



**APLIKASI PENERJEMAH BAHASA OGAN BERBASIS
MODEL *NO LANGUAGE LEFT BEHIND* (NLLB)**

SKRIPSI

KADEK SARASTIKA

211410044

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS SAINS TEKNOLOGI

UNIVERSITAS BINA DARMA

2025/2026



**APLIKASI PENERJEMAH BAHASA OGAN BERBASIS
MODEL *NO LANGUAGE LEFT BEHIND* (NLLB)**

KADEK SARASTIKA

211410044

**Skripsi ini diajukan sebagai syarat memperoleh gelar Sarjana
Komputer**

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS SAINS TEKNOLOGI

UNIVERSITAS BINA DARMA

2025/2026

HALAMAN PENGESAHAN

**APLIKASI PENERJEMAH BAHASA OGAN BERBASIS MODEL
NO LANGUAGE LEFT BEHIND (NLLB)**

KADEK SARASTIKA

211410044

**Telah diterima sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer pada Program Studi Sistem Informasi**

Pembimbing

**Palembang, 30 Juli 2025
Fakultas Sains Teknologi
Universitas Bina Darma
Dekan,**



Ilman Zuhri Yadi, S.Kom, M.M, M.Kom. Dr. Tata Sutabri, S.Kom., MMSI., MKM



Fakultas Sains Teknologi

HALAMAN PERSETUJUAN

Skripsi Berjudul "APLIKASI PENERJEMAH BAHASA OGAN BERBASIS MODEL NO LANGUAGE LEFT BEHIND (NLLB)" Oleh "Kadek Sarastika", telah dipertahankan di depan komisi penguji pada hari Rabu tanggal 30 Juli 2025.

Komisi Penguji

1. Ketua : Ilman Zuhri Yadi, S.Kom, M.M, M.Kom.



(.....)

2. Anggota : Susan Dian Purnamasari, M.Kom




(.....)

3. Anggota : Edi Supratman, M.Kom



(.....)

Mengetahui,
Program Studi Sistem Informasi
Fakultas Sains Teknologi
Universitas Bina Darma
Ketua,



Dr. Ari Muzakir, S.Kom., M.Cs

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini :

Nama : Kadek Sarastika

NIM : 211410044

Dengan ini menyatakan bahwa :

1. Karya tulis saya (Skripsi) adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (Sarjana) di Universitas Bina Darma atau perguruan tinggi lainnya ;
2. Karya tulis ini murni gagasan, rumusan dan penelitian saya dengan arahan dari tim pembimbing ;
3. Di dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau di publikasikan orang lain, kecuali secara tertulis dengan jelas dikutip dengan mencantumkan nama pengarang dan memasukkan ke dalam daftar rujukan ;
4. Saya bersedia tugas skripsi, di cek keasliannya menggunakan plagiarism checker serta di unggah ke internet, sehingga dapat diakses secara daring ;
5. Surat pernyataan ini saya tulis dengan sungguh-sungguh dan apabila terbukti melakukan penyimpangan atau ketidakbenaran dalam pernyataan ini maka saya bersedia menerima sanksi dengan peraturan dan perundang-undangan yang berlaku ;

Demikian surat pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Palembang, Juli 2025

Yang membuat pernyataan,



Kadek Sarastika

211410044

MOTTO DAN PERSEMBAHAN

MOTTO

“Kesalahan terbesar seseorang adalah menganggap bahwa masih ada hari esok”

PERSEMBAHAN

Skripsi ini saya persembahkan untuk:

1. Kedua orang tua tercinta yang selalu menjadi sumber kekuatan dan semangat dalam hidup penulis, atas doa, cinta kasih, dan dukungan moril maupun materiil yang tak pernah berhenti.
2. Ibu Dr. Yesi Novaria Kunang, S.T., M.Kom. selaku Ketua Riset Intelligent System Research Group yang telah memberikan arahan dalam penentuan topik dan model pengembangan aplikasi.
3. Bapak Ilman Zuhri Yadi, S.Kom., M.M., M.Kom. selaku dosen pembimbing karya akhir yang telah membimbing penulis dengan sabar dan penuh perhatian selama proses penyusunan laporan ini.
4. Kakak laki-laki penulis, Wayan Agus Budianto, S.Psi yang selalu menjadi panutan dalam bekerja keras dan memberi motivasi di setiap langkah perjalanan studi ini.
5. Adik perempuan penulis, Komang Dinda Jenitalia yang senantiasa memberi semangat, perhatian, dan menjadi tempat berbagi di kala lelah dan butuh dukungan moral.

Abstract

Machine translation technology is an essential part of natural language processing that aims to facilitate cross-lingual communication. Oganese, as one of the regional languages in South Sumatra, has not yet received adequate technological support for translation. This research aims to apply the No Language Left Behind (NLLB) model for translating Indonesian into Oganese and vice versa, as well as to develop a bidirectional web-based translation application using the Flask framework. A total of 5,000 parallel Indonesian–Oganese sentence pairs were collected for this study. The NLLB model was fine-tuned on the dataset with tokenizer adjustments to recognize the Oganese language (ogan_Latn). Evaluation using BLEU and ChrF++ metrics achieved scores of BLEU 93.24 and ChrF++ 96.50 for Oganese-to-Indonesian translation, and BLEU 86.49 and ChrF++ 93.69 for the reverse direction. These results demonstrate that the model is capable of accurate bidirectional translation. The web application was successfully implemented locally and provides real-time translation with an interactive interface. This study proves that the NLLB approach can be adapted to support the preservation of low-resource regional languages.

Keywords: *Oganese, Machine Translation, No Language Left Behind (NLLB), BLEU Score, ChrF++, Regional Language.*

ABSTRAK

Teknologi penerjemahan mesin merupakan bagian penting dari pemrosesan bahasa alami yang bertujuan untuk mempermudah komunikasi antarbahasa. Bahasa Ogan, sebagai salah satu bahasa daerah di Sumatera Selatan, hingga saat ini belum memiliki dukungan teknologi penerjemahan yang memadai. Penelitian ini bertujuan untuk menerapkan model *No Language Left Behind* (NLLB) dalam menerjemahkan bahasa Indonesia ke bahasa Ogan dan sebaliknya, serta mengembangkan sebuah aplikasi penerjemah dua arah berbasis *web* menggunakan *framework Flask*. Dalam proses pengembangannya, dikumpulkan sebanyak 5.000 pasangan data paralel bahasa Indonesia–Ogan. Model NLLB di-*fine-tune* menggunakan data tersebut, dengan penyesuaian tokenizer agar dapat mengenali bahasa Ogan (ogan_Latn). Evaluasi dilakukan dengan metrik *BLEU* dan *ChrF++*, yang menunjukkan skor: *BLEU* 93.24 dan *ChrF++* 96.50 untuk arah Ogan ke Indonesia, serta *BLEU* 86.49 dan *ChrF++* 93.69 untuk arah sebaliknya. Hasil ini menunjukkan bahwa model mampu melakukan penerjemahan dua arah secara akurat. Aplikasi web berhasil diimplementasikan secara lokal dan dapat menampilkan hasil terjemahan secara *real-time* dengan antarmuka yang interaktif. Penelitian ini membuktikan bahwa pendekatan NLLB dapat diadaptasi untuk mendukung pelestarian bahasa daerah dengan sumber daya terbatas.

Kata kunci: Bahasa Ogan, Penerjemahan Mesin, *No Language Left Behind* (NLLB), *BLEU* Score, *ChrF++*, Bahasa Daerah.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa, karena atas rahmat, kasih, dan karunia-Nya, laporan penelitian ini dapat diselesaikan dengan baik. Laporan ini disusun sebagai salah satu syarat untuk menyelesaikan pendidikan Strata Satu (S1) pada Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Bina Darma.

Penyusunan laporan ini tentu bukanlah proses yang mudah. Di balik pencapaiannya, terdapat banyak tantangan dan keterbatasan yang penulis hadapi. Namun dengan semangat, kerja keras, dan doa yang terus mengiringi, akhirnya laporan ini berhasil diselesaikan. Penulis menyadari bahwa laporan ini masih jauh dari kata sempurna, oleh karena itu saran dan kritik yang membangun sangat penulis harapkan demi perbaikan di masa mendatang.

Ucapan terima kasih yang sebesar-besarnya penulis sampaikan kepada:

1. Ibu Prof. Sunda Ariana, M.Pd., M.M. selaku Rektor Universitas Bina Darma.
2. Bapak Dr. Tata Sutabri, S.Kom., MMSI., M.Kom., Ph.D. selaku Dekan Fakultas Sains dan Teknologi Universitas Bina Darma.
3. Bapak Dr. Ari Muzakir, S.Kom., M.Cs. selaku Kepala Program Studi Sistem Informasi Universitas Bina Darma.
4. Ibu Dr. Yesi Novaria Kunang, S.T., M.Kom. selaku Ketua Riset Intelligent System Research Group yang telah memberikan arahan dalam penentuan topik dan model pengembangan aplikasi.
5. Bapak Ilman Zuhri Yadi, S.Kom., M.M., M.Kom. selaku dosen pembimbing karya akhir yang telah membimbing penulis dengan sabar dan penuh perhatian selama proses penyusunan laporan ini.
6. Kedua orang tua tercinta yang selalu menjadi sumber kekuatan dan semangat dalam hidup penulis, atas doa, cinta kasih, dan dukungan moril maupun materiil yang tak pernah berhenti.

7. Kakak laki-laki penulis, Wayan Agus Budianto, S.Psi yang selalu menjadi panutan dalam bekerja keras dan memberi motivasi di setiap langkah perjalanan studi ini.
8. Adik perempuan penulis, Komang Dinda Jenitalia yang senantiasa memberi semangat, perhatian, dan menjadi tempat berbagi di kala lelah dan butuh dukungan moral.
9. Sahabat-sahabat seperjuangan penulis selama masa kuliah, yang telah banyak membantu dan menemani dalam suka dan duka, khususnya kepada Maryani Farwati, Dimas, Selda, Liza, Eca, Imel, Rama, Mba Marina dan Mba Dewi.

Akhir kata, penulis berharap laporan penelitian ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya dalam bidang teknologi penerjemahan bahasa daerah. Semoga apa yang telah dicapai ini dapat menjadi langkah awal dalam memberikan kontribusi nyata bagi pelestarian budaya bangsa melalui teknologi.

Palembang, 17 Juli 2025

DAFTAR ISI

MOTTO DAN PERSEMBAHAN.....	vi
<i>Abstract</i>	vii
ABSTRAK.....	viii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL	xv
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah	4
1.5 Manfaat Penelitian.....	4
BAB II.....	5
TINJAUAN PUSTAKA.....	5
2.1 Landasan Teori.....	5
2.1.1 Bahasa Ogan.....	5
2.1.2 Mesin Penerjemah	5
2.1.3 <i>No Language Left Behind (NLLB)</i>	5
2.1.4 <i>BLEU Score</i>	9
2.1.5 Penelitian Terdahulu	11
BAB III.....	14
METODOLOGI PENELITIAN	14
3.1 Metodologi Penelitian	14
3.2 Tahapan Pengembangan Aplikasi.....	14
BAB IV.....	19
HASIL DAN PEMBAHASAN.....	19
4.1 Pengumpulan Dataset	19
4.2 Pra Proses Data	21

4.3 Pelatihan Model	24
4.3.1 Eksplorasi Data	25
4.3.2 Evaluasi Kesesuaian Dataset dengan Tokenizer NLLB	27
4.3.3 Menambahkan Tag Bahasa Baru ke dalam Tokenizer Dan Model.	29
4.3.4 Menyiapkan Proses Pelatihan Model	32
4.3.5 Proses Pelatihan	33
4.3.6 Penerapan Model	36
4.4 Evaluasi Bleu Score	38
4.5 Implementasi Ke Dalam Aplikasi Penerjemah	42
4.6 Pengujian Dan Penyempurnaan	47
BAB V	51
KESIMPULAN DAN SARAN	51
1.2 Kesimpulan	51
1.3 Saran	51
DAFTAR PUSTAKA	53

DAFTAR GAMBAR

Gambar 3.1 Tahapan Pengembangan Aplikasi Penerjemah.....	14
Gambar 4.1 Kamus OGAN-INDONESIA.....	20
Gambar 4.2 Dataset Format <i>Excel</i>	21
Gambar 4.3 Dataset Format <i>Excel</i>	21
Gambar 4.4 Dataset Format <i>JSON</i>	22
Gambar 4.5 Proses Pembersihan Dataset.....	23
Gambar 4.6 Jumlah Dataset Setelah Pembersihan.....	23
Gambar 4.7 <i>Mounted at GDrive</i>	25
Gambar 4.8 Kode Membaca File Dataset <i>JSON</i>	26
Gambar 4.9 Kode Membaca File Dataset <i>JSON</i>	26
Gambar 4.10 Proses Pemanggilan <i>Tokenizer</i>	27
Gambar 4.11 Contoh Acak Lima Pasang Dataset.....	28
Gambar 4.12 Proses Impor Pustaka <i>Transformers</i>	29
Gambar 4.13 Pemanggilan Fungsi <i>Fix_Tokenizer</i>	30
Gambar 4.14 Penambahan Custom Token.....	31
Gambar 4.15 Perintah Model <i>Resize_Token_Embeddings</i>	31
Gambar 4.16 Pemindahan Model Dari CPU Ke GPU.....	32
Gambar 4.17 Proses Utama Pelatihan.....	33
Gambar 4.18 Model Berhasil Menyimpan <i>Checkpoint</i>	34
Gambar 4.19 Perintah <i>!pip Install Matplotlib</i>	35
Gambar 4.20 Kurva Penurunan Loss.....	35
Gambar 4.21 <i>Model yang Dipindahkan Ke GPU Dimuat Ulang</i>	37
Gambar 4.22 Pengujian Terhadap Model Penerjemah Dua Arah.....	37
Gambar 4.23 <i>json_to_pandas</i> Dataset Kalimat.....	38
Gambar 4.24 Mengimpor Pustaka <i>SacreBLEU</i>	40
Gambar 4.25 Contoh Kalimat yang digunakan untuk Pengujian <i>BLEU score</i>	40
Gambar 4.26 Proses Pengujian 70 Data Uji.....	41

Gambar 4.27 Performa Model Menggunakan <i>BLEU</i> dan <i>chrF++</i>	42
Gambar 4.28 Model Dipindahkan dari <i>Google Drive</i> ke <i>VSCode</i>	43
Gambar 4.29 Bagian <i>test.py</i>	44
Gambar 4.30 <i>Framework Flask</i>	45
Gambar 4.31 <i>Flask</i> Berhasil Dijalankan.....	46
Gambar 4.32 Hasil Terjemahan Aplikasi <i>Web Ogan</i> ↔ <i>Indonesia</i>	47
Gambar 4.33 Hasil Terjemahan Aplikasi <i>Web Indonesia</i> ↔ <i>Ogan</i>	48
Gambar 4.34 Peringatan Untuk Menghindari Penggunaan Huruf Kapital Secara Keseluruhan.....	49
Gambar 4.35 Peringatan Bahwa Sistem Tidak Mendukung Penerjemahan Angka.....	50



DAFTAR TABEL

Tabel 2.1 Interpretasi <i>Bleu Score</i>	10
Tabel 2. 2 Penelitian Terdahulu.....	13

