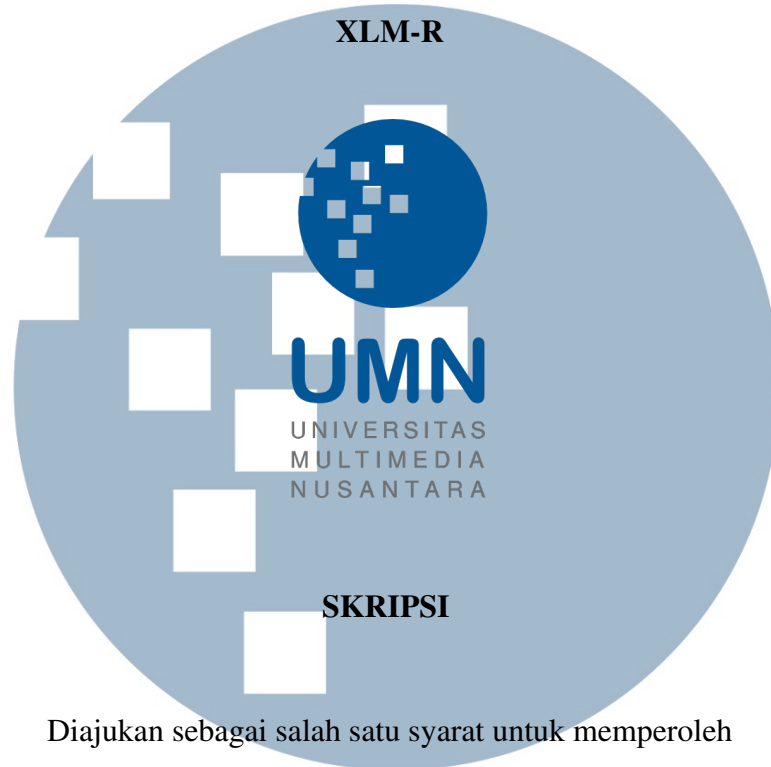
**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA**

**TANGERANG**

**2023**

**IMPLEMENTASI KLASIFIKASI SENTIMEN ULASAN PRODUK  
E-COMMERCE LAZADA BAHASA INDONESIA MENGGUNAKAN**

**XLM-R**



**SKRIPSI**

Diajukan sebagai salah satu syarat untuk memperoleh  
Gelar Sarjana Komputer (S.Kom.)

**Evan**

**00000035947**

**UMMN**

**UNIVERSITAS  
MULTIMEDIA  
NUSANTARA**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA**

**TANGERANG**

**2023**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Evan  
Nomor Induk Mahasiswa : 00000035947  
Program Studi : Informatika

Skripsi dengan judul:

**Implementasi Klasifikasi Sentimen Ulasan Produk E-Commerce Lazada Bahasa Indonesia Menggunakan XLM-R**

Merupakan hasil karya saya sendiri bukan plagiat dari karya 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 Skripsi maupun dalam penulisan laporan Skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas akhir yang telah saya tempuh.

Tangerang, 6 Juni 2023



(Evan)

U M M N  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## HALAMAN PENGESAHAN

Skripsi dengan judul  
**IMPLEMENTASI KLASIFIKASI SENTIMEN ULASAN PRODUK  
E-COMMERCE LAZADA BAHASA INDONESIA MENGGUNAKAN  
XLM-R**

oleh

Nama : Evan  
NIM : 00000035947  
Program Studi : Informatika  
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Selasa, 13 Juni 2023

Pukul 08.00 s/s 12.00 dan dinyatakan

**LULUS**

Dengan susunan penguji sebagai berikut

Ketua Sidang

  
(Farica Perdana Putri, S.Kom., M.Sc.)


NIDN: 331019301

Pembimbing I

  
(Andre Rusli, S.Kom., M.Sc.)


NIDN: 319069201

Penguji

  
(Eunike Endariahna Surbakti, S.Kom.,  
M.T.I.)

NIDN: 0322099401

Pembimbing II

  
(Alexander Waworuntu, S.Kom., M.T.I.)

NIDN: 309068503

Ketua Program Studi Informatika,

  
(Marlinda Vasty Overbeek, S.Kom., M.Kom.)

NIDN: 0818038501

iv

Implementasi Klasifikasi Sentimen..., Evan, Universitas Multimedia Nusantara

## KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Implementasi Klasifikasi Sentimen Ulasan Produk E-Commerce Lazada Bahasa Indonesia Menggunakan XLM-R dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

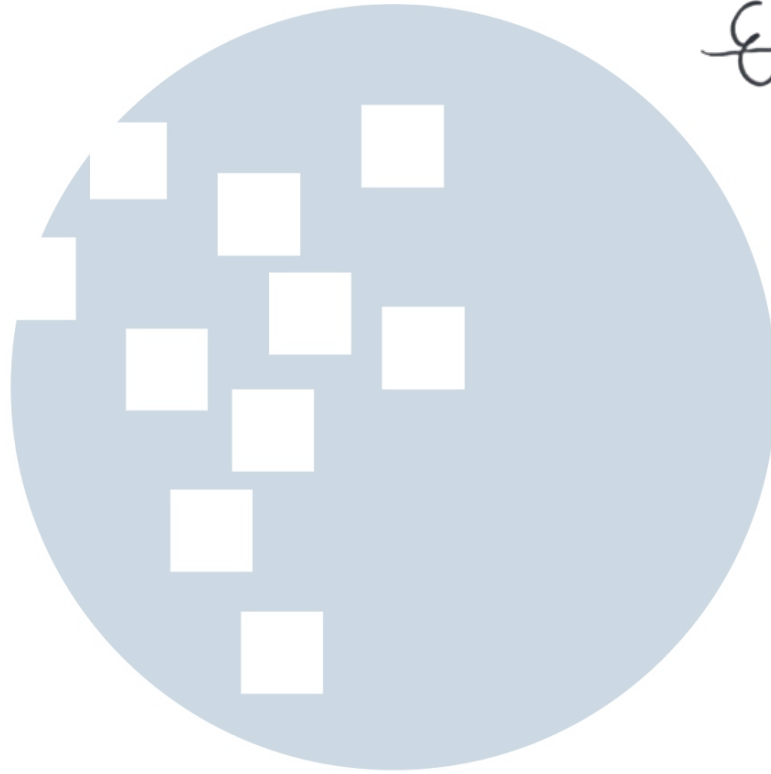
1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Marlinda Vasty Overbeek, S.Kom., M.Kom., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak Andre Rusli, S.Kom., M.Sc., sebagai Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya tesis ini.
5. Bapak Alexander Waworuntu, S.Kom., M.T.I., sebagai Pembimbing kedua yang telah banyak membantu dan memberikan bimbingan atas terselesainya Skripsi/Tesis ini.
6. Orang Tua saya yang telah memberikan bantuan dukungan moral, sehingga penulis dapat menyelesaikan tesis ini.
7. Vera Felia sebagai orang yang terus mendukung saya secara moral hingga akhirnya saya dapat menyelesaikan laporan ini.

Semoga skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 6 Juni 2023



Evan



UMMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA



**IMPLEMENTASI KLASIFIKASI SENTIMEN ULASAN PRODUK  
E-COMMERCE LAZADA BAHASA INDONESIA MENGGUNAKAN  
XLM-R**

**ABSTRAK**

Analisis sentimen merupakan salah satu cara untuk mengetahui pendapat seseorang atau sekelompok orang terkait suatu topik, produk, layanan, dan lain-lain. Analisis sentimen juga merupakan topik yang penting dalam *Natural Language Processing*. Dalam penelitian yang dilakukan ini, analisis sentimen dilakukan dengan menggunakan model XLM-R untuk melakukan analisis sentimen terhadap *review* produk berbahasa Indonesia. Model XLM-R merupakan model yang telah dilatih dengan 100 bahasa sebelumnya, sehingga memiliki kemampuan untuk melakukan klasifikasi sentimen teks multi bahasa. Pada penelitian ini, Bahasa Indonesia digunakan untuk jadi objek penelitian karena Bahasa Indonesia merupakan bahasa yang masuk ke dalam bahasa dengan sumber daya yang rendah. Maka dari itu, penelitian ini bertujuan untuk melihat performa model XLM-R dalam melakukan klasifikasi sentimen untuk bahasa yang sumber dayanya rendah. Hasil dari penelitian menunjukkan bahwa model XLM-R dapat dengan baik melakukan klasifikasi teks Bahasa Indonesia, meskipun dilatih hanya dengan bahasa selain Bahasa Indonesia. Bahasa-bahasa yang digunakan selama pelatihan adalah bahasa Jerman, Inggris, Spanyol, dan Prancis. Akurasi model tertinggi diraih oleh model XLM-R yang dilatih dengan bahasa Prancis dengan nilai akurasi sebesar 83.25%. Pada penelitian ini juga disimpulkan bahwa model dapat dengan baik melakukan *transfer learning*. Selain itu, dapat disimpulkan juga bahwa penambahan data berbahasa Indonesia untuk melatih model sangat berpengaruh terhadap akurasi yang di dapat, tapi model XLM-R yang dilatih tanpa campuran Bahasa Indonesia sendiri juga bisa dikatakan memiliki akurasi yang cukup bersaing.

**Kata kunci:** *Analisis Sentimen, Bahasa Indonesia, Natural Language Processing, XLM-R*

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A

# Implementation of Sentiment Classification for Product Reviews in Lazada E-Commerce using XLM-R

Evan

## ABSTRACT

Sentiment analysis is a crucial method for understanding people's opinions and sentiments on various topics, products, services, and more. It plays a significant role in Natural Language Processing (NLP). In this research, sentiment analysis is conducted using the XLM-R model to analyze the sentiment of Indonesian language product reviews. The XLM-R model has been trained on 100 languages, enabling it to perform multi-language sentiment classification. The focus of this study is on the Indonesian language, which falls into the category of low-resource languages. The objective is to evaluate the performance of the XLM-R model in sentiment classification for a low-resource language. The results of the research demonstrate that the XLM-R model performs well in classifying Indonesian language texts, even when trained solely on languages other than Indonesian. The languages used for training include German, English, Spanish, and French. The highest accuracy is achieved by the XLM-R model trained with the French language, achieving an accuracy score of 83.25%. The research also concludes that the XLM-R model exhibits effective transfer learning capabilities. Furthermore, it is observed that the addition of Indonesian language data during model training significantly influences the accuracy obtained. However, it is worth noting that the XLM-R model trained without incorporating Indonesian language data still exhibits competitive accuracy. In summary, this research highlights the use of the XLM-R model for sentiment analysis of Indonesian language product reviews. The findings demonstrate the model's ability to effectively classify Indonesian texts, despite being trained primarily on languages other than Indonesian.

**Keywords:** *Indonesian Language, Natural Language Processing, Sentiment Analysis, XLM-R*

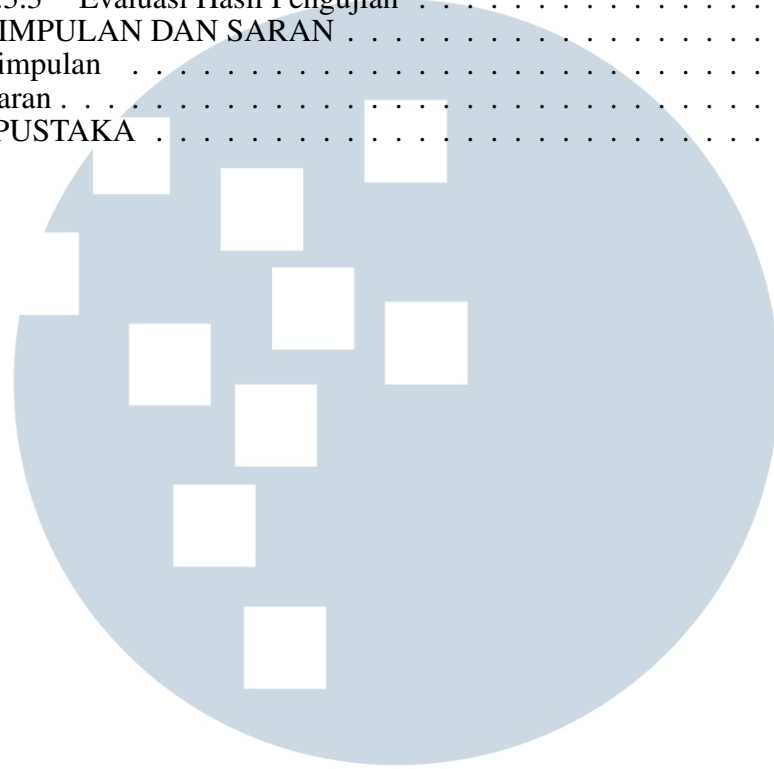
U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A



## DAFTAR ISI

HALAMAN JUDUL . . . . .	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT . . . . .	ii
HALAMAN PENGESAHAN . . . . .	iii
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH . . . . .	iv
HALAMAN PERSEMBAHAN/MOTTO . . . . .	v
KATA PENGANTAR . . . . .	vi
ABSTRAK . . . . .	viii
ABSTRACT . . . . .	ix
DAFTAR ISI . . . . .	x
DAFTAR GAMBAR . . . . .	xii
DAFTAR TABEL . . . . .	xiii
DAFTAR LAMPIRAN . . . . .	xiv
BAB 1 PENDAHULUAN . . . . .	1
1.1 Latar Belakang Masalah . . . . .	1
1.2 Rumusan Masalah . . . . .	3
1.3 Batasan Permasalahan . . . . .	4
1.4 Tujuan Penelitian . . . . .	4
1.5 Manfaat Penelitian . . . . .	4
1.6 Sistematika Penulisan . . . . .	5
BAB 2 LANDASAN TEORI . . . . .	6
2.1 <i>Multilingual Text Classification</i> . . . . .	6
2.2 XLM-R . . . . .	6
2.3 Transfer Learning . . . . .	7
2.4 Metrik Evaluasi . . . . .	7
BAB 3 METODOLOGI PENELITIAN . . . . .	10
3.1 Metodologi Penelitian . . . . .	10
3.2 Perancangan Sistem . . . . .	11
3.2.1 Flowchart Utama . . . . .	12
3.2.2 Pra Proses Teks . . . . .	13
3.2.3 Pembersihan Teks . . . . .	14
3.2.4 Pembagian Dataset . . . . .	15
3.2.5 Melatih Model . . . . .	17
3.2.6 Validasi Model . . . . .	19
3.2.7 Evaluasi . . . . .	20
BAB 4 HASIL DAN DISKUSI . . . . .	21
4.1 Spesifikasi sistem . . . . .	21
4.2 Potongan Kode . . . . .	21
4.2.1 <i>Text Cleaning</i> . . . . .	21
4.2.2 <i>Tokenizer</i> . . . . .	22
4.2.3 <i>Padding</i> . . . . .	23
4.2.4 <i>Create Attention Mask</i> . . . . .	24
4.2.5 <i>Split Dataset</i> . . . . .	24
4.2.6 <i>Convert to Tensor and Dataloader</i> . . . . .	25
4.2.7 <i>Load Model</i> . . . . .	25
4.2.8 <i>Scheduler</i> . . . . .	27
4.2.9 <i>Train and Validate Model</i> . . . . .	27
4.2.10 <i>Testing Model</i> . . . . .	31
4.3 Uji Sistem . . . . .	32

4.3.1	Skenario Pengujian . . . . .	32
4.3.2	Hasil Pengujian . . . . .	33
4.3.3	Evaluasi Hasil Pengujian . . . . .	35
BAB 5	SIMPULAN DAN SARAN . . . . .	36
5.1	Simpulan . . . . .	36
5.2	Saran . . . . .	37
	DAFTAR PUSTAKA . . . . .	38



UMMN  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

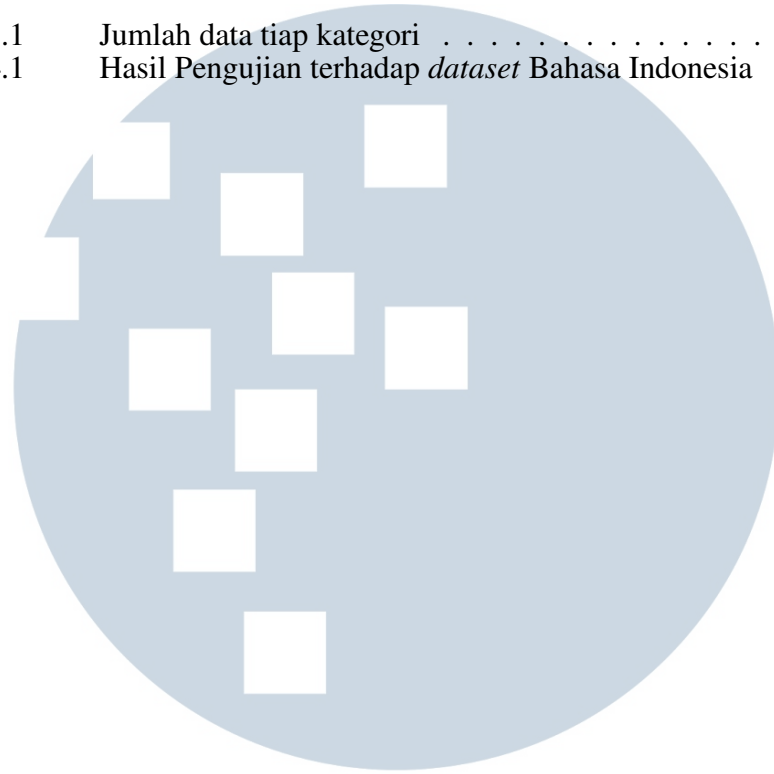
## DAFTAR GAMBAR

Gambar 2.1	Gambaran Masked Language Model . . . . .	7
Gambar 3.1	<i>Flowchart</i> Utama . . . . .	12
Gambar 3.2	<i>Flowchart</i> Pra Proses Teks . . . . .	13
Gambar 3.3	<i>Flowchart</i> Pembersihan Teks . . . . .	14
Gambar 3.4	<i>Flowchart</i> Pembagian Dataset . . . . .	15
Gambar 3.5	<i>Train Model</i> . . . . .	17
Gambar 3.6	Validasi Model . . . . .	19
Gambar 3.7	<i>Flowchart</i> Evaluasi . . . . .	20
Gambar 4.1	Kode <i>text cleaning</i> . . . . .	22
Gambar 4.2	Kode <i>tokenizer</i> . . . . .	22
Gambar 4.3	Contoh Penggunaan <i>Tokenizer</i> . . . . .	23
Gambar 4.4	Kode Proses Tokenisasi pada Dataset . . . . .	23
Gambar 4.5	Kode Proses <i>Padding</i> pada Token ID . . . . .	24
Gambar 4.6	Kode Proses pembuatan <i>Attention Mask</i> . . . . .	24
Gambar 4.7	Kode <i>Split Dataset</i> . . . . .	24
Gambar 4.8	Kode <i>Convert Dataset to Tensor</i> . . . . .	25
Gambar 4.9	Kode <i>Load Model</i> . . . . .	26
Gambar 4.10	Kode <i>Optimizer</i> . . . . .	26
Gambar 4.11	Kode <i>Scheduler</i> . . . . .	27
Gambar 4.12	Kode Perulangan <i>Epoch</i> . . . . .	27
Gambar 4.13	Kode Perulangan <i>Data Train</i> . . . . .	28
Gambar 4.14	Kode Menghitung Rata-Rata <i>Training Loss</i> . . . . .	29
Gambar 4.15	Kode Evaluasi Model . . . . .	30
Gambar 4.16	Kode Prediksi . . . . .	31



## DAFTAR TABEL

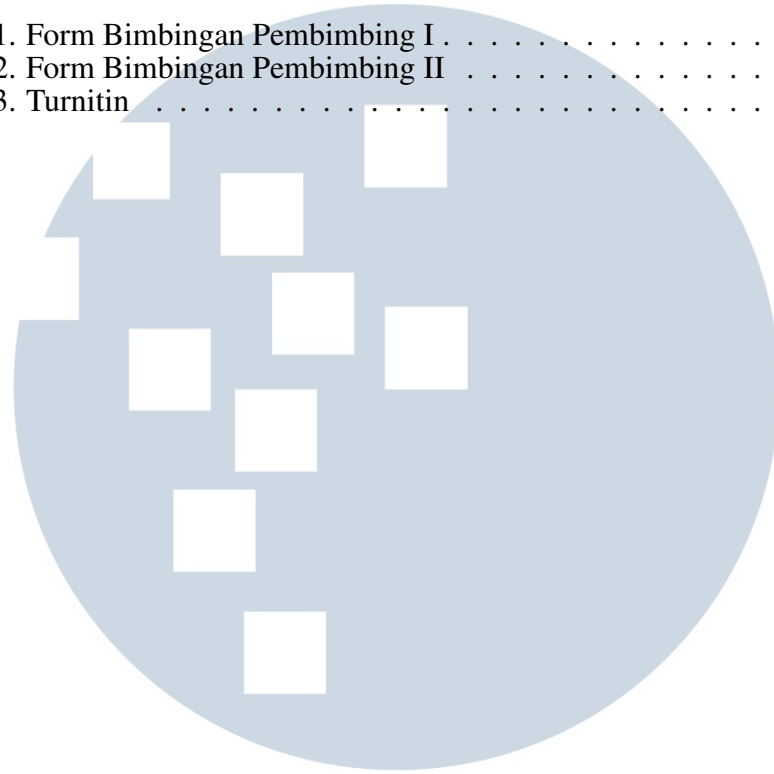
Tabel 3.1	Jumlah data tiap kategori . . . . .	10
Tabel 4.1	Hasil Pengujian terhadap <i>dataset</i> Bahasa Indonesia . . . . .	33



**UMMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## DAFTAR LAMPIRAN

Lampiran 1. Form Bimbingan Pembimbing I . . . . .	39
Lampiran 2. Form Bimbingan Pembimbing II . . . . .	41
Lampiran 3. Turnitin . . . . .	42



# UMMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA