

**KLASIFIKASI GENDER BERDASARKAN NAMA  
INDONESIA MENGGUNAKAN MODEL INDOBERT  
DARI INDONLU: STUDI KASUS PADA CUSTOMER  
PROFILE KOMPAS GRAMEDIA**



**LAPORAN MBKM MAGANG**

**LEONARDO TYOES HUIBU  
00000065503**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA  
TANGERANG  
2025**

**KLASIFIKASI GENDER BERDASARKAN NAMA  
INDONESIA MENGGUNAKAN MODEL INDOBERT  
DARI INDONLU: STUDI KASUS PADA CUSTOMER  
PROFILE KOMPAS GRAMEDIA**



**LAPORAN MBKM MAGANG**

**LEONARDO TYOES HUIBU  
00000065503**

**UNIVERSITAS  
MULTIMEDIA  
NUSANTARA**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA  
TANGERANG  
2025**

## HALAMAN PERNYATAAN ORISINALITAS TIDAK PLAGIAT

Dengan ini saya,

Nama : Leonardo Tyoes Huibu

NIM : 00000065503

Program Studi : Informatika

Menyatakan dengan sesungguhnya bahwa Laporan MBKM Magang saya yang berjudul:

**Klasifikasi Gender Berdasarkan Nama Indonesia Menggunakan Model IndoBERT dari IndoNLU: Studi Kasus pada Customer Profile Kompas Gramedia**

merupakan hasil karya saya sendiri, bukan merupakan hasil plagiat, dan tidak pula dituliskan oleh orang lain; Semua sumber, baik yang dikutip maupun dirujuk, telah saya cantumkan dan nyatakan dengan benar pada bagian Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan karya ilmiah, saya bersedia menerima konsekuensi untuk dinyatakan TIDAK LULUS. Saya juga bersedia menanggung segala konsekuensi hukum yang berkaitan dengan tindak plagiarisme ini sebagai kesalahan saya pribadi dan bukan tanggung jawab Universitas Multimedia Nusantara.

Tangerang, 19 Juni 2025



(Leonardo Tyoes Huibu)

## HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH MAHASISWA

Yang bertanda tangan di bawah ini:

Nama : Leonardo Tyoes Huibu

NIM : 00000065503

Program Studi : Informatika

Jenjang : S1

Jenis Karya : Laporan MBKM Magang

Menyatakan dengan sesungguhnya bahwa:

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya di repositori Knowledge Center, sehingga dapat diakses oleh Civitas Akademika/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial dan saya juga tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.
- ☐ Saya tidak bersedia karena dalam proses pengajuan untuk diterbitkan ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)\*\*.

Tangerang, 19 Juni 2025

Yang menyatakan



Leonardo Tyoes Huibu

UMN  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

\*\* Jika tidak bisa membuktikan LoA jurnal/HKI selama enam bulan ke depan, saya bersedia mengizinkan penuh karya ilmiah saya untuk diunggah ke KC UMN dan menjadi hak institusi UMN.

## Halaman Persembahan / Motto

"A good name is to be more desired than great wealth, Favor is better than silver and gold."

Proverbs 22:1 (NASB)



## KATA PENGANTAR

Puji dan syukur panjatkan ke hadirat Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya, sehingga dapat menyelesaikan laporan magang ini dengan baik. Tugas Akhir ini tidak hanya disusun untuk memenuhi syarat akademik maupun kebutuhan perusahaan, tetapi juga sebagai upaya untuk memperdalam pemahaman terhadap konsep deep learning dan mengeksplorasi bagaimana model dengan struktur yang kompleks mampu menangani permasalahan nyata secara efektif. Ucapan terima kasih disampaikan kepada pihak-pihak sebagai berikut:

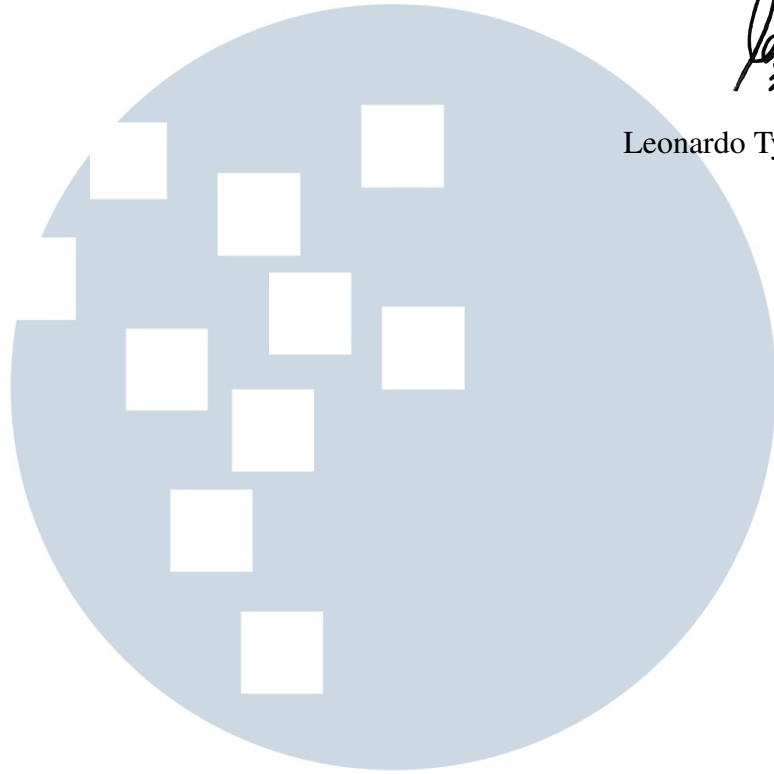
1. Bapak Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu Sy Yuliani Yakub, S.Kom., M.T. PhD, sebagai Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya tesis ini. memberikan bimbingan atas terselesainya laporan magang ini.
5. Kepada Bapak Agung Wicaksono sebagai Supervisor, serta Bapak Falah Novayanda Adlin sebagai Mentor, atas dukungan moral, bantuan dalam perancangan awal, dan masukan yang sangat membantu dalam proses diskusi dan penyusunan proyek ini.
6. Orang Tua saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tesis ini.

Harapannya, karya ilmiah ini dapat mendorong lebih banyak generasi muda untuk tertarik melakukan analisis dan mengembangkan proyek-proyek lanjutan berbasis deep learning dan machine learning, serta menjadi referensi yang bermanfaat bagi penelitian selanjutnya dalam menerapkan model ini pada berbagai jenis tugas yang berbeda.

Tangerang, 19 Juni 2025



Leonardo Tyoes Huibu



UMN  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

**KLASIFIKASI GENDER BERDASARKAN NAMA INDONESIA  
MENGUNAKAN MODEL INDOBERT DARI INDONLU: STUDI KASUS  
PADA CUSTOMER PROFILE KOMPAS GRAMEDIA**

Leonardo Tyoes Huibu

**ABSTRAK**

Profil pelanggan memberikan keuntungan strategis bagi sebuah toko dalam merancang pendekatan yang tepat guna meningkatkan profit. Salah satu elemen penting dalam profil pelanggan adalah distribusi jenis kelamin. Memahami distribusi ini dapat memberikan wawasan yang bernilai dalam menentukan jenis promosi serta produk yang perlu diutamakan. Kompas Gramedia mengidentifikasi kebutuhan ini dalam tim pemasaran mereka, namun data jenis kelamin yang tersedia masih sangat terbatas. Pada tahap awal, model *Random Forest* telah dikembangkan untuk klasifikasi jenis kelamin berdasarkan nama-nama Indonesia, namun akurasi dan *F1 score* yang dicapai hanya berkisar antara 70%–72%, yang dianggap belum memadai. Maka dengan alasan tersebut, diperlukan model yang bisa mengatasi permasalahan yang ada. Dikembangkanlah model baru berbasis *Deep Learning* bernama *IndoBERT*. *IndoBERT* adalah model berbasis arsitektur *Transformer* pada bagian *Encoder*, yang terbukti efektif dalam berbagai tugas klasifikasi, terutama dalam konteks Bahasa Indonesia. Dengan menggunakan *IndoBERT*, akurasi dan *F1 score* dalam klasifikasi jenis kelamin meningkat secara signifikan hingga mencapai 94%.

**Kata kunci:** *IndoBERT*, Jenis Kelamin, Klasifikasi Teks, Kompas Gramedia, Nama Indonesia



***GENDER CLASSIFICATION BASED ON INDONESIAN NAMES USING  
THE INDOBERT MODEL FROM INDONLU: A CASE STUDY ON KOMPAS  
GRAMEDIA'S CUSTOMER PROFILE***

Leonardo Tyoes Huibu

***ABSTRACT***

*Customer profiling provides a strategic advantage for a store in designing the right approach to increase profit. One important element in customer profiling is gender distribution. Understanding this distribution can offer valuable insights in determining the types of promotions and products that should be prioritized. Kompas Gramedia identified this need within their marketing team; however, the available gender data was still very limited. In the initial phase, a Random Forest model was developed to classify gender based on Indonesian names, but the achieved accuracy and F1 score only ranged between 70%–72%, which was considered insufficient. For this reason, a more capable model was needed to address the issue. A new deep learning-based model called IndoBERT was then developed. IndoBERT is a model based on a transformer architecture at the encoder level, which has proven effective in various classification tasks, especially in the context of the Indonesian language. By using IndoBERT, the accuracy and F1 score in gender prediction significantly improved, reaching up to 94%.*

***Keywords:*** Gender Prediction, IndoBERT, Indonesian Names, Kompas Gramedia, Text Classification



## DAFTAR ISI

|                                                           |      |
|-----------------------------------------------------------|------|
| HALAMAN JUDUL . . . . .                                   | i    |
| HALAMAN PERNYATAAN ORISINALITAS . . . . .                 | ii   |
| HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH . . . . .      | iii  |
| HALAMAN PERSEMBAHAN/MOTO . . . . .                        | iv   |
| KATA PENGANTAR . . . . .                                  | v    |
| ABSTRAK . . . . .                                         | vii  |
| ABSTRACT . . . . .                                        | viii |
| DAFTAR ISI . . . . .                                      | ix   |
| DAFTAR TABEL . . . . .                                    | x    |
| DAFTAR GAMBAR . . . . .                                   | xi   |
| DAFTAR KODE . . . . .                                     | xii  |
| DAFTAR RUMUS . . . . .                                    | xiii |
| DAFTAR LAMPIRAN . . . . .                                 | xiv  |
| BAB 1 PENDAHULUAN . . . . .                               | 1    |
| 1.1 Latar Belakang Masalah . . . . .                      | 1    |
| 1.2 Maksud dan Tujuan Kerja Magang . . . . .              | 5    |
| 1.3 Waktu dan Prosedur Pelaksanaan Kerja Magang . . . . . | 5    |
| BAB 2 GAMBARAN UMUM PERUSAHAAN . . . . .                  | 7    |
| 2.1 Sejarah Singkat Perusahaan . . . . .                  | 7    |
| 2.2 Visi dan Misi Perusahaan . . . . .                    | 11   |
| 2.3 Struktur Organisasi Perusahaan . . . . .              | 12   |
| BAB 3 PELAKSANAAN KERJA MAGANG . . . . .                  | 15   |
| 3.1 Kedudukan dan Koordinasi . . . . .                    | 15   |
| 3.2 Tugas yang Dilakukan . . . . .                        | 16   |
| 3.3 Uraian Pelaksanaan Magang . . . . .                   | 17   |
| 3.3.1 pengumpulan data . . . . .                          | 18   |
| 3.3.2 Pra-pemrosesan . . . . .                            | 21   |
| 3.3.3 Fungsi dan Utilitas Evaluasi . . . . .              | 36   |
| 3.3.4 Fase pelatihan dan validasi . . . . .               | 43   |
| 3.3.5 Fase testing dan penyimpanan mode . . . . .         | 47   |
| 3.3.6 Evaluasi Model . . . . .                            | 48   |
| 3.3.7 Penerapan server . . . . .                          | 52   |
| 3.3.8 Pelaksanaan Magang . . . . .                        | 61   |
| 3.4 Kendala dan Solusi yang Ditemukan . . . . .           | 62   |
| BAB 4 SIMPULAN DAN SARAN . . . . .                        | 67   |
| 4.1 Simpulan . . . . .                                    | 67   |
| 4.2 Saran . . . . .                                       | 67   |
| DAFTAR PUSTAKA . . . . .                                  | 69   |

## DAFTAR TABEL

|           |                                                                                                     |    |
|-----------|-----------------------------------------------------------------------------------------------------|----|
| Tabel 1.1 | Informasi Magang di Kompas Gramedia . . . . .                                                       | 6  |
| Tabel 3.1 | Confusion Matrix dengan Penandaan Klasifikasi (Kelas 1:<br>Laki-laki, Kelas 0: Perempuan) . . . . . | 41 |
| Tabel 3.2 | Classification Report buat dataset training . . . . .                                               | 49 |
| Tabel 3.3 | Classification Report buat dataset validasi . . . . .                                               | 51 |
| Tabel 3.4 | Classification Report buat dataset pengujian . . . . .                                              | 52 |
| Tabel 3.5 | Pekerjaan yang dilakukan tiap minggu selama<br>pelaksanaan kerja magang . . . . .                   | 61 |



## DAFTAR GAMBAR

|             |                                                                                                                                                                             |    |
|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Gambar 1.1  | Divisi Kompas Gramedia, Grup Retail & Publishing . . . .                                                                                                                    | 1  |
| Gambar 1.2  | The 5 vs of <i>Big Data</i> . . . . .                                                                                                                                       | 2  |
| Gambar 1.3  | Hierarki metode dalam kecerdasan buatan: Artificial Intelligence → Machine Learning → Pembelajaran Fitur (Representation Learning) → Deep Learning → Generative AI. . . . . | 3  |
| Gambar 2.1  | Logo Kompas Gramedia . . . . .                                                                                                                                              | 7  |
| Gambar 2.2  | Logo KG Media . . . . .                                                                                                                                                     | 8  |
| Gambar 2.3  | Logo KG Hospitality . . . . .                                                                                                                                               | 9  |
| Gambar 2.4  | Logo KG Education . . . . .                                                                                                                                                 | 9  |
| Gambar 2.5  | Logo KG Retail & Publishing . . . . .                                                                                                                                       | 9  |
| Gambar 2.6  | Logo KG Manufacture . . . . .                                                                                                                                               | 10 |
| Gambar 2.7  | Logo KG Event & Venue . . . . .                                                                                                                                             | 10 |
| Gambar 2.8  | Logo KG Property . . . . .                                                                                                                                                  | 10 |
| Gambar 2.9  | Logo KG Digital . . . . .                                                                                                                                                   | 11 |
| Gambar 2.10 | Struktur dari Grup Retail dan Publishing . . . . .                                                                                                                          | 12 |
| Gambar 2.11 | Struktur Finance dan Management Directorate . . . . .                                                                                                                       | 13 |
| Gambar 3.1  | Bagan proses pembuatan model IndoBERT prediksi kelamin dari nama Indonesia . . . . .                                                                                        | 18 |
| Gambar 3.2  | Bagan Kerja untuk pengumpulan data . . . . .                                                                                                                                | 19 |
| Gambar 3.3  | Sebagian dataset untuk prediksi <i>gender</i> berdasarkan nama . . . . .                                                                                                    | 19 |
| Gambar 3.4  | Bagan Kerja untuk Pra-pemrosesan . . . . .                                                                                                                                  | 21 |
| Gambar 3.5  | Total nilai yang hilang dalam dataset, baik pada kolom gender, nama, atau keduanya . . . . .                                                                                | 21 |
| Gambar 3.6  | Total karakter khusus dalam nama pelanggan . . . . .                                                                                                                        | 21 |
| Gambar 3.7  | Tabel dataset yang sudah bersih . . . . .                                                                                                                                   | 28 |
| Gambar 3.8  | pie chart distribusi kelamin pada dataset yang sudah bersih . . . . .                                                                                                       | 28 |
| Gambar 3.9  | Bagan kerja untuk fungsi dan utilitas evaluasi . . . . .                                                                                                                    | 36 |
| Gambar 3.10 | Bagan kerja magang fase pelatihan dan validasi . . . . .                                                                                                                    | 43 |
| Gambar 3.11 | Bagan kerja magang fase testing . . . . .                                                                                                                                   | 47 |
| Gambar 3.12 | Beberapa versi model yang telah dibuatkan . . . . .                                                                                                                         | 48 |
| Gambar 3.13 | Beberapa versi model yang telah dibuatkan . . . . .                                                                                                                         | 49 |
| Gambar 3.14 | Hasil confusion matrix untuk Dataset pelatihan . . . . .                                                                                                                    | 49 |
| Gambar 3.15 | Hasil confusion matrix untuk Dataset Validasi . . . . .                                                                                                                     | 50 |
| Gambar 3.16 | Hasil confusion matrix untuk Dataset Pengujian . . . . .                                                                                                                    | 51 |
| Gambar 3.17 | Bagan kerja untuk penerapan server . . . . .                                                                                                                                | 53 |
| Gambar 3.18 | File sumber Python yang berisi prediksi gender terjadwal. . . . .                                                                                                           | 53 |
| Gambar 3.19 | Keluaran di prompt saat memulai dari command prompt/anaconda prompt . . . . .                                                                                               | 60 |
| Gambar 3.20 | Keluaran di file log saat program dimulai . . . . .                                                                                                                         | 60 |
| Gambar 3.21 | Perbedaan data dari dataset lama dengan dataset baru adalah sebanyak 596 entri . . . . .                                                                                    | 63 |
| Gambar 3.22 | Pada folder klasifikasi gender terdapat proyek yang telah dikerjakan serta dataset. . . . .                                                                                 | 65 |
| Gambar 3.23 | Laporan pengerjaan IndoBERT yang telah dibuat sebelumnya. . . . .                                                                                                           | 66 |

## DAFTAR KODE

|           |                                                                                                                     |    |
|-----------|---------------------------------------------------------------------------------------------------------------------|----|
| Kode 3.1  | Potongan kode SQL untuk pengambilan data . . . . .                                                                  | 20 |
| Kode 3.2  | Potongan kode untuk membersihkan dataset KGGenderDataset .                                                          | 22 |
| Kode 3.3  | Potongan kode untuk library yang diperlukan . . . . .                                                               | 24 |
| Kode 3.4  | Potongan kode untuk mendefinisikan Dataset . . . . .                                                                | 25 |
| Kode 3.5  | Potongan kode untuk menghilangkan data yang NULL . . . . .                                                          | 25 |
| Kode 3.6  | Potongan kode untuk membersihkan data yang memiliki karakter khusus . . . . .                                       | 26 |
| Kode 3.7  | Potongan kode untuk mengubah semua huruf jadi huruf kecil . .                                                       | 27 |
| Kode 3.8  | Potongan kode untuk menyimpan dataset yang sudah diproses . .                                                       | 27 |
| Kode 3.9  | Potongan kode untuk label M dan F menjadi class 1 dan 0 . . . .                                                     | 28 |
| Kode 3.10 | Kode ini mengubah kolom data menjadi list Python dan menyimpannya ke dalam variabel baru bernama texts dan labels.  | 30 |
| Kode 3.11 | Potongan kode untuk membagikan data menjadi data pelatihan .                                                        | 30 |
| Kode 3.12 | Potongan kode untuk mengubah kalimat jadi token dan label Laki-laki dan perempuan menjadi '1' dan '0' . . . . .     | 31 |
| Kode 3.13 | kode def __init__ untuk class GenderDataset . . . . .                                                               | 32 |
| Kode 3.14 | kode def __len__ untuk class GenderDataset . . . . .                                                                | 33 |
| Kode 3.15 | kode def __len__ untuk class GenderDataset . . . . .                                                                | 33 |
| Kode 3.16 | Kode yang membantu pelabelan untuk dataset pelatihan . . . . .                                                      | 35 |
| Kode 3.17 | Kode untuk Evaluasi model IndoBERT . . . . .                                                                        | 36 |
| Kode 3.18 | Kode untuk persiapan evaluasi untuk testing dataset . . . . .                                                       | 37 |
| Kode 3.19 | Kode untuk persiapan evaluasi untuk testing dataset . . . . .                                                       | 38 |
| Kode 3.20 | Kode ini untuk buat report klasifikasi . . . . .                                                                    | 39 |
| Kode 3.21 | Kode dibuat untuk fase training untuk model . . . . .                                                               | 43 |
| Kode 3.22 | Argument untuk memberikan kepada model untuk training . . .                                                         | 44 |
| Kode 3.23 | Kode untuk menginisiasi Trainer . . . . .                                                                           | 45 |
| Kode 3.24 | Kode untuk menjalankan evaluasi untuk training . . . . .                                                            | 46 |
| Kode 3.25 | Kode untuk menjalankan evaluasi untuk training . . . . .                                                            | 47 |
| Kode 3.26 | Library yang perlu di import untuk file Server . . . . .                                                            | 53 |
| Kode 3.27 | Import model dan tokenizer yang sudah dibuat . . . . .                                                              | 54 |
| Kode 3.28 | Di program python menggunakan library pyodbc sql diatas dipakai untuk ambil data langsung dari sql server . . . . . | 55 |
| Kode 3.29 | Preprocessing data yang didapatkan dari SQL server . . . . .                                                        | 55 |
| Kode 3.30 | Pembuatan fungsi prediksi di dalam program prediksi kelamin . .                                                     | 56 |
| Kode 3.31 | pengiriman data kembali ke SQL server dengan mengirim ke store procedure . . . . .                                  | 57 |
| Kode 3.32 | kode yang buat program tersebut berjalan otomatis pada jam 22.00 WIB . . . . .                                      | 57 |
| Kode 3.33 | Pembuatan sistem pencatat untuk menyimpan status program prediksi kelamin . . . . .                                 | 58 |

## DAFTAR RUMUS

|           |                                                |    |
|-----------|------------------------------------------------|----|
| Rumus 3.1 | <i>Akurasi</i> . . . . .                       | 39 |
| Rumus 3.2 | <i>Presisi untuk kelas tertentu</i> . . . . .  | 39 |
| Rumus 3.3 | <i>Recall untuk kelas tertentu</i> . . . . .   | 40 |
| Rumus 3.4 | <i>F1 Score untuk kelas tertentu</i> . . . . . | 40 |
| Rumus 3.5 | <i>Macro Average</i> . . . . .                 | 40 |
| Rumus 3.6 | <i>Weighted Average</i> . . . . .              | 41 |



## DAFTAR LAMPIRAN

|            |                                                                                     |    |
|------------|-------------------------------------------------------------------------------------|----|
| Lampiran 1 | MBKM-01 Cover Letter MBKM Internship Track 1 . . . . .                              | 73 |
| Lampiran 2 | MBKM-02 MBKM Internship Track 1 Card . . . . .                                      | 74 |
| Lampiran 3 | MBKM-03 Daily Task - Internship Track 1 . . . . .                                   | 75 |
| Lampiran 4 | MBKM-04 Verification Form of Internship Report MBKM<br>Internship Track 1 . . . . . | 83 |
| Lampiran 5 | Form Bimbingan . . . . .                                                            | 84 |
| Lampiran 6 | Hasil Turnitin . . . . .                                                            | 85 |

