



**OPTIMALISASI KONFIGURASI PIPELINE GUNA MENINGKATKAN
AKURASI INTENT DAN ENTITY PADA KONDISI DATA YANG TIDAK
SEIMBANG DI CHATBOT REKOMENDASI MOBIL BERBASIS RASA
FRAMEWORK**

**TUGAS AKHIR
SKRIPSI**

**MOCHAMMAD IRFAN KURNIAWAN
41822010028**

**UNIVERSITAS
MERCU BUANA**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2026**



**OPTIMALISASI KONFIGURASI PIPELINE GUNA MENINGKATKAN
AKURASI INTENT DAN ENTITY PADA KONDISI DATA YANG TIDAK
SEIMBANG DI CHATBOT REKOMENDASI MOBIL BERBASIS RASA
FRAMEWORK**

**TUGAS AKHIR
SKRIPSI**

**MOCHAMMAD IRFAN KURNIAWAN
41822010028**

MERCU BUANA

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2026**

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Mochammad Irfan Kurniawan

NIM : 41822010028

Program Studi : Sistem Informasi

Menyatakan dengan sebenar-benarnya bahwa Tugas Akhir berjudul:

“ Optimalisasi Konfigurasi Pipeline Guna Meningkatkan Akurasi Intent Dan Entity Pada Kondisi Data Yang Tidak Seimbang Di Chatbot Rekomendasi Mobil Berbasis Rasa Framework”

adalah hasil karya saya sendiri, tidak mengandung unsur plagiarisme, pelanggaran hak cipta, atau konten ilegal dalam bentuk apapun dan tidak melanggar hukum atau hak pihak manapun.

Apabila di kemudian hari ditemukan pelanggaran terhadap pernyataan ini, saya bersedia menanggung seluruh konsekuensi hukum dan membebaskan Universitas Mercu Buana dari segala bentuk tuntutan hukum dan saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Demikian pernyataan ini saya buat dengan sebenar-benarnya untuk digunakan sebagaimana mestinya.

Jakarta, 15 Agustus 2026



Mochammad Irfan Kurniawan

PERNYATAAN SIMILARITY CHECK
/SIMILARITY CHECK STATEMENT

Saya yang bertanda tangan dibawah ini menyatakan, bahwa karya ilmiah yang ditulis oleh
(The undersigned, hereby declare that the scientific work written by

Nama /Name : Mochammad Irfan Kurniawan
NIM /Student ID Number : 41822010028
Program Studi /Study Program : Sistem Informasi

Dengan Judul Tugas Akhir

(The title:

**“OPTIMALISASI KONFIGURASI PIPELINE GUNA MENINGKATKAN AKURASI
INTENT DAN ENTITY PADA KONDISI DATA YANG TIDAK SEIMBANG DI
CHATBOT REKOMENDASI MOBIL BERBASIS RASA FRAMEWORK”**

telah dilakukan pengecekan *similarity* dengan sistem Turnitin pada tanggal:

(Similarity checks have been carried out with the Turnitin system on the date:

30 Juli 2026

dengan nilai persentase sebesar :

(with a percentage value of:

5%

dinyatakan memenuhi standar sesuai dengan ketentuan berlaku di **Fakultas Ilmu Komputer**
Universitas Mercu Buana. *(declared to meet standards in accordance with applicable*
regulations at the Faculty of Computer Science, Universitas Mercu Buana.

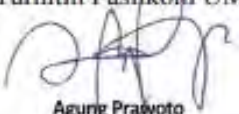
File hasil cek similarity turnitin:

(Turnitin similarity report file

https://drive.google.com/file/d/1mmLAWceF2vQ4BSSbLnCTy8sJB118rIZz/view?usp=drive_link



Jakarta, 30 Juli 2026
Admin Turnitin Fasilkom UMB



Agung Prawoto

Agung Prawoto, S.Kom., B.Sc
NIK : 322970503

Fakultas Ilmu Komputer

KAMPUS MENARA BHAKTI

Jl. Raya Meruya Selatan No. 1 Kembangan, Jakarta Barat 11650

Telp. 021-5840816 (Hunting), Psw : 5700 Fax. 021-5840813

<http://www.mercubuana.ac.id>, e-mail : fasilkom@mercubuana.ac.id

HALAMAN PENGESAHAN

Laporan Skripsi ini diajukan oleh:

Nama : Mochammad Irfan Kurniawan
NIM : 41822010028
Program Studi : Sistem Informasi
Judul Laporan Skripsi : Optimalisasi Konfigurasi Pipeline Guna
Meningkatkan Akurasi Intent Dan Entity Pada
Kondisi Data Yang Tidak Seimbang Di Chatbot
Rekomendasi Mobil Berbasis Rasa Framework

Telah berhasil dipertahankan pada sidang di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Strata 1 pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer Universitas Mercu Buana.

Disahkan oleh:

Pembimbing,



Riri Fajriah, S.Kom., MM., M.Kom
NIDN/NUPTK: 0321108502

Jakarta, 15 Agustus 2026

Mengetahui,

Dekan Fakultas Ilmu Komputer



Dr. Bambang Jokonowo, S.Si., M.T.I.
NIDN: 0320037002

Ketua Program Studi
Sistem Informasi



Wawan Gunawan, S.Kom., M.T., M.Kom.
NIDN: 0424108104

KATA PENGANTAR

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, saya dapat menyelesaikan Laporan Skripsi ini. Penulisan Laporan Skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer pada Fakultas Ilmu Komputer Universitas Mercu Buana. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan Laporan Skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Ir. Andi Adriansyah, M Eng., selaku Rektor Universitas Mercu Buana.
2. Bapak Dr. Bambang Jokonowo, S.Si., MTL., selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Wawan Gunawan, S.Kom., MT., M.Kom., selaku Ketua Program Studi Sistem Informasi.
4. Ibu Riri Fajriah, S.Kom., MM., M.Kom., selaku Dosen Pembimbing yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan saya dalam penyusunan skripsi ini.
5. Teruntuk Orang Tua dan Keluarga Besar Saya. Terima kasih atas doa, dukungan moral, materi, dan pengorbanan tanpa henti. Gelar sarjana ini adalah hasil dari doa kalian.
6. Teruntuk Johnny Silverhand. Terima kasih karena telah mengajarkan penulis untuk selalu berpikir kritis.

Akhir kata, saya berharap Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga Laporan Skripsi ini membawa manfaat bagi pengembangan ilmu.

Jakarta, 22 Juli 2026



Mochammad Irfan Kurniawan

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR DI REPOSITORI UMB**

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : Mochammad Irfan Kurniawan
NIM : 41822010028
Fakultas/Program Studi : Ilmu Komputer/Sistem Informasi
Judul Tugas Akhir : Optimalisasi Konfigurasi Pipeline Guna
Meningkatkan Akurasi Intent Dan Entity Pada
Kondisi Data Yang Tidak Seimbang Di Chatbot
Rekomendasi Mobil Berbasis Rasa Framework

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 15 Agustus 2026

Yang menyatakan,



Mochammad Irfan Kurniawan

ABSTRAK

Pertumbuhan pesat industri otomotif Indonesia mendorong kebutuhan akan chatbot rekomendasi mobil yang mampu memahami maksud pengguna secara akurat. Komponen inti dari chatbot tersebut, yaitu Natural Language Understanding (NLU), sering menghadapi tantangan berupa extreme class imbalance, di mana kelas intent dan entity minoritas kurang terwakili dalam data latih sehingga rentan menghasilkan kesalahan interpretasi preferensi pengguna. Penelitian ini bertujuan mengevaluasi secara komparatif tujuh konfigurasi pipeline NLU berbasis framework Rasa dengan DIETClassifier, terdiri atas satu konfigurasi baseline (Default) dan enam konfigurasi eksperimental (TAv1–TAv6), guna mengidentifikasi arsitektur yang paling optimal dalam mengatasi dampak ketidakseimbangan data pada domain rekomendasi otomotif berbahasa Indonesia. Dataset yang digunakan berjumlah 903 kalimat berbahasa Indonesia yang tersebar dalam enam intent dengan rasio ketimpangan antar kelas mencapai 1:16. Evaluasi dilakukan menggunakan 5-fold cross-validation dan diukur berdasarkan empat kriteria (KPI): performa Macro F1-Score agregat, performa pada kelas minoritas, stabilitas model, dan kemampuan generalisasi. Hasil penelitian menunjukkan bahwa konfigurasi TAv6, yang mengintegrasikan LanguageModelFeaturizer berbasis IndoBERT, fitur leksikal, dan RegexEntityExtractor secara sinergis, merupakan satu-satunya konfigurasi yang memenuhi seluruh kriteria KPI (6/6). TAv6 mencatatkan peningkatan signifikan pada F1-Score intent minoritas `ask_similar_car` dan entitas `feature.negated` dibandingkan baseline, dengan ukuran efek Cohen's d sebesar +1,1112 (sangat besar) serta stabilitas standar deviasi yang minimal. Temuan ini membuktikan bahwa integrasi representasi semantik kontekstual dan fitur leksikal-aturan secara bersamaan mampu menghasilkan model yang akurat, stabil, dan tahan terhadap ketidakseimbangan data ekstrem.

Kata Kunci : NLU, data imbalance, IndoBERT, Optimasi, DIETClassifier

ABSTRACT

The rapid growth of Indonesia's automotive industry has increased the need for car recommendation chatbots capable of accurately understanding user intent. The core component of such chatbots, Natural Language Understanding (NLU), often faces extreme class imbalance, where minority intent and entity classes are underrepresented in training data, leading to misinterpretation of user preferences. This study comparatively evaluates seven Rasa-based NLU pipeline configurations using DIETClassifier, consisting of one baseline (Default) and six experimental configurations (TAv1–TAv6), to identify the most optimal architecture for mitigating data imbalance in an Indonesian-language automotive recommendation domain. The dataset comprises 903 Indonesian sentences across six intents with a class imbalance ratio reaching 1:16. Evaluation was conducted using 5-fold cross-validation and assessed against four key performance indicators (KPIs): aggregate Macro F1-Score, minority class performance, model stability, and generalization capability. Results show that the TAv6 configuration, which synergistically integrates an IndoBERT-based LanguageModelFeaturizer, lexical features, and RegexEntityExtractor, is the only configuration meeting all KPI criteria (6/6). TAv6 achieved significant F1-Score improvements on the minority intent ask_similar_car and the feature.negated entity compared to the baseline, with a Cohen's d effect size of +1.1112 (very large) and minimal standard deviation. These findings demonstrate that integrating contextual semantic representation with lexical and rule-based features simultaneously produces a model that is accurate, stable, and resilient to extreme data imbalance.

Keywords: NLU, data imbalance, IndoBERT, optimization, DIETClassifier

U N I V E R S I T A S
M E R C U B U A N A

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERNYATAAN KARYA SENDIRI.....	ii
HALAMAN SURAT KETERANGAN UJI TURNITIN.....	iii
HALAMAN PENGESAHAN	iv
KATA PENGANTAR	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR DI REPOSITORI UMB.....	vi
ABSTRAK.....	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 Teori Utama.....	6
2.2 Teori Pendukung.....	10
2.3 Penelitian Terdahulu	18
BAB III METODE PENELITIAN.....	42
3.1 Jenis Penelitian	42
3.2 Subjek Penelitian.....	42
3.3 Instrumen Penelitian.....	50
3.4 Tahapan Penelitian.....	52
BAB IV PEMBAHASAN	60
4.1 Pre-Processing Data	60
4.2 Split Data.....	64
4.3 Proses Training	65
4.4 Evaluasi.....	71
BAB V KESIMPULAN DAN SARAN.....	99
5.1 Kesimpulan	99
5.2 Saran.....	101
DAFTAR PUSTAKA.....	102
LAMPIRAN	105

DAFTAR TABEL

Tabel 2. 1 Penelitian terdahulu	18
Tabel 3. 1 Distribusi Kalimat Intent	43
Tabel 3. 2 Contoh Intent start_search	44
Tabel 3. 3 Contoh Intent out_of_scope	45
Tabel 3. 4 Contoh Intent choose_preference.....	45
Tabel 3. 5 Contoh Intent ask_recommendation.....	46
Tabel 3. 6 Contoh Intent ask_comparison.....	47
Tabel 3. 7 Contoh Intent ask_similiar_car	48
Tabel 3. 8 Jumlah Lookup per entity	49
Tabel 3. 9 Spesifikasi Hardware.....	51
Tabel 3. 10 Daftar Software dan Library	51
Tabel 3. 11 Skrip Yang Digunakan	52
Tabel 4. 1 Distribusi Intent	61
Tabel 4. 2 Distribusi Entity	62
Tabel 4. 3 Konfigurasi Default	66
Tabel 4. 4 Konfigurasi TAv1	66
Tabel 4. 5 Konfigurasi TAv2	67
Tabel 4. 6 Konfigurasi TAv3	68
Tabel 4. 7 Konfigurasi TAv4	68
Tabel 4. 8 Konfigurasi TAv5	69
Tabel 4. 9 Konfigurasi TAv6	70
Tabel 4. 10 F1-Score Intent ask_similiar_car	74
Tabel 4. 11 F1-Score Entity featured.negated.....	76
Tabel 4. 12 Rata-rata Overall Macro F1-Score Intent dan Entity.....	78
Tabel 4. 13 Hasil Uji Normalitas (Shapiro-Wilk Test).....	82
Tabel 4. 14 Uji Signifikansi Intent	84
Tabel 4. 15 Uji Effect Size Intent.....	85
Tabel 4. 16 Uji Signifikansi Entity	86
Tabel 4. 17 Uji Effect Size Entity	87
Tabel 4. 18 Uji Signifikansi dan Effect Size Pada Kelas Minoritas (Wilcoxon Test)	87
Tabel 4. 19 Hasil Perbandingan Waktu Pengujian	91

DAFTAR GAMBAR

Gambar 3. 1 Tahapan Penelitian	53
Gambar 4. 1 Heatmap Overall Intent.....	72
Gambar 4. 2 Heatmap Overall Entity	73
Gambar 4. 3 F1-Score Intent ask_similar	75
Gambar 4. 4 F1-Score Entity feature.negated	77
Gambar 4. 5 Overall F1-Score Rata-Rata Intent dan Entity	79
Gambar 4. 6 Scorecard KPI	95



DAFTAR LAMPIRAN

Lampiran 1 Kartu Asistensi	105
Lampiran 2 Curriculum Vitae	106
Lampiran 3 Bukti Daftar Bnsp	107
Lampiran 4 Hasil Cek Turnitin	108
Lampiran 5 Function untuk menjalankan eval di terminal secara otomatis	109
Lampiran 6 Function untuk training dan overfit test.....	109

