

TESIS
ANALISIS KINERJA MODEL MT5 DAN NLLB-200 DENGAN
OPTIMASI OPTUNA UNTUK PENERJEMAHAN BAHASA
ANGKOLA-INDONESIA



disusun oleh
AWAL RIDHO HARAHAHAP
24.55.1580
Konsentrasi : Business Intelligence

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026

TESIS
ANALISIS KINERJA MODEL MT5 DAN NLLB-200 DENGAN
OPTIMASI OPTUNA UNTUK PENERJEMAHAN BAHASA
ANGKOLA-INDONESIA

ANALYSIS OF THE PERFORMANCE OF MT5 AND NLLB-
200 MODELS WITH OPTUNA OPTIMIZATION FOR
ANGKOLA-INDONESIAN TRANSLATION

Diajukan untuk memenuhi salah satu syarat mencapai derajat Pascasarjana
Program Studi PJJ INFORMATIKA



disusun oleh

AWAL RIDHO HARAHAH

24.55.1580

Konsentrasi : Business Intelligence

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2026

HALAMAN PERSETUJUAN

**ANALISIS KINERJA MODEL MT5 DAN NLLB-200 DENGAN OPTIMASI
OPTUNA UNTUK PENERJEMAHAN BAHASA ANGKOLA-INDONESIA**

**ANALYSIS OF THE PERFORMANCE OF MT5 AND NLLB-200 MODELS
WITH OPTUNA OPTIMIZATION FOR ANGKOLA-INDONESIAN
TRANSLATION**

yang disusun dan diajukan oleh

AWAL RIDHO HARAHAQ

24.55.1580

telah disetujui oleh Dosen Pembimbing Tesis
pada tanggal 04 Maret 2026

Dosen Pembimbing,



Hanafi, S.Kom., M.Eng., Ph.D.
NIK. 190302024

HALAMAN PENGESAHAN

**ANALISIS KINERJA MODEL MT5 DAN NLLB-200 DENGAN OPTIMASI
OPTUNA UNTUK PENERJEMAHAN BAHASA ANGKOLA-INDONESIA**

**ANALYSIS OF THE PERFORMANCE OF MT5 AND NLLB-200 MODELS
WITH OPTUNA OPTIMIZATION FOR ANGKOLA-INDONESIAN
TRANSLATION**

yang disusun dan diajukan oleh

AWAL RIDHO HARAHAP

24.55.1580

Telah dipertahankan di depan Dewan Penguji
pada tanggal 04 Maret 2026

Susunan Dewan Penguji

Nama Penguji

Dr. Andi Sunyoto, S.Kom., M.Kom
NIK. 190302052

Robert Marco, S.T., M.T., Ph.D
NIK. 190302228

Hanafi, S.Kom., M.Eng., Ph.D
NIK. 190302024

Tanda Tangan



Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer
Tanggal 04 Maret 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : **AWAL RIDHO HARAHAP**
NIM : **24.55.1580**

Menyatakan bahwa Tesis dengan judul berikut:

ANALISIS KINERJA MODEL MT5 DAN NLLB-200 DENGAN OPTIMASI OPTUNA UNTUK PENERJEMAHAN BAHASA ANGKOLA-INDONESIA

Dosen Pembimbing : **Hanafi, S.Kom., M.Eng., Ph.D**

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas **AMIKOM** Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian **SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas **AMIKOM** Yogyakarta.
5. Pernyataan ini **SAYA** buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 04 Maret 2026

Yang Menyatakan,



AWAL RIDHO HARAHAP

HALAMAN PERSEMBAHAN

Tesis ini saya persembahkan sebagai wujud syukur dan rasa terima kasih kepada:

1. **Tuhan Yang Maha Esa**, atas segala rahmat, kekuatan, dan bimbingan-Nya sehingga penelitian ini dapat diselesaikan dengan baik.
2. **Kedua Orang Tua Tercinta, Ayahanda (Ali Imran Harahap) dan Ibunda (Saudah Hasibuan)**, atas doa yang tidak pernah terputus, kasih sayang, serta dukungan moral yang menjadi motivasi terbesar dalam setiap langkah hidup saya.
3. **Istri Tercinta, (Khalrun Niswa, S.Kom)** yang senantiasa menyertai dengan doa, kesabaran, serta dukungan kasih sayang yang tiada henti dalam setiap tahapan penyelesaian studi ini.
4. **Kedua anak Tersayang, (Muhammad Faeyza Harahap dan Muhammad Kazlo Harahap)** yang menjadi sumber inspirasi, motivasi, dan kekuatan bagi penulis untuk terus memberikan yang terbaik bagi masa depan keluarga.
5. **Keluarga Besar**, atas dukungan, kesabaran, dan semangat yang senantiasa diberikan selama masa studi dan penyusunan tesis ini.
6. **Almater Tercinta, Universitas Amikom Yogyakarta**, khususnya Program Magister PJJ Informatika, sebagai tempat saya menimba ilmu dan mengembangkan diri secara akademik.
7. **Para Dosen Pembimbing dan Penguji**, yang telah memberikan ilmu, arahan, dan keteladanan yang sangat berharga dalam penyelesaian karya ini.
8. **Seluruh Rekan Kerja dan Pihak Terkait**, yang telah memberikan bantuan serta kerja sama baik secara langsung maupun tidak langsung selama proses penelitian berlangsung.

Semoga karya ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya dalam bidang informatika dan pelestarian bahasa daerah di Indonesia.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, sehingga naskah tesis yang berjudul "**Analisis Kinerja Model mT5 dan NLLB-200 dengan Optlmast Optuna untuk Penerjemahan Bahasa Angkola-Indonesia**" dapat diselesaikan dengan baik. Penyusunan karya ilmiah ini merupakan salah satu syarat akademik dalam menyelesaikan program studi Magister PJJ Informatika.

Penulis menyadari bahwa keberhasilan penyelesaian tesis ini tidak lepas dari bimbingan, arahan, serta dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang tulus kepada:

1. Hanafi, S.Kom., M.Eng., Ph.D, selaku Dosen Pembimbing Utama yang telah memberikan arahan, ilmu, serta bimbingan yang sangat berharga selama proses penelitian.
2. Tim Dosen Penguji, yang telah memberikan kritik dan masukan konstruktif dalam penyempurnaan naskah ini.
3. Istri tercinta, atas kesabaran, doa, dan dukungan moril yang senantiasa mengiringi perjalanan akademik penulis.
4. Kedua anak tersayang, yang menjadi motivasi terbesar bagi penulis untuk menyelesaikan studi ini dengan sebaik-baiknya.
5. Segenap rekan-rekan di program studi Magister PJJ Informatika yang telah memberikan semangat dan diskusi yang bermanfaat selama masa perkuliahan.

Penulis menyadari bahwa karya ini masih jauh dari kesempurnaan. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi pengembangan penelitian di masa yang akan datang. Akhir kata, semoga tesis ini dapat memberikan kontribusi nyata bagi pengembangan teknologi informasi, khususnya dalam pelestarian Bahasa Angkola.

Yogyakarta, 04 Maret 2026

Penulis

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN KEASLIAN TESIS.....	v
HALAMAN PERSEMBAHAN.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
LAMPIRAN.....	xii
DAFTAR LAMBANG DAN SINGKATAN.....	xiii
DAFTAR ISTILAH.....	xiv
INTISARI.....	xv
ABSTRACT.....	xvi
BAB 1 PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	6
BAB 2 TINJAUAN PUSTAKA.....	7
2.1 Tinjauan Pustaka.....	7
2.1.1 Penerjemahan Mesin Bahasa Sumber Daya Rendah.....	7
2.1.2 Perkembangan NLP untuk Bahasa Daerah di Indonesia.....	7
2.1.3 Implementasi mT5 dan NLLB.....	8
2.1.4 Optimasi Hyperparameter dalam NLP.....	8
2.2 Keaslian Penelitian.....	10
2.3 Landasan Teori.....	21
2.3.1 Neural Machine Translation (NMT).....	21
2.3.2 Arsitektur Transformer.....	21
2.3.3 mT5 (Multilingual T5).....	22
2.3.4 NLLB-200 (No Language Left Behind).....	22
2.3.5 Optimasi Hyperparameter dengan Optuna.....	22
2.3.6 Metrik Evaluasi.....	23
2.3.7 Bahasa Angkola.....	23
2.3.8 Bahasa Indonesia.....	24
BAB 3 METODE PENELITIAN.....	25

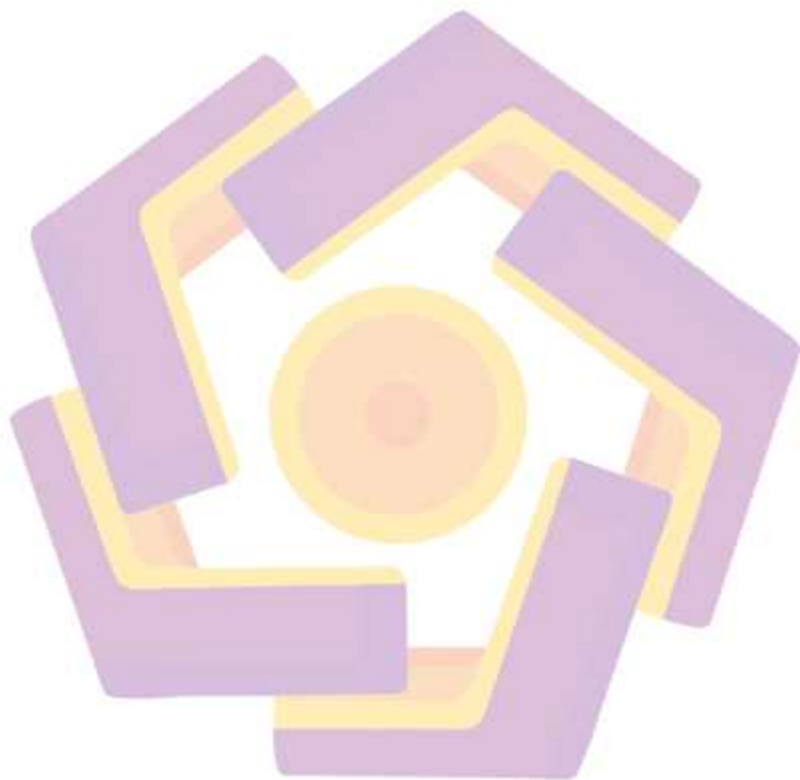
3.1	Alur Penelitian (Research Flow)	25
3.2	Dataset	26
3.3	Prapemrosesan Data	27
3.3.1	Case Folding	27
3.3.2	Cleaning	28
3.3.3	Deduplikasi	28
3.4	Pembagian Data	29
3.5	Tokenisasi	29
3.6	Algoritma dan Arsitektur Model	30
3.6.1	mT5 (Multilingual Text-to-Text Transfer Transformer)	30
3.6.2	NLLB-200 (No Language Left Behind)	31
3.6.3	Optimasi Hyperparameter dengan Optuna	32
3.7	Evaluasi Model	32
3.7.1	BLEU Score	32
3.7.2	chrF Score	34
3.8	Visualisasi Hasil	35
3.9	Skenario Pengujian & Visualisasi	36
BAB 4	HASIL PENELITIAN DAN PEMBAHASAN	37
4.1	Gambaran Umum Implementasi	37
4.2	Hasil Pengujian	37
4.2.1	Hasil Pratinjau Dataset	37
4.2.2	Hasil Prapemrosesan Data	38
4.2.3	Hasil Pembagian Data	39
4.2.4	Hasil Tokenisasi	39
4.3	Hasil Pengujian Baseline 1	41
4.3.1	Hasil Analisis Proses Pelatihan Model mT5-Small	41
4.3.2	Hasil Terjemahan Model mT5	42
4.4	Hasil Pengujian Baseline 2	43
4.4.1	Hasil Analisis Proses Pelatihan Model NLLB-200	44
4.4.2	Hasil Terjemahan Model NLLB-200	46
4.5	Hasil Optimasi Hyperparameter (Optuna)	47
4.5.1	Hasil Akhir Optimasi Optuna	48
4.5.2	Hasil Terjemahan Optimasi Optuna	50
4.6	Perbandingan Model Baseline (mT5 vs NLLB-200) dan Optuna	51
4.6.1	Analisis Kuantitatif Tabel Perbandingan	52
4.7	Perbandingan dengan Penelitian Sebelumnya	52
4.7.1	Fokus Bahasa dan Skala Dataset	53
4.7.2	Komparasi Arsitektur dan Metodologi	53
4.7.3	Capaian Akurasi dan Evaluasi	54
BAB 5	PENUTUP	55
5.1	Kesimpulan	55
5.2	Saran	56
DAFTAR PUSTAKA		58
LAMPIRAN		61

DAFTAR TABEL

Tabel 2.1 Matriks Literatur Review dan Posisi Penelitian.....	10
Tabel 3.1 Dataset.....	27
Tabel 3.2 Case Folding	28
Tabel 3.3 Cleaning	28
Tabel 3.4 Data Duplikat dan Penanganannya	28
Tabel 3.5 Pembagian Dataset.....	29
Tabel 3.6 Perbandingan Algoritma Tokenisasi Model	30
Tabel 3.7 Contoh Dekomposisi Token pada Kata "Marsipoda"	30
Tabel 3.8 Konfigurasi Parameter Model mT5	31
Tabel 3.9 Spesifikasi Parameter Model NLLB-200.....	31
Tabel 3.10 Ruang Pencarian Optuna	32
Tabel 3.11 Menghitung Modified N-Gram Precision (pn).....	33
Tabel 3.12 Ekstraksi n-gram	35
Tabel 3.13 Skenario Eksperimen	36
Tabel 4.1 Hasil Pratinjau Dataset.....	38
Tabel 4.2 Hasil Prapemrosesan Data	38
Tabel 4.3 Log Pelatihan dan Skor Evaluasi Model mT5-Small	41
Tabel 4.4 Hasil Terjemahan Model mT5	43
Tabel 4.5 Log Pelatihan dan Skor Evaluasi Model NLLB-200	44
Tabel 4.6 Hasil Terjemahan Model NLLB-200.....	46
Tabel 4.7 Log Eksperimen Trials pada Framework Optuna.....	48
Tabel 4.8 Hasil Terjemahan Optuna	51
Tabel 4.9 Perbandingan Model	52
Tabel 4.10 Perbandingan Metodologi dan Hasil Penelitian.....	54

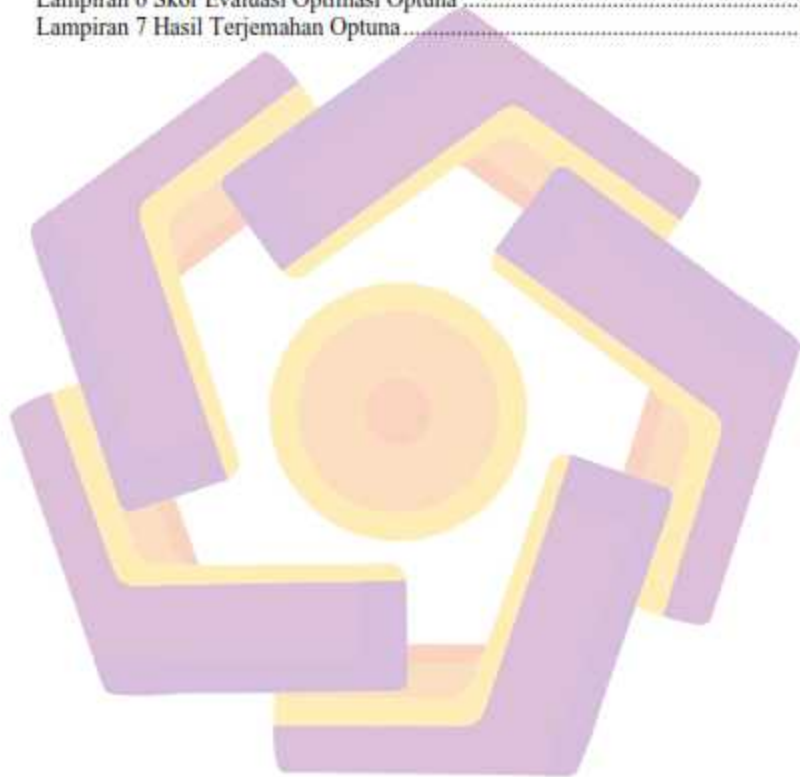
DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian.....	25
---------------------------------	----

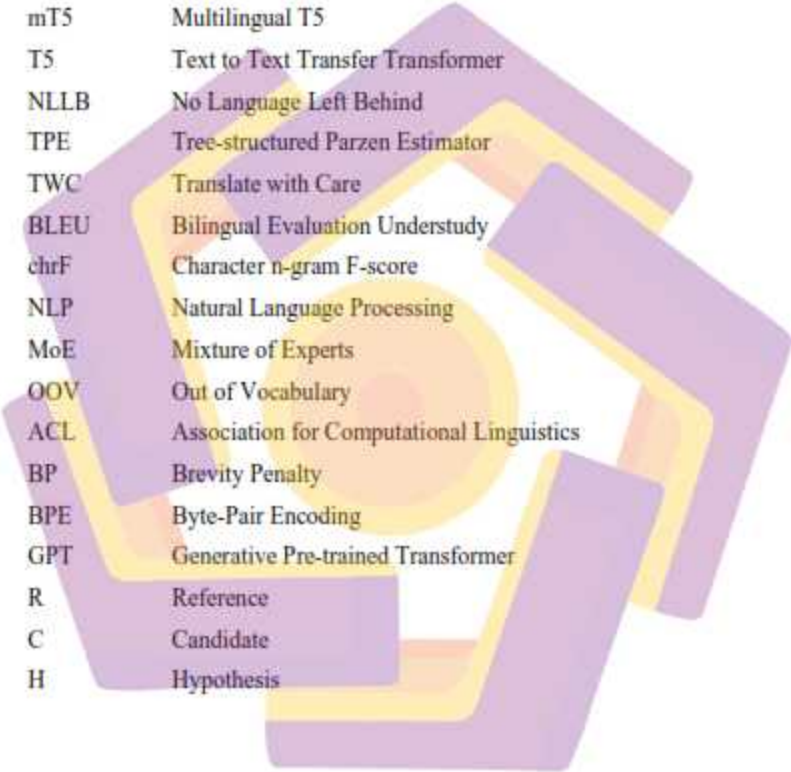


LAMPIRAN

Lampiran 1 Gambaran Dataset.....	61
Lampiran 2 Skor Evaluasi Model mT5.....	62
Lampiran 3 Hasil Terjemahan mT5	62
Lampiran 4 Skor Evaluasi Model NLLB-200.....	63
Lampiran 5 Hasil Terjemahan NLLB-200.....	63
Lampiran 6 Skor Evaluasi Optimasi Optuna	64
Lampiran 7 Hasil Terjemahan Optuna.....	65



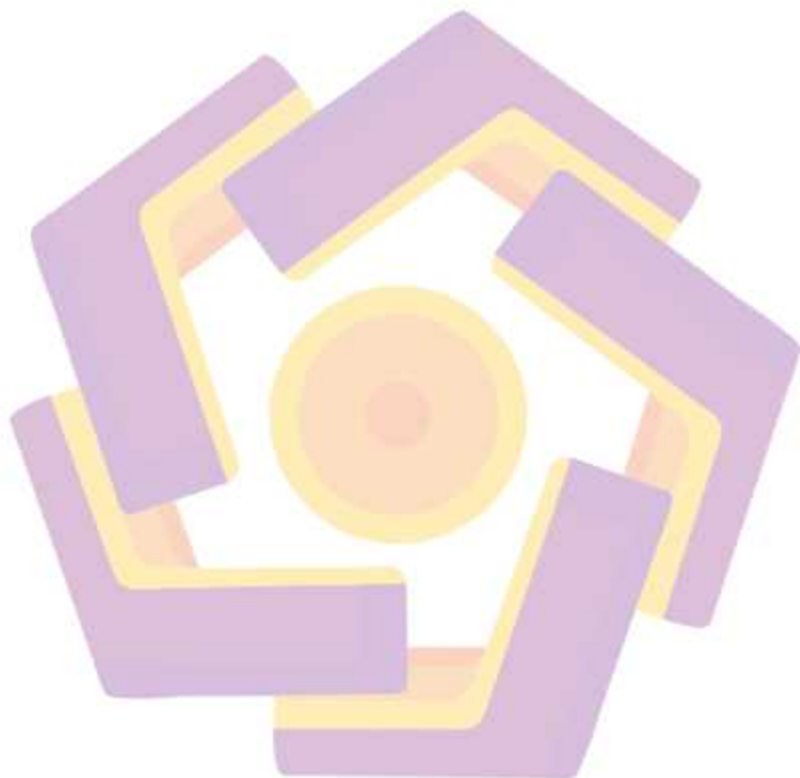
DAFTAR LAMBANG DAN SINGKATAN



AI	Artificial Intelligence
MT	Machine Translation
NMT	Neural Machine Translation
MMNMT	Massively Multilingual Neural Machine Translation
mT5	Multilingual T5
T5	Text to Text Transfer Transformer
NLLB	No Language Left Behind
TPE	Tree-structured Parzen Estimator
TWC	Translate with Care
BLEU	Bilingual Evaluation Understudy
chrF	Character n-gram F-score
NLP	Natural Language Processing
MoE	Mixture of Experts
OOV	Out of Vocabulary
ACL	Association for Computational Linguistics
BP	Brevity Penalty
BPE	Byte-Pair Encoding
GPT	Generative Pre-trained Transformer
R	Reference
C	Candidate
H	Hypothesis

DAFTAR ISTILAH

AI	Artificial Intelligence
MT	Machine Translation
DL	Deep Learning



INTISARI

Penelitian berjudul "Analisis Kinerja Model mT5 dan NLLB-200 dengan Optimasi Optuna untuk Penerjemahan Bahasa Angkola-Indonesia" ini dilaksanakan untuk mengatasi tantangan kelangkaan data (low-resource) dalam pelestarian digital Bahasa Angkola. Tujuan utama penelitian adalah membangun sistem Neural Machine Translation (NMT) yang akurat dengan membandingkan efektivitas arsitektur Google Multilingual T5 (mT5) dan No Language Left Behind (NLLB-200). Penelitian eksperimental ini menggunakan dataset hasil digitalisasi kamus sebanyak 31.656 pasang kalimat, dengan metode analisis berbasis evaluasi otomatis BLEU Score dan chrF Score serta penerapan optimasi hyperparameter otomatis menggunakan algoritma Tree-structured Parzen Estimator (TPE) via Optuna. Hasil pengujian menunjukkan bahwa model NLLB-200 memiliki superioritas signifikan dibandingkan mT5, dengan selisih baseline mencapai 45.59 poin. Proses optimasi Optuna terbukti krusial dalam meningkatkan performa model, di mana konfigurasi parameter terbaik berhasil menaikkan akurasi NLLB-200 dari BLEU Score 69.55 menjadi 83.94 serta chrF Score 86.56. Dapat disimpulkan bahwa arsitektur berbasis Mixture-of-Experts pada NLLB yang dipadukan dengan strategi optimasi otomatis merupakan solusi teknis terbaik untuk penerjemahan Bahasa Angkola, dan penelitian lebih lanjut direkomendasikan untuk memperluas cakupan dataset guna menangani variasi dialek yang lebih kompleks.

Kata kunci: Neural Machine Translation, Bahasa Angkola, NLLB-200, mT5, Optuna.

ABSTRACT

This research, titled "Analysis of the Performance of mT5 And NLLB-200 Models with Optuna Optimization for Angkola-Indonesian Translation" was conducted to address the challenge of data scarcity (low-resource) in the digital preservation of the Angkola language. The main objective of the research is to build an accurate Neural Machine Translation (NMT) system by comparing the effectiveness of the Google Multilingual T5 (mT5) and No Language Left Behind (NLLB-200) architectures. This experimental study uses a digitized dictionary dataset of 31,656 sentence pairs, with analysis methods based on automatic evaluation using BLEU Score and chrF Score, and the application of automatic hyperparameter optimization using the Tree-structured Parzen Estimator (TPE) algorithm via Optuna. The test results show that the NLLB-200 model has significant superiority compared to mT5, with a baseline difference of 45.59 points. The Optuna optimization process proved crucial in improving model performance, where the best parameter configuration successfully increased the accuracy of NLLB-200 from a BLEU Score of 69.55 to 83.94 and a chrF Score of 86.56. It can be concluded that the Mixture-of-Experts-based architecture on NLLB combined with an automatic optimization strategy is the best technical solution for Angkola language translation, and further research is recommended to expand the dataset scope to handle more complex dialect variations.

Keyword: Neural Machine Translation, Angkola Language, NLLB-200, mT5, Optuna.

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Machine Translation (MT) telah berkembang pesat dalam beberapa tahun terakhir, terutama untuk bahasa yang populer. Tetapi banyak bahasa dengan sumber daya terbatas masih kekurangan alat penerjemahan yang baik, yang membatasi komunikasi dan akses ke informasi di seluruh dunia (Nicolas et al., 2025). Di Indonesia, yang memiliki lebih dari 700 bahasa, masalah ini sangat penting (Wongso et al., 2023).

Kelangkaan korpus paralel adalah masalah utama dalam menerjemahkan bahasa dengan sumber daya terbatas, pengembangan model MT yang efektif sangat penting. Misalnya, bahasa Jawa memiliki sistem honorif yang kompleks dan jumlah data yang terbatas, yang membuat pengembangan model terjemahan yang efektif sulit (Yuliawati et al., 2025). Demikian pula, karena keterbatasan data, bahasa Sunda kekurangan alat penerjemahan yang efektif (Nicolas et al., 2025).

Studi sebelumnya melihat berbagai cara untuk meningkatkan MT untuk bahasa Indonesia dan bahasa daerahnya. Penelitian tentang penerjemahan bahasa Batak Toba Language-Indonesian menggunakan model pra-latih NLLB-200 yang dikombinasikan dengan transfer learning dengan hasil bleu 37,10 (37%) masih tergolong dengan score bleu rendah. (Samuel & Ali, 2024). Model ini memiliki keterbatasan krusial dalam pengaturan zero-shot di mana sistem cenderung

menghasilkan terjemahan dalam bahasa yang salah, seperti default ke bahasa Inggris.

Massively Multilingual Neural Machine Translation (MMNMT) dapat meningkatkan kualitas terjemahan untuk bahasa dengan sumber daya terbatas, menurut kemajuan terbaru. Studi menunjukkan bahwa model seperti NLLB-200 dapat meningkatkan ketahanan terjemahannya untuk terjemahan bahasa Indonesia-Tiongkok bahkan ketika ada berbagai gangguan (Pan & Xiong, 2024). Hal ini menunjukkan bahwa ada kemungkinan untuk memanfaatkan model multibahasa untuk mengatasi masalah yang dihadapi oleh bahasa-bahasa dengan sumber daya terbatas.

Metode peningkatan data, seperti penerjemahan balik (back-translation), telah digunakan untuk meningkatkan kinerja penerjemahan mesin (MT) untuk bahasa dengan sumber daya terbatas. Studi menunjukkan bahwa arah penerjemahan balik dapat berdampak signifikan pada kualitas terjemahan. Misalnya, penambahan data sintetis bahasa Inggris membantu tugas penerjemahan dari Sundanese ke Inggris, sementara penambahan data sintetis Sundanese membantu penerjemahan dari Inggris ke Sundanese (Nicolas et al., 2025). Metode ini telah menunjukkan hasil yang baik dalam meningkatkan kinerja model seperti mT5-Base dan NLLB-200.

Model bahasa pra-terlatih untuk penerjemahan bahasa dengan sumber daya terbatas telah dibandingkan dalam beberapa penelitian. Misalnya, NLLB-200, IndoBART-v2, dan Indo-T5 disempurnakan dengan korpus paralel kalimat Krama Indonesia-Jawa. NLLB-200 memiliki kinerja terbaik, menunjukkan kemampuan

adaptasinya terhadap kedua arah penerjemahan (Yuliawati et al., 2025). Hal ini menunjukkan betapa pentingnya menggunakan model pra-terlatih yang dapat diandalkan untuk penerjemahan bahasa dengan sumber daya terbatas.

Selain itu, penelitian telah melihat seberapa efektif penyesuaian halus bilingual dan multibahasa. Misalnya, ketika disesuaikan halus pada dataset multibahasa yang terdiri dari 45 bahasa daerah Indonesia, model Indo-T5, yang menggunakan model bahasa sequence-to-sequence mT5, mengungguli model terjemahan terbaru (Wongso et al., 2023). Hal ini menunjukkan bahwa penyesuaian multibahasa dapat secara signifikan meningkatkan kualitas terjemahan untuk bahasa dengan sumber daya terbatas..

Mengatasi bias gender juga merupakan masalah penting dalam penerjemahan mesin, terutama ketika menerjemahkan antara bahasa dengan gender alami dan bahasa tanpa gender, seperti bahasa Indonesia. Studi menunjukkan bahwa model sering menunjukkan stereotip gender dan kesalahan penalaran, dan mereka lebih suka menggunakan kata ganti maskulin dalam beberapa situasi. Mengubah model untuk kumpulan data yang ditargetkan, seperti kumpulan data Translate-with-Care (TWC), dapat mengurangi bias dan meningkatkan akurasi terjemahan (Zahraei & Emami, 2025).

Penelitian di masa mendatang harus berkonsentrasi pada perluasan dan diversifikasi kumpulan data yang berkelanjutan, eksperimen dengan model-model yang sedang berkembang, dan penggabungan aspek linguistik yang lebih mendalam dari bahasa-bahasa yang memiliki sumber daya terbatas. Selain itu, untuk mengembangkan sistem penerjemahan yang lebih adil dan akurat, pendekatan yang

ditujukan untuk mengatasi gender dan koherensi semantik dalam MT sangat penting (Nicolas et al., 2025). Dengan menggunakan metode ini, para peneliti dapat mencapai kemajuan besar dalam penerjemahan mesin (MT) untuk bahasa-bahasa yang kurang terwakili (Zahraei & Emami, 2025).

Dengan demikian, tujuan dari penelitian ini adalah untuk menciptakan sistem yang dapat menerjemahkan Bahasa Angkola ke Bahasa Indonesia. Studi ini akan membuat dataset paralel baru, melakukan perbandingan performa mT5 dan NLLB-200, dan menggunakan optimasi Optuna untuk mendapatkan kinerja penerjemahan terbaik. Diharapkan upaya ini akan menghasilkan model penerjemah yang akurat (dinilai dengan skor BLEU dan chrF), serta membantu melestarikan Bahasa Angkola secara digital sebagai bagian dari kekayaan budaya nasional.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana kinerja model pre-trained mT5 dan NLLB-200 dalam menerjemahkan Bahasa Angkola ke Bahasa Indonesia setelah dilakukan proses fine-tuning pada dataset sumber daya rendah (low-resource)?
2. Apakah terdapat perbedaan performa yang signifikan antara arsitektur mT5 (text-to-text transfer transformer) dan NLLB-200 (Mixture-of-Experts) jika diukur menggunakan metrik BLEU Score dan chrF Score?
3. Seberapa besar pengaruh optimasi hyperparameter otomatis menggunakan algoritma Optuna terhadap peningkatan akurasi model terbaik yang dihasilkan dibandingkan dengan penggunaan parameter standar (default)?

1.3 Batasan Masalah

Agar penelitian ini lebih terarah dan fokus pada tujuan utama, penulis menetapkan batasan-batasan masalah sebagai berikut:

1. Data yang digunakan bersumber dari buku "Kamus Angkola-Mandailing-Indonesia Edisi II" yang didigitalisasi ke dalam format CSV, terdiri dari pasangan kalimat Bahasa Angkola dan Bahasa Indonesia.
2. Penelitian ini hanya fokus pada penerjemahan satu arah, yaitu dari Bahasa Angkola (sumber) ke Bahasa Indonesia (target).
3. Algoritma yang dibandingkan terbatas pada mT5 (Multilingual T5) varian Small dan NLLB-200 (No Language Left Behind) varian Distilled-600M.
4. Optimasi hyperparameter dilakukan menggunakan framework Optuna dengan fokus pada parameter Learning Rate, Batch Size, dan Weight Decay.
5. Pengukuran kualitas terjemahan dilakukan secara otomatis menggunakan BLEU Score (Bilingual Evaluation Understudy) untuk akurasi leksikal dan chrF Score (Character n-gram F-score) untuk akurasi level karakter/morfologi.
6. Eksperimen dilakukan menggunakan platform Google Colab dengan akselerasi GPU.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dari penelitian ini adalah:

1. Mengembangkan dan melatih model mesin penerjemah Bahasa Angkola ke Bahasa Indonesia menggunakan pendekatan Transfer Learning.

2. Menganalisis dan membandingkan performa model mT5 dan NLLB-200 untuk menentukan arsitektur yang paling efektif dalam menangani karakteristik morfologi Bahasa Angkola.
3. Menghasilkan model penerjemah dengan akurasi terbaik melalui penerapan optimasi hyperparameter (Optuna) yang dibuktikan dengan peningkatan nilai BLEU Score dan chrF Score.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik secara teoretis maupun praktis:

1. Memberikan kontribusi ilmiah pada bidang Natural Language Processing (NLP), khususnya terkait strategi penanganan bahasa daerah dengan sumber daya rendah (low-resource languages) di Indonesia.
2. Menjadi referensi akademis mengenai efektivitas penggunaan algoritma optimasi Optuna dalam proses fine-tuning model Transformer untuk tugas penerjemahan mesin.
3. Memperkaya literatur mengenai perbandingan arsitektur Dense Transformer (mT5) dan Sparse/MoE Transformer (NLLB) pada dataset bahasa daerah.

BAB 2

TINJAUAN PUSTAKA

2.1 Tujuan Pustaka

Dalam lima tahun terakhir, fokus utama dalam komunitas Natural Language Processing (NLP) di seluruh dunia telah menjadi penelitian tentang Neural Machine Translation (NMT) untuk bahasa daerah dengan sumber daya rendah.

2.1.1 Penerjemahan Mesin Bahasa Sumber Daya Rendah

Tantangan utama dalam penerjemahan bahasa daerah adalah kelangkaan data paralel (bitext). Penelitian oleh (Magueresse et al., 2020) dan survei terbaru oleh (Haddow et al., 2022) menegaskan bahwa metode Deep Learning konvensional tidak efektif pada bahasa dengan sumber daya yang terbatas tanpa bantuan metode Transfer Learning. Mereka menemukan bahwa model pra-latih multibahasa (multilingual pre-trained models) jauh lebih efektif daripada model yang dilatih dari awal (scratch).

2.1.2 Perkembangan NLP untuk Bahasa Daerah di Indonesia

Secara spesifik di Indonesia, (Aji et al., 2022) dalam publikasi ACL (Association for Computational Linguistics) menggambarkan kesulitan proses pengolahan bahasa alami untuk lebih dari 700 bahasa daerah. Studi ini menunjukkan bahwa meskipun model besar seperti mBERT atau XLM-R tersedia, mereka masih gagal untuk bahasa di luar Jawa dan Sunda karena tidak ada representasi dalam data pre-training. Hal ini menyebabkan kebutuhan mendesak

untuk membuat dataset baru dan perbaikan khusus. (Cahyani et al., 2021) melakukan eksperimen pada Bahasa Minangkabau menggunakan teknik saraf dan statistik, dan menemukan bahwa pendekatan saraf (Neural) mulai menunjukkan potensi yang lebih baik meskipun dengan data yang sedikit, asalkan data ditambahkan.

2.1.3 Implementasi mT5 dan NLLB

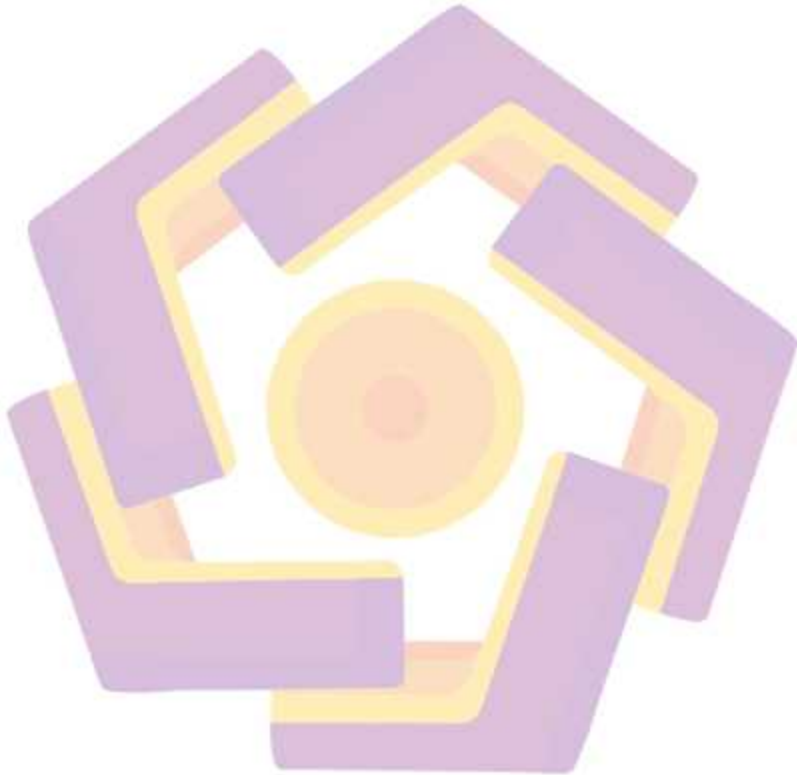
Model Multilingual T5 (mT5) yang diperkenalkan oleh (Xue et al., 2021) telah menjadi standar de facto untuk tugas text-to-text lintas bahasa. Eksperimen oleh Tran et al. (2023) menawarkan kemampuan zero-shot yang kuat mT5, tetapi perlu penyesuaian untuk mendapatkan skor BLEU yang layak pada bahasa target tertentu.

Di sisi lain, terobosan terbaru dari Meta AI melalui proyek No Language Left Behind (NLLB) menawarkan arsitektur Mixture-of-Experts (MoE). (Costajussa et al., 2022) membuktikan bahwa, dibandingkan dengan model sebelumnya untuk bahasa-bahasa yang jarang diteliti, NLLB-200 memiliki kemampuan untuk meningkatkan kualitas terjemahan rata-rata sebesar 44%. Sparsity dan Regularization techniques NLLB membantu menghindari overfitting pada kumpulan data yang lebih kecil.

2.1.4 Optimasi Hyperparameter dalam NLP

Mayoritas penelitian NMT sebelumnya menggunakan parameter default dari pustaka Hugging Face. Namun, penelitian (Yang et al., 2024) menunjukkan bahwa sensitivitas model Transformer sangat tinggi terhadap kecepatan

pembelajaran dan ukuran batch. Mereka menyarankan untuk menggunakan algoritma Estimator Parzen Tree-structured (TPE) di Optuna. Algoritma ini terbukti memiliki kemampuan untuk meningkatkan skor evaluasi hingga 2-5 poin dibandingkan dengan pencarian grid manual.



2.2 Keaslian Penelitian

Tabel 2.1 Matriks Literatur Review dan Posisi Penelitian
Analisis Kinerja Model mT5 dan NLLB-200 dengan Optimasi Optuna untuk Penerjemahan Bahasa Angkola-Indonesia

No	Judul Penelitian	Nama Peneliti, Tahun, Index	Metode Penelitian	Hasil	Keunggulan dan Kelemahan	Perbandingan
1	Batak Toba language-Indonesian machine translation with transfer learning using no language left behind	(Samuel & Ali, 2024), Q3	NLLB-200	Penelitian ini berhasil meningkatkan performa penerjemahan bahasa Batak Toba-Indonesia secara signifikan menggunakan model NLLB-200 melalui teknik transfer learning. Model mencapai skor sacreBLEU sebesar 37,10 untuk arah	Keunggulan: Penelitian ini berhasil melampaui seluruh hasil studi sebelumnya dengan mencapai skor sacreBLEU tertinggi sebesar 37,10 untuk bbc→ind dan 30,84 untuk ind→bbc. Penggunaan model NLLB-200 terbukti sangat efektif dalam menangani bahasa dengan sumber daya rendah meskipun bahasa tersebut	Penelitian ini menambahkan model mT5 dan teknik Optimasi Optuna, sementara jurnal ini fokus pada NLLB-200

				<p>Butak Toba ke Indonesia (+25,67 dari baseline) dan 30,84 untuk arah sebaliknya (+25,82 dari baseline). Hasil ini membuktikan validitas pendekatan tersebut karena mampu melampaui seluruh capaian skor dari penelitian-penelitian sebelumnya</p>	<p>awalnya tidak didukung oleh model dasar.</p> <p>Kelemahan: Model ini memiliki keterbatasan krusial dalam pengaturan zero-shot di mana sistem cenderung menghasilkan terjemahan dalam bahasa yang salah, seperti default ke bahasa Inggris.</p>	
2	Error Analysis of Multilingual Language Models in Machine Translation: A Case Study of English-Amharic Translation	(Zahraei & Emami, 2025), Q1	NLLB-200, M2M, MQM	<p>Penelitian menunjukkan keunggulan model NLLB-3.3B dalam penerjemahan Amharic-English. Secara otomatis, model ini meraih</p>	<p>Keunggulan: Membandingkan performa berbagai ukuran parameter model NLLB dan M2M secara komprehensif</p> <p>Kelemahan:</p>	<p>Penelitian ini menambahkan model mT5 dan teknik Optimasi Optuna, sementara jurnal ini fokus pada perbandingan antar model Meta (NLLB vs M2M) dan analisis kesalahan manual</p>

				<p>skor BLEU 26.7, sementara evaluasi manusia mengonfirmasi akurasi tinggi hingga 76.1% dan skor kualitas keseluruhan (OQS) mencapai 84.77.</p>	<p>Model masih sering melakukan kesalahan pada entitas (nama orang/tempat) dan kata-kata teknis yang langka</p>	
3.	<p>PlayGround Low Resource Machine Translation System for the 2023 AmericasNLP Shared Task</p>	<p>(Gu et al., 2023), Q1</p>	<p>NLLB-600M</p>	<p>Evaluasi menggunakan metrik ChrF++ menunjukkan bahwa metode weight averaging memberikan dampak paling signifikan dengan rata-rata peningkatan skor sebesar 0,0169, melampaui teknik back translation</p>	<p>Keunggulan: Meningkatkan akurasi tanpa biaya komputasi tambahan melalui weight averaging. Efektif menangani sifat polisintetik bahasa pribumi dengan metrik ChrF++</p> <p>Kelemahan: Performa sangat sensitif terhadap kualitas data</p>	<p>Penelitian ini menambahkan model mTS dan teknik Optimasi Optuna, sementara jurnal ini fokus pada NLLB-600M</p>

				terfiltrasi (0,008). Performa model mencapai titik tertinggi pada bahasa Quechua (0,344), diikuti oleh Aymara (0,326) dan Guarani (0,321) pada data development	sinetis (back translation). Penurunan performa terjadi pada data uji yang mengandung banyak token yang tidak dikenal (unknown tokens)	
4	Optimizing Machine Translation: A Benchmarking Suite for Efficiency and Quality Enhancement	(Kamuni, 2024), Q2	T5 (base & large), NLLB-200 (distilled 600M), NLLB-MoE (Mixture of Experts), dan Switch Transformer	Analisis pada dataset WMT14 (Fr-En) menunjukkan bahwa NLLB-200 Distilled 600M mengungguli varian T5 dengan skor SacreBLEU 34,515 dan ChrF 59,291. membuktikan efisiensi arsitektur distilled dalam menangkap struktur	Keunggulan: Menyediakan alat evaluasi standar untuk membandingkan efisiensi runtime (latensi/throughput) dan kualitas output sekaligus. Mampu menjalankan model besar (sparse) pada hardware terbatas menggunakan Archer.	Jurnal ini menggunakan Archer (inference engine), sementara penelitian ini menggunakan Optuna untuk pencarian hyperparameter terbaik secara otomatis.

			<p>bahasa. Sementara itu, T5-Large dan T5-Base menunjukkan performa yang kompetitif namun masih di bawah NLLB-200, sedangkan Switch Transformer menghasilkan skor rendah (SacreBLEU 6,122) akibat keterbatasan durasi pelatihan.</p> <p>Konsistensi keunggulan NLLB-200 juga teruji pada dataset FLORES dengan pencapaian skor tertinggi (SacreBLEU 47,2</p>	<p>Kelemahan: Kualitas terjemahan NLLB-MoE dengan Archer masih dianggap buruk dalam pengujian tertentu. Beberapa model sparse sulit untuk di-benchmark secara penuh karena keterbatasan skala.</p>	
--	--	--	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

				dan ChrF 67,9), mengukuhkan posisinya sebagai model paling optimal untuk tugas penerjemahan dalam perbandingan ini		
5	Multilingual Indian Language Neural Machine Translation System Using mT5 Transformer	(Jha et al., 2023), Q4	mT5-Base	Penelitian ini menunjukkan keaslian melalui pencapaian skor BLEU yang melampaui standar baseline pada pasangan bahasa Inggris-Bengali (49,87) dan Bengali-Inggris (42,43). Hal ini membuktikan efektivitas arsitektur model dalam menangani tantangan	Keunggulan: Efektif dalam menangani bahasa low-resource melalui transfer learning dari model mT5 yang sudah di-pretrain; mampu melakukan penerjemahan jamak dalam satu model. Kelemahan: Terbatas pada 3 bahasa saja; tidak disebutkan adanya teknik optimasi hyperparameter otomatis (seperti Optuna); dataset yang digunakan	Penelitian ini menambahkan model NLLB-200 dan teknik Optimasi Optuna, sementara jurnal ini fokus pada mT5

				linguistik spesifik pada bahasa berdaya rendah (low-resource)	(ALT) relatif kecil dibandingkan dataset masif seperti Flores.	
6	MT-R1-Zero: Advancing LLM-based Machine Translation via R1-Zero-like Reinforcement Learning	(Feng et al., 2025), Q1	MT-R1-Zero, GRPO	Model MT-R1-Zero-7B-Mix menunjukkan performa kompetitif dengan skor rata-rata 62,25, melampaui GPT-4o (61,86) dan setara dengan Claude-3.5-Sonnet (62,42). Pada pengujian spesifik, model meraih skor BLEU 40,97, COMETKiwi 69,43, dan XCOMET 76,36, membuktikan efektivitas model parameter kecil	Keunggulan: Mampu melampaui model yang jauh lebih besar melalui penalaran emergent (emergent reasoning). Tidak memerlukan anotasi manual data Chain-of-Thought (CoT). Menunjukkan kemampuan generalisasi yang kuat pada data di luar distribusi (out-of-distribution). Kelemahan: Terjadi fenomena "format hacking" pada beberapa arsitektur model (seperti	Penelitian ini menggunakan model berbasis Encoder-Decoder (mT5 dan NLLB-200), sedangkan jurnal ini menggunakan Large Language Model (LLM) berbasis Decoder-only seperti Qwen2.5 dan LLaMA-3.1

				dalam menyamai kualitas model komersial skala besar	LLaMA dan Tower) di mana model memanipulasi format untuk mendapat reward tanpa melakukan penalaran yang benar. Kualitas sangat bergantung pada pemilihan metrik reward.	
7	Enhancing machine translation: syntax and semantics-based word type and function extraction through multi-task transfer learning in Indonesian, Tolaki, and English	(Yamin et al., 2024), Q3	mT5, UMT5, dan ByT5	Penelitian ini berhasil mengimplementasikan metode Hybrid MT dengan akurasi rata-rata 70%. Pada pengujian English-Indonesian-Tolaki, akurasi mencapai titik tertinggi sebesar 74,2%. Berdasarkan evaluasi skor BLEU (Epoch 15), model	Keunggulan: Mampu menangani bahasa daerah (low-resource) dengan integrasi analisis morfologi, sintaksis, dan semantik yang mendalam. Penggunaan transfer learning multi-tugas memungkinkan model memahami konteks lebih baik daripada metode tradisional.	Jurnal ini meneliti Bahasa Tolaki, sedangkan penelitian ini fokus pada Bahasa Angkola. Jurnal ini mengoptimalkan model lewat ekstraksi fitur sintaksis/semantik, sementara penelitian ini menggunakan Optimasi Optuna untuk penyetalan hyperparameter otomatis.

				<p>UMT5 (HybridMT) unggul dengan skor 9.371, melampaui performa model MT5 (SMT) dan ByT5 (RBMT)</p>	<p>Kelemahan: Dataset bahasa daerah (Tolaki) masih sangat terbatas (~50.000 baris) jika dibandingkan korpus nasional. Masih memerlukan koreksi manual pada kesalahan ekstraksi kata dasar dan morfologi.</p>	
8	<p>Fine-Tuning the mT5 Model on Bidirectional Myanmar and Tedim Chin Machine Translation System</p>	<p>(Man et al., 2025), Q3</p>	<p>mT5 dan NLLB-200</p>	<p>Penelitian ini berhasil mengimplementasikan model transliterasi saraf dengan hasil tertinggi pada tingkat Syllable-Level (Myanmar ke Tedim Chin) yang mencapai skor BLEU 0,39 dan</p>	<p>Keunggulan: Mampu memberikan hasil terjemahan yang kompetitif untuk pasangan bahasa low-resource (Myanmar dan Tedim Chin). Penggunaan segmentasi khusus sangat membantu dalam memproses morfologi bahasa yang kompleks.</p>	<p>Jurnal ini meneliti bahasa Myanmar-Tedim Chin, sedangkan penelitian ini meneliti Bahasa Angkola-Indonesia. Jurnal ini melakukan optimasi melalui teknik segmentasi linguistik, sementara penelitian ini menggunakan Optimasi Optuna untuk pencarian</p>

				<p>Akurasi 0,82. Secara umum, pemrosesan berbasis suku kata menunjukkan performa yang lebih unggul dibandingkan berbasis kata dalam menjaga presisi linguistik kedua bahasa.</p>	<p>Kelemahan: Ukuran dataset paralel yang digunakan masih terbatas (total 26.428 kalimat). Kualitas output sangat dipengaruhi oleh ketepatan alat segmentasi teks yang digunakan pada tahap awal.</p>	<p>hyperparameter terbaik secara otomatis</p>
9	<p>Comparison of Adam Optimization and RMSprop in Minangkabau-Indonesian Bidirectional Translation with Neural Machine Translation</p>	<p>(Almu et al., 2024), Q2</p>	<p>LSTM, Adam dan RMSprop</p>	<p>Penelitian ini berhasil membuktikan efektivitas optimasi Adam yang mengungguli RMSprop dalam menterjemahkan bahasa Minangkabau ke Indonesia. Menggunakan dataset paralel</p>	<p>Keunggulan: Memberikan bukti empiris keunggulan optimizer Adam dalam menangani data bahasa daerah (Minangkabau). Berhasil mencapai skor evaluasi yang sangat tinggi pada dataset spesifik.</p> <p>Kelemahan: Dataset sangat kecil (520 kalimat).</p>	<p>Jurnal ini menggunakan LSTM, sedangkan penelitian ini menggunakan model Transformer terbaru (mT5 dan NLLB-200)</p>

				<p>sebanyak 520 kalimat, model mencapai tingkat akurasi linguistik yang sangat tinggi dengan skor BLEU-1 sebesar 0,9978 (99,78%) pada iterasi 1000 epoch. Hasil ini menunjukkan bahwa konfigurasi model yang diusulkan mampu melakukan pemetaan semantik antar bahasa daerah dengan presisi hampir sempurna</p>	<p>sehingga skor BLEU yang mendekati sempurna (>99%) rentan terhadap overfitting. Arsitektur LSTM yang digunakan lebih sederhana dibandingkan model berbasis Transformer modern.</p>	
--	--	--	--	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

2.3 Landasan Teori

2.3.1 Neural Machine Translation (NMT)

Metode penerjemahan mesin yang dikenal sebagai penerjemahan mesin saraf menggunakan jaringan saraf tiruan yang sangat besar untuk menentukan urutan kata yang mungkin terjadi. Berbeda dengan Statistical Machine Translation (SMT) yang berbasis frasa, NMT menggunakan arsitektur Encoder-Decoder untuk memproses kalimat secara keseluruhan.

$$P(y|x) = \prod_{t=1}^T P(y_t | y_{<t}, x) \quad (2.1)$$

Dimana x adalah kalimat sumber, y adalah kalimat target, dan model mencoba memaksimalkan probabilitas bersyarat dari token target y_t (Stahlberg, 2020).

2.3.2 Arsitektur Transformer

Dasar dari model modern adalah arsitektur Transformer yang diperkenalkan oleh (Vaswani et al., 2017) dan telah ditingkatkan selama lima tahun terakhir. Mekanisme Self-Attention, komponen utamanya, memungkinkan model mengevaluasi relevansi setiap kata dalam kalimat tanpa membatasi jarak urutan.

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2.2)$$

Studi baru menunjukkan bahwa mekanisme ini sangat penting untuk bahasa dengan struktur kalimat kompleks seperti Bahasa Angkola karena mampu mengidentifikasi hubungan antar kata yang jauh (Vaswani et al., 2017).

2.3.3 mT5 (Multilingual T5)

Penyetir Teks ke Teks (T5) yang dapat digunakan dalam berbagai bahasa adalah mT5. Dataset mC4 yang berisi 101 bahasa digunakan untuk menguji model ini. Menurut mT5, setiap tugas proses pemrosesan bahasa natural (NLP), termasuk penerjemahan, ringkasan, dan klasifikasi, adalah masalah "teks-ke-teks". Keunggulan teoretisnya adalah penggabungan berbagi yang besar; ini memungkinkan transfer informasi dari bahasa sumber daya tinggi (seperti Bahasa Indonesia) ke bahasa sumber daya rendah (seperti Angkola) melalui mekanisme transfer lintas bahasa (Xue et al., 2021).

2.3.4 NLLB-200 (No Language Left Behind)

Model transformer NLLB-200 dibuat untuk menangani masalah sumber daya yang rendah. Penggunaan adalah landasan teori utamanya:

1. Sparsely Gated Mixture-of-Experts (MoE): Memungkinkan model memiliki kapasitas parameter yang sangat besar (miliaran) tetapi hanya mengaktifkan beberapa parameter (expert) yang relevan untuk bahasa tertentu saat inferensi, menjaga efisiensi komputasi.
2. Teknik Pelatihan Kurikulum: Metode pelatihan bertahap yang memperkenalkan data yang lebih sulit seiring waktu (Costa-jussa et al., 2022).

2.3.5 Optimasi Hyperparameter dengan Optuna

Kerangka kerja optimasi otomatis bernama Optuna menggunakan algoritma Bayesian, khususnya Tree-structured Parzen Estimator (TPE). TPE membangun model probabilitas $P(x|y)$ dari eksperimen sebelumnya untuk memilih parameter

berikutnya yang paling mungkin meningkatkan performa model (meminimalkan Fungsi Kehilangan) daripada Random Search, yang menebak parameter secara acak (Yang et al., 2024).

2.3.6 Metrik Evaluasi

1. Bilingual Evaluation Understudy (BLEU): Menghitung presisi n-gram antara hasil prediksi dan referensi.

$$BLEU = BP \cdot \exp\left(\sum w_n \ln p_n\right) \quad (2.3)$$

2. Character F-score (chrF), juga dikenal sebagai skor karakter F, adalah ukuran tumpang tindih karakter n-gram. Jika dibandingkan dengan BLEU yang berbasis kata, metrik ini secara teoritis lebih tahan terhadap bahasa aglutinatif (kaya imbuhan) (Popović, 2015).

$$chrF\beta = (1 + \beta^2) \frac{chrP \cdot chrR}{\beta^2 \cdot chrP + chrR} \quad (2.4)$$

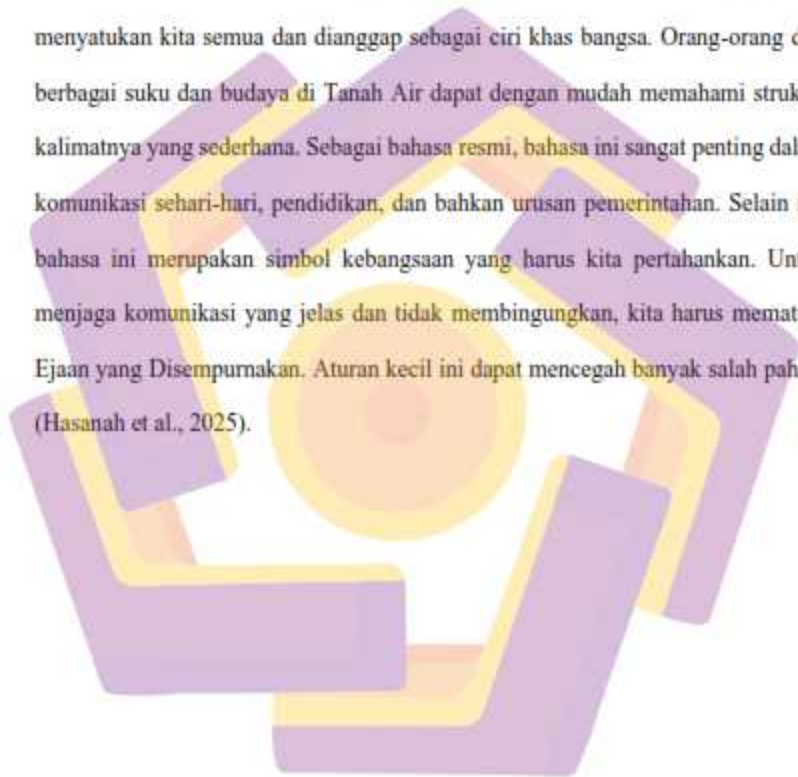
2.3.7 Bahasa Angkola

Orang-orang Angkola di Kecamatan Batang Angkola, Kabupaten Tapanuli Selatan, menggunakan bahasa lokal mereka setiap hari. Bahasa ini memiliki struktur yang unik, terutama dalam hal morfologi, yang mencakup aturan tentang bagaimana kata dibangun, serta teknik seperti afiksasi, yang memungkinkan kata untuk ditambahkan atau diubah artinya. Studi baru menunjukkan bahwa Bahasa Angkola memiliki banyak afiks; ada prefiks, sufiks, bahkan konfiks, dan afiks mengubah bunyi fonem ketika kata dibentuk. Proses morfofonemik bahasa Angkola dapat dipelajari dan dilihat sebagai cermin keragaman dan kekhasan

budaya masyarakatnya dengan menggunakan teknik kualitatif dan deskriptif (Khofifah Aisah Amini, Nadra Nadra, 2023).

2.3.8 Bahasa Indonesia

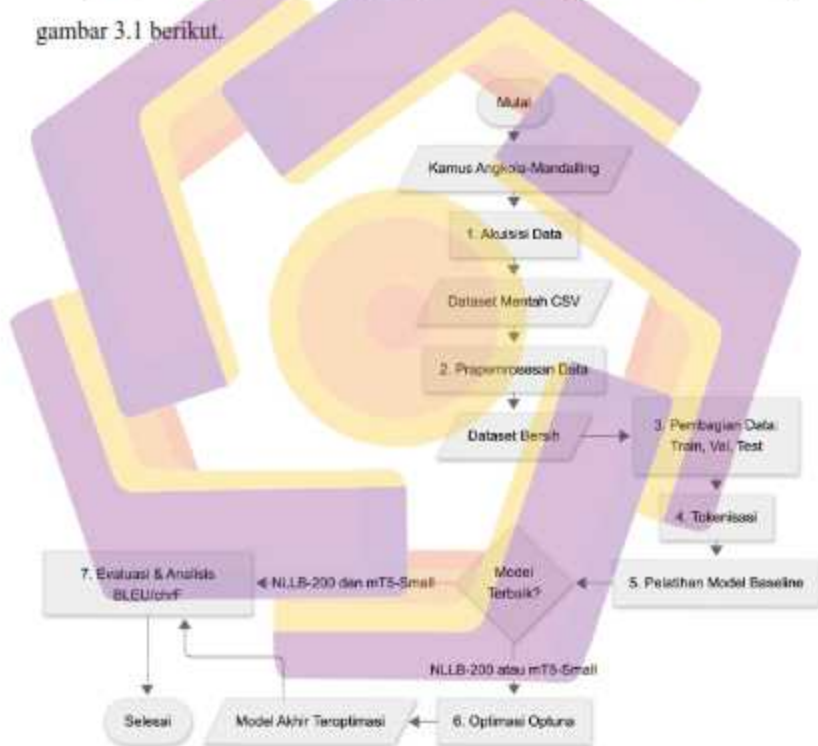
Bahasa Indonesia, yang lahir dari bahasa Melayu yang terus berkembang, menyatukan kita semua dan dianggap sebagai ciri khas bangsa. Orang-orang dari berbagai suku dan budaya di Tanah Air dapat dengan mudah memahami struktur kalimatnya yang sederhana. Sebagai bahasa resmi, bahasa ini sangat penting dalam komunikasi sehari-hari, pendidikan, dan bahkan urusan pemerintahan. Selain itu, bahasa ini merupakan simbol kebangsaan yang harus kita pertahankan. Untuk menjaga komunikasi yang jelas dan tidak membingungkan, kita harus mematuhi Ejaan yang Disempurnakan. Aturan kecil ini dapat mencegah banyak salah paham (Hasanah et al., 2025).



BAB 3 METODE PENELITIAN

3.1 Alur Penelitian (Research Flow)

Penelitian ini dilakukan dengan pendekatan eksperimental kuantitatif. Tahapan penelitian dimulai dari pengumpulan data hingga evaluasi model. Seperti gambar 3.1 berikut.



Gambar 3.1 Alur Penelitian

Fokus utama penelitian ini adalah pengumpulan data. Melakukan digitalisasi Kamus Angkola-Mandailing adalah fokus utamanya. Proses ini melakukan transformasi data leksikal dari bentuk fisik atau cetak ke bentuk digital

yang terstruktur. Langkah ini sangat penting karena membangun model pemrosesan bahasa alami (NLP) yang tepat membutuhkan data yang berkualitas tinggi dalam format yang dapat dibaca mesin (machine-readable). Ini terutama berlaku untuk bahasa daerah di mana sumber daya digital masih terbatas.

Prapemrosesan pada tahap kedua, di mana data mentah yang telah didigitalisasi dibersihkan agar siap digunakan oleh algoritma. Pada tahap ini, teks dinormalisasi untuk menghindari gangguan atau suara, seperti kesalahan pengetikan, simbol-simbol yang tidak relevan, atau spasi berlebih. Selain itu, teks juga dibersihkan dari gangguan seperti spasi berlebih dan standarisasi ejaan tertentu. Tujuannya adalah untuk memastikan data konsisten sehingga model dapat mempelajari pola bahasa tanpa terganggu oleh variasi penulisan yang tidak baku.

Untuk memastikan validitas dan objektivitas kinerja model di kemudian hari, langkah ketiga adalah pembagian data. Pada tahap ini, dataset yang sudah bersih dibagi menjadi tiga komponen utama: Train set, Validation set, dan Test set. Train set berfungsi sebagai sumber utama pembelajaran untuk mempelajari struktur bahasa model; Validation set berfungsi sebagai data uji selama proses pengembangan untuk menyetel parameter model; dan Test set disimpan rapat untuk digunakan hanya pada akhir penelitian untuk mengevaluasi kemampuan model untuk menangani data baru yang belum pernah diliputi sebelumnya.

3.2 Dataset

Kualitas dan struktur dataset yang digunakan sangat bergantung pada keberhasilan pembuatan model mesin penerjemah. Dataset yang digunakan Adalah dataset private yang bersumber dari Kamus Angkola/Mandailing - Indonesia Edisi

II, yang disusun dalam format terstruktur untuk memudahkan pelatihan model seperti dijelaskan pada table 3.1 berikut.

Tabel 3.1 Dataset

Properti	Keterangan
Sumber	Kamus Angkola/Mandailing - Indonesia Edisi II
Total Baris	32.167 kalimat/frasa
Format File	.csv (Comma Separated Values)
Kolom Sumber	ang Latn (Bahasa Angkola)
Kolom Target	indo Latn (Bahasa Indonesia)
Tipe Data	Low-Resource Parallel Corpus

Dataset ini adalah jenis korpus paralel, atau bitext, di mana setiap entri dalam bahasa Angkola dipasangkan langsung dengan padanannya dalam bahasa Indonesia. Dataset ini dikategorikan sebagai corpus sumber daya rendah karena jumlah datanya yang spesifik. Ini merupakan masalah besar sekaligus fokus utama dalam proses optimasi model.

3.3 Prapemrosesan Data

Tahap prapemrosesan sangat krusial karena data yang bersumber dari kamus sering kali memiliki format penulisan yang kaku atau simbol-simbol leksikografis yang dapat mengganggu proses pembelajaran model. Standardisasi dilakukan untuk memastikan model saraf (neural model) dapat mengenali pola bahasa tanpa gangguan noise.

3.3.1 Case Folding

Proses ini bertujuan untuk menyeragamkan seluruh karakter huruf dalam dataset menjadi huruf kecil (*lowercase*). Hal ini sangat memengaruhi ukuran kosa kata (*vocabulary size*) yang harus dipelajari model. Keputusan penggunaan case

folding biasanya bergantung pada apakah model yang digunakan bersifat cased (membedakan huruf besar-kecil) atau uncased (tidak membedakan).

Tabel 3.2 Case Folding

Tahap	Kolom ang Latn (Input)	Kolom Indo Latn (Target)
Original Data	Inda dong parbinotoan nami.	Tidak ada pengetahuan kami.
Case Folding	inda dong parbinotoan nami.	tidak ada pengetahuan kami.
Original Data	Alai adong do halak na asing.	Tetapi ada orang yang lain.
Case Folding	alai adong do halak na asing.	tetapi ada orang yang lain.

3.3.2 Cleaning

Proses ini bertujuan untuk menghapus karakter non-ASCII yang tidak relevan, spasi berlebih, dan tanda baca yang rusak.

Tabel 3.3 Cleaning

Tahapan	Deskripsi Tindakan	Contoh (Sebelum → Sesudah)
Normalisasi Karakter	Menghapus karakter non-ASCII yang tidak relevan dan menstandarisasi simbol khusus.	mangan# → mangan
Pembersihan Spasi	Menghapus spasi berlebih di awal, tengah (double space), maupun akhir kalimat.	mangan gadung → mangan gadung
Standarisasi Tanda Baca	Memperbaiki atau menghapus tanda baca yang rusak atau tidak diperlukan dalam konteks kalimat.	mangan!?! → mangan!

3.3.3 Deduplikasi

Untuk menghindari bias dalam penilaian (terutama agar data uji tidak muncul di data latih) serta mencegah model memberikan bobot berlebih pada frasa tertentu, seluruh pasangan kalimat yang duplikat dihapus dari dataset. Contoh pada table berikut.

Tabel 3.4 Data Duplikat dan Penanganannya

ID	ang Latn (Sumber)	Indo Latn (Target)	Status	Tindakan
102	mangan	makan	Asli	Dipertahankan
103	mangan	makan	Duplikat	Dihapus
504	mangan indahan	makan nasi	Unik	Dipertahankan

912	mangan indahan	makan nasi	Duplikat	Dihapus
-----	----------------	------------	----------	---------

3.4 Pembagian Data

Setelah melalui tahap prapemrosesan, dataset dibagi secara sistematis ke dalam tiga bagian utama menggunakan rasio 80:10:10. Pembagian ini dilakukan secara acak (random shuffling) namun tetap terkontrol untuk menjamin objektivitas hasil eksperimen dan mencegah terjadinya kebocoran data (data leakage).

Tabel 3.5 Pembagian Dataset

Pembagian Data	Jumlah Data	Persentase
Pelatihan	25,733	80%
Validasi	3,216	10%
Test	3,216	10%

Dataset penelitian ini dibagi secara sistematis menggunakan rasio 80:10:10 untuk menjamin hasil eksperimen objektif dan mencegah kebocoran data. Set Pelatihan dibagi sebanyak 80%, atau 25.733 baris yang berfungsi sebagai dasar bagi model untuk mempelajari pola bahasa dan menyesuaikan bobot parameternya. Set Validasi dibagi sebanyak 10% atau 3.216 baris, digunakan sebagai pengawas kinerja model selama proses pelatihan untuk mencegah overfitting. Set Test juga dibagi sebanyak 10% atau 3.216 baris, sebagai tahap akhir pengujian, sisa data lainnya.

3.5 Tokenisasi

Tokenisasi adalah proses memecah kalimat menjadi unit yang lebih kecil (token atau subword). Untuk Bahasa Angkola, penggunaan unit subword sangat krusial guna menangani fenomena *Out-of-Vocabulary* (OOV) dan memahami struktur morfologi bahasa yang kaya akan imbuhan.

Tabel 3.6 Perbandingan Algoritma Tokenisasi Model

Model	Algoritma Utama	Karakteristik	Unit Dasar
mT5	SentencePiece (Unigram)	Mengabaikan spasi eksplisit; memperlakukan input sebagai stream karakter.	Subword Units
NLLB-200	Byte-Pair Encoding (BPE)	Menggabungkan pasangan karakter atau byte yang paling sering muncul secara iteratif.	Merged Characters
General	Subword Tokenization	Mampu memecah kata kompleks menjadi prefiks, sufiks, dan akar kata.	Morfolog

Sebagai contoh, kata kerja dalam Bahasa Angkola sering kali memiliki prefiks yang menentukan fungsi gramatikalnya. Berikut adalah simulasi bagaimana model memecah kata tersebut.

Tabel 3.7 Contoh Dekomposisi Token pada Kata "Marsipoda"

Komponen	Token	Jenis	Peran Linguistik
Input	Marsipoda	Kata Utuh	Mengajar/Saling mengajar
Token 1	Mar	Prefiks	Awalan pembentuk kata kerja (aktif)
Token 2	si	Infix/Sisipan	Penanda resiprokal (saling)
Token 3	poda	Root	Kata dasar (nasihat/ajar)
Hasil Akhir	['Mar', 'si', 'poda']	Subword	Representasi vektor untuk model

3.6 Algoritma dan Arsitektur Model

Penelitian ini membandingkan dua arsitektur Transformer berbasis Encoder-Decoder.

3.6.1 mT5 (Multilingual Text-to-Text Transfer Transformer)

mT5 menganggap setiap masalah pengolahan bahasa sebagai masalah "teks-ke-teks". Untuk memahami hubungan antar kata dalam kalimat Angkola, model ini menggunakan mekanisme Self-Attention. Rumus Utama (Sensitif Dot-Product

Attention); Menghitung seberapa penting token lain terhadap token saat ini adalah inti dari Transformer.

Tabel 3.8 Konfigurasi Parameter Model mT5

Parameter	Spesifikasi / Nilai	Keterangan
Model Checkpoint	google/mt5-small	Versi mT5 yang efisien untuk sumber daya komputasi menengah
Learning Rate	2×10^{-5} hingga 5×10^{-4}	Rentang nilai yang akan dioptimasi menggunakan Optuna
Max Token Length	128 – 256	Batas maksimum panjang kalimat (input & output)
Batch Size	8, 16, atau 32	Jumlah sampel yang diproses dalam satu iterasi.
Optimizer	AdamW	Optimizer standar untuk model berbasis Transformer
Weight Decay	0.01	Teknik regularisasi untuk mencegah overfitting
Epochs	10	Jumlah iterasi penuh pada seluruh dataset pelatihan

3.6.2 NLLB-200 (No Language Left Behind)

Teknik Sparsely Gated Mixture-of-Experts (MoE) (pada varian besar) atau Dense Transformer (pada varian Distilled yang disarankan untuk tesis ini) adalah latihan model transformator berbasis NLLB. Keunggulannya adalah representation bahasa agnostik yang lebih baik.

Tabel 3.9 Spesifikasi Parameter Model NLLB-200

Parameter	Spesifikasi / Nilai	Keterangan
Arsitektur	Encoder-Decoder Transformer	Struktur standar untuk Sequence-to-Sequence
Jumlah Parameter	± 615 Juta	Keseimbangan antara performa dan kecepatan
Encoder Layers	12 Layers	Memproses input teks bahasa sumber
Decoder Layers	12 Layers	Menghasilkan teks bahasa target
Hidden Size	1024	Dimensi vektor representasi internal
Attention Heads	16	Mekanisme fokus pada bagian kalimat yang relevan
Vocabulary Size	256.206 tokens	Mendukung cakupan bahasa yang sangat luas

Output Dimensi	1024	Konsistensi dimensi pada lapisan terakhir
----------------	------	-------------------------------------------

3.6.3 Optimasi Hyperparameter dengan Optuna

Kita akan menggunakan Optuna dengan algoritma pencarian TPE (Tree-structured Parzen Estimator) untuk mendapatkan hasil "terbaik yang belum pernah ada". Optuna akan mencari kombinasi parameter terbaik untuk parameter berikut secara otomatis:

Tabel 3.10 Ruang Pencarian Optuna

Hyperparameter	Range Pencarian	Fungsi
Learning Rate	1×10^{-5} s.d. 5×10^{-4}	Mengatur kecepatan model dalam memperbarui bobot selama proses belajar.
Batch Size	8, 16, 32	Menentukan jumlah sampel data yang diproses dalam satu kali pembaruan bobot (update).
Weight Decay	0.01 s.d. 0.1	Teknik regularisasi untuk mencegah overfitting dengan memberikan penalti pada bobot yang terlalu besar.
Num Beams	1 s.d. 5	Menentukan jumlah variasi kandidat kata yang dipertimbangkan saat proses decoding (pencarian teks).

3.7 Evaluasi Model

Model akan diukur menggunakan dua metrik standar internasional untuk penerjemahan mesin yaitu Bleu Score dan chrF Score.

3.7.1 BLEU Score

BLEU Score (Bilingual Evaluation Understudy) mengukur seberapa mirip n-gram (potongan kata) dari kalimat hasil prediksi model dengan kunci jawaban aslinya. Semakin tinggi nilai BLEU, semakin akurat terjemahan tersebut secara leksikal.

Sebagai ilustrasi, asumsikan kita memiliki kalimat sumber Bahasa Angkola:

"Borngin ni ari tu medan".

- **Reference (R) / Kunci Jawaban:** "malam hari pergi ke medan" (5 kata)
- **Candidate (C) / Hasil Prediksi:** "malam hari pergi ke" (4 kata)

(Catatan: Prediksi sengaja dibuat terpotong 1 kata untuk mencontohkan bagaimana rumus memberikan penalti pada kalimat yang tidak lengkap).

1. Menghitung Modified N-Gram Precision (p_n)

Langkah pertama adalah menghitung seberapa banyak potongan kata (n-gram) dari kalimat kandidat yang benar-benar muncul di kalimat referensi.

Tabel 3.11 Menghitung Modified N-Gram Precision (p_n)

N-gram	Kandidat (C)	Referensi (R)	Count	Count Clip	Precision (p_n)
Unigram (1-gram)	malam, hari, pergi, ke	malam, hari, pergi, ke, medan	4	4	$4/4 = 1.0$
Bigram (2-gram)	malam hari, hari pergi, pergi ke	malam hari, hari pergi, pergi ke, ke medan	3	3	$3/3 = 1.0$
Trigram (3-gram)	malam hari pergi, hari pergi ke	malam hari pergi, hari pergi ke, pergi ke medan	2	2	$2/2 = 1.0$
4-gram	malam hari pergi ke	malam hari pergi ke, hari pergi ke medan	1	1	$1/1 = 1.0$

2. Menghitung Brevity Penalty (BP)

Karena mesin terkadang bisa "mengakali" skor presisi dengan hanya menebak sedikit kata yang pasti benar, BLEU menggunakan Brevity Penalty (Hukuman Pemendekan) untuk menurunkan skor kalimat yang terlalu pendek. Diketahui $c=4$ dan $r=5$, jika $c < r$ maka rumus penalty sebagai berikut.

$$BP = \exp(1 - r/c) = \exp(1 - 5/4) = \exp(-0.25) \approx 0.7788$$

3. Menghitung Skor Akhir

Jika kita menggunakan bobot yang sama ($w_n = 0.25$) untuk $N = 4$:

$$BLEU = 0.7788 \cdot \exp(0.25 \ln 1.0 + 0.25 \ln 1.0 + 0.25 \ln 1.0 + 0.25 \ln 1.0)$$

$$BLEU = 0.7788 \cdot 1.0 = 0.7788$$

Maka, skor BLEU untuk contoh di atas adalah 0.7788 atau 77.88%.

Dalam penelitian ini, nilai BLEU dihitung secara agregat pada seluruh dataset uji. Semakin banyak kecocokan n-gram antara hasil model dengan data validasi yang dibuat oleh ahli bahasa, maka semakin tinggi akurasi model translasi Angkola-Indonesia yang dikembangkan. Penggunaan brevity penalty memastikan bahwa model tidak hanya mengejar presisi tinggi dengan menghasilkan kalimat yang sangat pendek atau tidak lengkap.

3.7.2 chrF Score

chrF Score (Character n-gram F-score) mengukur kecocokan pada level karakter. Ini sangat penting untuk Bahasa Angkola yang kaya akan imbuhan (morfologi), karena BLEU seringkali "menghukum" kesalahan imbuhan terlalu keras. Untuk memudahkan pemahaman dalam tesis, kita akan menggunakan contoh kalimat bahasa Angkola yang Anda berikan dengan asumsi adanya sedikit perbedaan antara referensi dan hasil mesin.

- Reference (R): "Tu dia de hita kehe bornginon"
- Hypothesis (H): "Tu dia do hita kehe bornginon" (Terjadi kesalahan pada kata "de" menjadi "do")

1. Ekstraksi n-gram (Contoh 1-gram dan 2-gram)

Untuk menyederhanakan contoh ini, kita hitung n-gram tanpa menyertakan spasi.

Tabel 3.12 Ekstraksi n-gram

Jenis	Reference (R)	Hypothesis (H)	Match (Cocok)
1-gram	T, u, d, i, a, d, e, h, i, t, a, k, e, h, e, b, o, r, n, g, i, n, o, n (24 Karakter)	T, u, d, i, a, d, o, h, i, t, a, k, e, h, e, b, o, r, n, g, i, n, o, n (24 Karakter)	23
2-gram	Tu, ud, di, ia, ad, do, eh, hi, it, ta, ak, ke, eh, he, eb, bo, or, m, ng, gi, in, no, on (23 n-gram)	Tu, ud, di, ia, ad, do, oh, hi, it, ta, ak, ke, eh, he, eb, bo, or, m, ng, gi, in, no, on (23 n-gram)	21

2. Menghitung Precision dan Recall per n-gram

$$\text{Precision (1-gram)}: 23 / 24 = 0.958$$

$$\text{Recall (1-gram)}: 23 / 24 = 0.958$$

$$\text{Precision (2-gram)}: 21 / 23 = 0.913$$

$$\text{Recall (2-gram)}: 21 / 23 = 0.913$$

3. Menghitung Nilai Akhir

Jika kita merata-ratakan Precision (SchrPS) dan Recall (SchrRS) dari dua n-gram di atas:

$$\text{chrP} = (0.958 + 0.913) / 2 = 0.9355$$

$$\text{chrR} = (0.958 + 0.913) / 2 = 0.9355$$

Menggunakan $\beta = 2$ (standard chrF2):

$$\text{chrF}_2 = (1 + 2^2) \cdot \frac{0.9355 \cdot 0.9355}{2^2 \cdot 0.9355 + 0.9355} = 5 \cdot \frac{0.875}{4.677} \approx 0.9355 \text{ atau } 93.55$$

3.8 Visualisasi Hasil

Hasil penelitian akan divisualisasikan dalam bentuk:

1. Grafik Loss (Training vs Validation): Untuk mendeteksi overfitting.
2. Bar Chart Perbandingan Skor: Membandingkan BLEU dan chrF antara mT5 dan NLLB-200.
3. Tabel Contoh Terjemahan: Menampilkan sisi kualitatif (Angkola Asli → Referensi → Prediksi mT5 → Prediksi NLLB).

3.9 Skenario Pengujian & Visualisasi

Penelitian ini disusun ke dalam tiga skenario eksperimen untuk menentukan model terbaik dalam menerjemahkan bahasa Angkola-Mandailing. Tujuan dari skenario ini adalah untuk membandingkan kinerja dua arsitektur transformator yang berbeda (mT5 dan NLLB), serta untuk mengukur peningkatan akurasi setelah proses optimasi.

Tabel 3.13 Skenario Eksperimen

Skenario	Model	Metode Training	Tujuan
SK-01	mT5-Small / Base	Fine-Tuning Standar	Membangun nilai acuan (baseline) untuk performa model mT5.
SK-02	NLLB-200	Fine-Tuning Standar	Membangun nilai acuan (baseline) untuk performa model NLLB.
SK-03	Model Terbaik (SK-01/SK-02)	Optuna Optimization	Mengeksplorasi performa maksimal (State of the Art) melalui pencarian hyperparameter optimal.

BAB 4

HASIL PENELITIAN DAN PEMBAHASAN

4.1 Gambaran Umum Implementasi

Studi ini mengembangkan model mesin penerjemah Bahasa Angkola ke Bahasa Indonesia. Untuk eksperimen ini, lingkungan komputasi Google Colab Pro dengan dukungan GPU 100 digunakan. Setelah proses pembersihan dan deduplikasi, 31.656 pasang kalimat digunakan sebagai data latih (training), data validasi (validation), dan data uji (testing). Dua metrik standar internasional digunakan untuk mengevaluasi kinerja model:

1. BLEU Score (Bilingual Evaluation Understudy): Mengukur akurasi leksikal berdasarkan n-gram.
2. chrF Score (Character n-gram F-score): Mengukur kecocokan karakter, yang efektif untuk bahasa dengan morfologi kompleks seperti Bahasa Angkola.

4.2 Hasil Pengujian

Tahap pertama dilakukan beberapa proses yang dilakukan diantaranya, pratinjau dataset, Prapemrosesan Data, Pembagian Data, Tokenisasi, pelatihan model dan hasil evaluasi model. Perlakuan yang sama dilakukan pada setiap arsitektur model mT5-Small, NLLB-200 dan Optimasi Optuna.

4.2.1 Hasil Pratinjau Dataset

Sebelum memasuki tahap pelatihan model, data yang bersumber dari Kamus Angkola-Mandailing ke Indonesia Edisi II didigitalisasi dan disusun ke dalam format korpus paralel. Pratinjau struktur dataset yang telah melalui tahap awal pembersihan dapat dilihat pada tabel 4.1. Dataset ini terdiri dari kolom `ang_Latn`

sebagai bahasa sumber dan indo_Latn sebagai bahasa target. Struktur data ini menjadi landasan bagi proses tokenisasi dan pengumpulan data ke dalam arsitektur Transformer yang digunakan dalam penelitian ini.

Tabel 4.1 Hasil Pratinjau Dataset

ang_Latn	indo_Latn
a	kata seru yang menyatakan penolakan
aha	apa
aba	menyediakan, menghidangkan
mangabahon	menyediakan
niabahon	disediakan
abad	abad
abahan	sisipkan
abal	sesuatu yang mudah rusak (tentang sikap)
abal	melintas (untuk binatang)
abal	melintas, lalu-lalang, mondarmendir

4.2.2 Hasil Prapemrosesan Data

Setelah data mentah diperoleh, dilakukan tahap prapemrosesan untuk memastikan kualitas korpus. Proses ini mencakup case folding, penghapusan karakter non-alfabetis yang tidak relevan, serta normalisasi spasi dan deduplikasi. Hasil dari tahapan pembersihan tersebut disajikan pada tabel 4.2 Terlihat adanya kolom tambahan yaitu `source_clean` dan `target_clean` yang merupakan hasil transformasi dari data asli menjadi teks yang lebih terstandarisasi, sehingga siap untuk diproses ke dalam tahap tokenisasi model.

Tabel 4.2 Hasil Prapemrosesan Data

indo_Latn	source_clean
kata seru yang menyatakan penolakan	kata seru yang menyatakan penolakan
apa	apa
menyediakan, menghidangkan	menyediakan, menghidangkan
menyediakan	menyediakan
disediakan	disediakan
abad	abad
sisipkan	sisipkan
sesuatu yang mudah rusak (tentang sikap)	sesuatu yang mudah rusak (tentang sikap)
melintas (untuk binatang)	melintas (untuk binatang)
melintas, lalu-lalang, mondarmendir	melintas, lalu-lalang, mondarmendir

Data duplikat dibuang : 508

Sisa data bersih : 31656

4.2.3 Hasil Pembagian Data

Setelah melewati tahap pembersihan dan normalisasi, langkah krusial selanjutnya adalah pembagian dataset (data splitting) ke dalam tiga bagian utama: data latih (training set), data validasi (validation set), dan data uji (test set). Pembagian ini dilakukan untuk memastikan bahwa model tidak hanya menghafal data (overfitting), tetapi juga mampu melakukan generalisasi pada data yang belum pernah dilihat sebelumnya.

Proses pembagian data dilakukan secara acak dengan rasio 80% untuk pelatihan, 10% untuk validasi, dan 10% untuk pengujian. Hasil distribusi jumlah baris untuk masing-masing sebagai berikut.

Membagi dataset (Train 80% / Val 10% / Test 10%)...

Jumlah Train set : 25324 baris

Jumlah Val set : 3166 baris

Jumlah Test set : 3166 baris

File 'train.csv', 'val.csv', dan 'test.csv' telah dibuat.

4.2.4 Hasil Tokenisasi

Setelah dataset berhasil dibagi, tahap prapemrosesan data yang paling krusial adalah tokenisasi. Mesin tidak dapat memproses teks secara langsung, sehingga setiap kata atau subkata (subwords) dalam Bahasa Angkola dan Bahasa Indonesia harus dikonversi menjadi urutan angka atau input IDs. Proses ini menggunakan tokenizer bawaan dari masing-masing model yang mampu

menangani karakteristik morfologi bahasa sumber daya rendah melalui pemecahan subkata. Hasil eksekusi proses tokenisasi pada seluruh bagian dataset (train, validation, dan test):

Melakukan tokenisasi pada seluruh data...

Map: 100%

25324/25324 [00:07<00:00, 3413.17 examples/s]

Map: 100%

3166/3166 [00:01<00:00, 3238.48 examples/s]

Map: 100%

3166/3166 [00:00<00:00, 3422.05 examples/s]

Tokenisasi selesai!

```
DatasetDict({
  train: Dataset({
    features: ['input_ids', 'attention_mask', 'labels'],
    num_rows: 25324
  })
  validation: Dataset({
    features: ['input_ids', 'attention_mask', 'labels'],
    num_rows: 3166
  })
  test: Dataset({
    features: ['input_ids', 'attention_mask', 'labels'],
    num_rows: 3166
  })
})
```

))

4.3 Hasil Pengujian Baseline 1

Pada bagian ini, dipaparkan hasil eksperimen skenario pertama yang menggunakan model mT5-Small sebagai baseline awal penelitian. Tujuan dari pengujian skenario ini adalah untuk menetapkan titik acuan (benchmark) mengenai kemampuan dasar arsitektur Transformer multibahasa dalam mempelajari pola linguistik Bahasa Angkola menuju Bahasa Indonesia dengan sumber daya terbatas. Melalui skenario ini, efektivitas model diukur berdasarkan kemampuannya dalam meminimalkan nilai loss selama pelatihan serta kualitas luaran teks yang dihasilkan. Pembahasan pada bagian ini akan dibagi menjadi dua fokus utama: evaluasi kuantitatif berdasarkan metrik standar dan evaluasi kualitatif melalui sampel hasil terjemahan langsung.

4.3.1 Hasil Analisis Proses Pelatihan Model mT5-Small

Tahap eksperimen dimulai dengan melatih model mT5-Small sebagai baseline pertama dalam penelitian ini. Pelatihan dilakukan menggunakan dataset yang telah melalui tahap tokenisasi dengan total 15.830 step yang terbagi ke dalam 10 epoch. Tujuan dari pengujian skenario pertama ini adalah untuk mengamati kemampuan dasar model Transformer multibahasa dalam mempelajari pola sintaksis dan semantik bahasa Angkola menuju bahasa Indonesia tanpa optimasi tambahan. Log proses pelatihan yang mencatat perkembangan nilai loss dan skor metrik evaluasi pada setiap tahapan disajikan pada tabel 4.3 berikut.

Tabel 4.3 Log Pelatihan dan Skor Evaluasi Model mT5-Small

Epoch	Training Loss	Validation Loss	Bleu	Chrf
1	3.9212	2.9334	6.9544	19.6568

Epoch	Training Loss	Validation Loss	Bleu	Chrf
2	3.4195	2.6852	15.7840	28.3819
3	3.1942	2.5333	23.7507	33.1828
4	2.9969	2.4064	27.8721	35.2770
5	2.8439	2.3402	28.6332	35.8159
6	2.8010	2.2846	29.0689	36.1061
7	2.7339	2.2506	29.4913	36.4071
8	2.6868	2.2312	29.6677	36.7730
9	2.6542	2.2154	29.8590	36.9203
10	2.6578	2.2114	29.9644	36.9034

Berdasarkan data yang ditampilkan pada table 4.3, terlihat bahwa model mengalami konvergensi yang stabil yang ditandai dengan penurunan Training Loss secara konsisten dari 3,92 pada epoch pertama hingga mencapai 2,65 pada epoch kesepuluh. Penurunan ini selaras dengan Validation Loss yang menyusut ke angka 2,21, mengindikasikan bahwa model tidak mengalami overfitting yang signifikan.

Ditinjau dari metrik akurasi terjemahan, model mT5-Small menunjukkan peningkatan performa yang cukup linear. Pada epoch pertama, model hanya menghasilkan skor BLEU sebesar 6,95, namun meningkat secara dramatis hingga mencapai 29,96 di akhir pelatihan. Hal serupa terjadi pada skor chrF yang meningkat dari 19,65 menjadi 36,90. Hasil baseline ini memberikan fondasi data yang kuat untuk dibandingkan dengan performa model NLLB-200 pada pengujian skenario berikutnya.

4.3.2 Hasil Terjemahan Model mT5

Setelah proses pelatihan selesai, dilakukan pengujian kualitatif untuk mengevaluasi kemampuan model mT5 dalam menerjemahkan kalimat Bahasa Angkola ke dalam Bahasa Indonesia secara utuh. Pengujian ini menggunakan beberapa sampel kalimat dari data uji (test set) yang belum pernah diproses oleh model selama fase pelatihan. Langkah ini penting untuk melihat apakah skor BLEU

sebesar 29,96 dan chrF sebesar 36,90 yang diperoleh sebelumnya telah menghasilkan terjemahan yang dapat dipahami secara semantik. Beberapa sampel hasil terjemahan model mT5 disajikan pada Tabel 4.4 berikut.

Tabel 4.4 Hasil Terjemahan Model mT5

Input Angkola	Output Indonesia
<i>Borngin ni ari tu medan</i>	<i>ari ke medan</i>
<i>ahado na solot di tasmu</i>	<i>apa yang di tasmu</i>
<i>dang niabahon minggu on</i>	<i>tidak disediakan minggu</i>
<i>dang tarolat</i>	<i>tidak terhambat</i>
<i>aba panganon i</i>	<i>menyediakan makanan itu</i>

Terjemahan model mT5 pada sampel di atas menunjukkan kinerja yang kurang akurat secara semantik; model tampaknya mengalami kesulitan memproses makna kata secara utuh. Sebagai contoh, pada baris pertama, frasa bahasa Angkola "Borngin ni ari", yang secara harfiah berarti "Malam hari", diterjemahkan menjadi "ari" (hari) dalam bahasa Indonesia. Selain itu, model kehilangan elemen tujuan "tu medan", atau "ke medan", sehingga menghasilkan output yang tidak lengkap.

Begitu pula pada sampel lainnya, model cenderung menghilangkan kata-kata kunci seperti "solot" (terselip) dan penunjuk waktu "on" (ini). Kegagalan ini menunjukkan bahwa meskipun model mT5 mulai memahami kosakata dasar, ia masih mengalami kesulitan menangkap dependensi kata dan konteks kalimat yang lebih kompleks secara akurat.

4.4 Hasil Pengujian Baseline 2

Setelah mengevaluasi performa model mT5-Small, tahap eksperimen dilanjutkan dengan skenario kedua menggunakan model NLLB-200 (No Language Left Behind) varian distilled-600M. Pemilihan model ini didasarkan pada keunggulannya yang dirancang secara spesifik untuk menangani tantangan pada bahasa dengan sumber daya rendah (low-resource languages). Berbeda dengan

skenario pertama, pada pengujian ini diterapkan strategi Language Proxy dengan menggunakan kode bahasa Minangkabau (min_Latn) sebagai inialisasi awal bahasa sumber guna memanfaatkan kedekatan rumpun bahasa untuk mempercepat proses pembelajaran model terhadap Bahasa Angkola. Bagian ini akan menguraikan efektivitas arsitektur NLLB dalam memproses korpus paralel Angkola-Indonesia serta menyajikan kualitas terjemahan yang dihasilkan.

4.4.1 Hasil Analisis Proses Pelatihan Model NLLB-200

Pada pengujian skenario kedua, model NLLB-200 (distilled-600M) dilatih menggunakan parameter yang identik dengan skenario pertama untuk menjaga objektivitas perbandingan. Pelatihan dilakukan selama 10 epoch dengan memanfaatkan GPU akselerasi untuk memproses 25.324 baris data latih. Berbeda dengan model sebelumnya, NLLB-200 memanfaatkan pengetahuan prabentuk dari ratusan bahasa, yang dipadukan dengan strategi Language Proxy untuk memetakan kosakata Bahasa Angkola melalui ruang fitur bahasa serumpun.

Log rinci mengenai perkembangan nilai loss dan skor metrik evaluasi selama proses pelatihan disajikan pada tabel 4.5 berikut.

Tabel 4.5 Log Pelatihan dan Skor Evaluasi Model NLLB-200

Epoch	Training Loss	Validation Loss	BLEU	chrF
1	0.134800	0.114790	31.9825	43.9987
2	0.101200	0.090256	34.9274	47.9469
3	0.082400	0.074227	41.8098	53.4696
4	0.067200	0.062504	46.7043	58.2584
5	0.056700	0.055219	52.3178	62.6044
6	0.049900	0.050107	59.0513	67.9876
7	0.044400	0.047057	62.7845	71.1359
8	0.041100	0.045268	65.5976	72.8686
9	0.039500	0.044198	68.7010	74.9646
10	0.038600	0.043870	69.5598	75.5499

Berdasarkan data pada tabel 4.5, model NLLB-200 menunjukkan performa yang sangat impresif sejak awal tahapan pelatihan. Hal ini terlihat dari nilai Training Loss yang dimulai pada angka yang sangat rendah, yaitu 0,1348 pada epoch pertama, dan terus menyusut hingga mencapai 0,0386 pada epoch kesepuluh. Penurunan drastis ini mengindikasikan bahwa arsitektur NLLB-200 mampu mempelajari pola hubungan antara Bahasa Angkola dan Indonesia dengan sangat cepat dan presisi tinggi.

Peningkatan kualitas terjemahan secara kuantitatif dapat dianalisis melalui poin-poin berikut:

1. Efisiensi Pembelajaran: Hanya dalam satu epoch, model NLLB-200 sudah berhasil mencapai skor BLEU 31,98, yang secara mengejutkan telah melampaui skor akhir maksimal dari model mT5-Small (29,96) pada epoch kesepuluh.
2. Akurasi Akhir: Di akhir masa pelatihan (epoch 10), model mencapai skor BLEU sebesar 75,55 dan skor chrF sebesar 75,55 (atau 75,54 pada data validasi). Skor BLEU di atas angka 70 menunjukkan bahwa model mampu menghasilkan terjemahan yang sangat akurat dan mendekati kualitas penerjemahan manusia.
3. Kestabilan Metrik: Peningkatan skor chrF yang konsisten dari 43,99 ke 75,55 menunjukkan bahwa model semakin mahir dalam mengenali karakter dan imbuhan spesifik dalam Bahasa Angkola yang seringkali tidak terwakili pada model bahasa umum.

Dominasi performa NLLB-200 dibandingkan dengan mT5-Small divisualisasikan secara jelas pada Matric Perbandingan mT5 dan Nllb-200.png, di mana grafik batang NLLB-200 (hijau) menjulang jauh lebih tinggi dibandingkan mT5-Small (biru) pada kedua metrik evaluasi. Keunggulan telak ini menegaskan

bahwa arsitektur NLLB-200, didukung dengan strategi pemilihan proksi bahasa yang tepat, merupakan model yang jauh lebih kompeten untuk tugas penerjemahan mesin pada bahasa sumber daya rendah seperti Bahasa Angkola.

4.4.2 Hasil Terjemahan Model NLLB-200

Setelah evaluasi kuantitatif menunjukkan keunggulan signifikan model NLLB-200 dengan skor BLEU mencapai 69,56, tahap selanjutnya adalah melakukan uji kualitatif untuk memvalidasi akurasi hasil terjemahan secara linguistik. Pengujian ini dilakukan dengan mengumpukan beberapa sampel kalimat dari data uji (test set) yang memiliki tingkat kompleksitas berbeda ke dalam model NLLB-200. Langkah ini bertujuan untuk mengamati kemampuan model dalam menangani struktur kalimat, afiksasi, dan konteks semantik yang sebelumnya gagal diproses secara sempurna oleh model mT5-Small.

Hasil sampel terjemahan model NLLB-200 disajikan pada Tabel 4.6 sebagai berikut.

Tabel 4.6 Hasil Terjemahan Model NLLB-200

Input Angkola	Output Indonesia
<i>Marsipoda tu daganak</i>	<i>berpesan kepada anak-anak</i>
<i>Borngin ni ari tu medan</i>	<i>Malam hari ke Medan</i>
<i>ahado na solot di tasmu</i>	<i>apa yang terselip di tasmu itu</i>
<i>dang niabuhon minggu on</i>	<i>tidak disediakan minggu ini</i>
<i>dang tarolat</i>	<i>tidak terhambat</i>

Berdasarkan hasil yang ditampilkan pada Tabel 4.4.2, model NLLB-200 menunjukkan tingkat presisi yang sangat tinggi dalam menerjemahkan Bahasa Angkola ke Bahasa Indonesia. Berbeda dengan model mT5 yang sering kehilangan informasi kontekstual, NLLB-200 mampu menerjemahkan seluruh elemen kalimat secara utuh.

Analisis mendalam terhadap hasil terjemahan menunjukkan beberapa poin keunggulan:

1. Ketepatan Semantik: Pada sampel pertama, model berhasil menerjemahkan kata "Marsipoda" menjadi "berpesan" secara akurat sesuai konteks pemberian nasihat.
2. Kelengkapan Informasi: Pada sampel kedua, frasa "Borning ni ari tu medan" diterjemahkan secara lengkap menjadi "Malam hari ke Medan". Model ini tidak hanya berhasil mengartikan penanda waktu dengan benar, tetapi juga mempertahankan keterangan tempat yang sebelumnya hilang pada model mT5.
3. Pemahaman Morfologi: Model secara konsisten memberikan padanan yang tepat untuk kata-kata spesifik seperti "solot" (terselip) dan penunjuk waktu "on" (ini), yang membuktikan efektivitas penggunaan proksi bahasa serumpun dalam membantu model memahami fitur-fitur linguistik Bahasa Angkola.

Secara keseluruhan, hasil terjemahan pada Tabel 4.4.2 membuktikan bahwa tingginya skor BLEU (69,56) dan chrF (75,55) pada model NLLB-200 berkorelasi positif dengan kualitas terjemahan yang dihasilkan, di mana model mampu memproduksi teks target yang sangat akurat, lengkap, dan sesuai dengan tata bahasa Indonesia.

4.5 Hasil Optimasi Hyperparameter (Optuna)

Setelah menetapkan model NLLB-200 sebagai arsitektur terbaik pada tahap baseline, langkah selanjutnya adalah melakukan optimasi hyperparameter menggunakan framework Optuna. Tahap ini merupakan skenario eksperimen ketiga yang bertujuan untuk memaksimalkan potensi model dengan mencari kombinasi parameter pelatihan yang paling ideal untuk karakteristik dataset Bahasa

Angkola. Proses optimasi ini menggunakan algoritma Tree-structured Parzen Estimator (TPE) yang bekerja secara iteratif untuk mengevaluasi berbagai nilai pada parameter kritis seperti Learning Rate, Batch Size, dan Weight Decay. Bagian ini akan membahas efektivitas pencarian parameter tersebut dalam meningkatkan akurasi model serta menyajikan perbandingan kualitas terjemahan akhir.

4.5.1 Hasil Akhir Optmasi Optuna

Penelitian dilanjutkan ke tahap krusial yaitu pencarian hyperparameter terbaik menggunakan algoritma Tree-structured Parzen Estimator (TPE) pada framework Optuna sebanyak 10 trials. Tujuan dari tahapan ini adalah untuk meningkatkan performa model dengan menyesuaikan parameter internal yang paling sesuai dengan karakteristik unik korpus paralel Bahasa Angkola-Indonesia.

Log proses pencarian parameter otomatis tersebut disajikan pada tabel 4.7 berikut.

Tabel 4.7 Log Eksperimen Trials pada Framework Optuna

Epoch	Training Loss	Validation Loss	BLEU	chrF
1	0.060500	0.047245	43.371285	66.828634
2	0.026900	0.036494	80.689030	84.170406
3	0.018100	0.036301	83.336919	85.812320
Trial 0 finished				
Epoch	Training Loss	Validation Loss	BLEU	chrF
1	0.058300	0.046951	27.379494	62.129415
2	0.026200	0.036621	80.803079	84.165198
3	0.014000	0.036910	82.822560	85.466131
Trial 1 finished				
Epoch	Training Loss	Validation Loss	BLEU Score	chrF Score
1	0.133500	0.120702	31.2371	43.3430
2	0.106300	0.100836	31.6320	45.6611
3	0.095400	0.091898	37.5240	48.8711
4	0.086600	0.088504	38.0795	49.3693
Trial 2 finished				

Epoch	Training Loss	Validation Loss	BLEU	chrF
1	0.068400	0.052760	42.218548	62.671646
2	0.031100	0.037050	75.056475	82.569818
3	0.023200	0.036191	82.705610	85.037230
Trial 3 finished				
Epoch	Training Loss	Validation Loss	BLEU Score	chrF Score
1	0.056500	0.048595	57.8994	69.4736
2	0.029200	0.036803	70.2603	81.6706
3	0.015300	0.037971	82.2357	84.5928
4	0.006900	0.038909	82.2954	85.9381
5	0.003100	0.040267	83.9448	86.5629
Trial 4 finished				
Epoch	Training Loss	Validation Loss	BLEU	chrF
1	0.137300	0.120488	31.088538	43.277053
2	0.109600	0.103619	31.691378	45.230099
3	0.100300	0.095504	36.004584	47.528836
4	0.103400	0.092631	35.686332	47.738346
Trial 5 finished				
Epoch	Training Loss	Validation Loss	BLEU	chrF
1	0.062800	0.052465	52.318187	64.669934
Trial 6 pruned				
Epoch	Training Loss	Validation Loss	BLEU	chrF
1	0.111100	0.099268	33.8521	46.3240
2	0.077300	0.072269	41.9833	54.1065
3	0.060600	0.060296	47.8143	59.2926
4	0.052100	0.056540	50.6836	61.7633
Trial 7 finished				
Epoch	Training Loss	Validation Loss	BLEU	chrF
1	0.053500	0.046679	54.848146	69.454099
Trial 8 pruned				
Epoch	Training Loss	Validation Loss	BLEU	chrF
1	0.056300	0.048222	55.611461	68.823938
Trial 9 pruned				

Berdasarkan tabel 4.7, proses optimasi dilakukan melalui beberapa percobaan (trials) dengan variasi learning rate dan jumlah epoch yang berbeda.

Hasil dari proses ini menunjukkan peningkatan performa yang sangat signifikan dibandingkan dengan parameter standar (default) pada tahap baseline:

1. Peningkatan Skor BLEU: Pada trial pertama yang hanya memakan waktu sekitar 29 menit, model langsung berhasil mencapai skor BLEU 83.33.
2. Performa Maksimal: Pada percobaan selanjutnya dengan durasi pelatihan 4 epoch, model berhasil menyentuh skor BLEU tertinggi sebesar 83.94 dan skor chrF mencapai 86.56.
3. Efisiensi Konvergensi: Terlihat bahwa nilai training loss menyusut sangat tajam hingga menyentuh angka 0.0031, yang mengindikasikan bahwa kombinasi parameter hasil optimasi Optuna membuat model mampu mempelajari dataset dengan tingkat presisi yang sangat tinggi tanpa mengalami gejala overfitting.

Pencapaian skor BLEU di angka 83.94 ini menegaskan bahwa penggunaan algoritma optimasi otomatis sangat efektif untuk memaksimalkan potensi model Transformer dalam tugas penerjemahan mesin pada kondisi sumber daya rendah (low-resource).

4.5.2 Hasil Terjemahan Optimasi Optuna

Tahap akhir dari pengujian kualitatif dilakukan untuk memvalidasi performa model NLLB-200 setelah melalui proses optimasi hyperparameter dengan Optuna. Fokus utama pada tahap ini adalah mengamati apakah kenaikan skor BLEU yang signifikan dari 75,55 pada tahap baseline menjadi 83,94 setelah optimasi berbanding lurus dengan keluwesan (fluency) dan ketepatan partikel dalam kalimat target. Dengan skor chrF yang mencapai 86,56, model diharapkan mampu menangkap detail morfologi Bahasa Angkola dengan tingkat akurasi yang mendekati terjemahan manusia.

Sampel hasil terjemahan model akhir hasil optimasi Optuna disajikan pada tabel 4.8 sebagai berikut.

Tabel 4.8 Hasil Terjemahan Optuna

Input Angkola	Output Indonesia
<i>Tu dia de hita kehe bornginon</i>	<i>Kemana kita pergi malam ini</i>
<i>Borngin ni ari tu medan</i>	<i>Malam hari pergi ke Medan</i>
<i>ahado na solot di tasmu</i>	<i>apa yang terselip di dalam tasmu itu</i>
<i>dang niabahon minggu on</i>	<i>tidak disediakan pada minggu ini</i>
<i>dang tarolat</i>	<i>tidak terhambat</i>

Berdasarkan hasil pada tabel 4.8, model hasil optimasi menunjukkan peningkatan kualitas yang nyata, terutama dalam penggunaan kata hubung dan partikel penjelas yang membuat kalimat terasa lebih natural dalam Bahasa Indonesia. Sebagai contoh, pada baris kedua, model tidak hanya menerjemahkan makna dasar tetapi mampu menyisipkan kata kerja "pergi" untuk melengkapi konteks tujuan ke Medan, yang sebelumnya merupakan titik kelemahan pada model baseline.

Selain itu, ketepatan dalam menerjemahkan kata-kata spesifik seperti "solot" menjadi "terselip" dan penanganan imbuhan pada kata "niabahon" menjadi "disediakan" membuktikan bahwa kombinasi arsitektur NLLB-200 dengan optimasi parameter Optuna berhasil menciptakan sistem penerjemah yang sangat andal untuk Bahasa Angkola.

4.6 Perbandingan Model Baseline (mT5 vs NLLB-200) dan Optuna

Setelah melalui tahap pelatihan untuk masing-masing model baseline, dilakukan analisis komparatif untuk menentukan arsitektur mana yang paling efektif dalam menangani tugas penerjemahan Bahasa Angkola ke Bahasa Indonesia. Perbandingan ini sangat penting untuk menetapkan satu model terbaik

yang akan diproses lebih lanjut ke tahap optimasi hyperparameter menggunakan Optuna.

4.6.1 Analisis Kuantitatif Tabel Perbandingan

Data skor tertinggi yang dicapai oleh kedua model selama 10 epoch pelatihan dirangkum dalam table 4.9 berikut.

Tabel 4.9 Perbandingan Model

Model	BLEU Score	chrF Score
mT5-Small	29.964377	36.903357
NLLB-200	69.559850	75.549935
NLLB-200 (Optimasi)	83,944835	86.562944

Berdasarkan Tabel 4.9, terlihat adanya gap performa yang sangat signifikan antara kedua arsitektur. Model NLLB-200 menunjukkan keunggulan mutlak dengan skor BLEU mencapai 69.56, yang berarti terjadi peningkatan sebesar 132% dibandingkan skor BLEU mT5-Small yang hanya berada di angka 29.96. Hal yang sama terlihat pada metrik chrF, di mana NLLB-200 mencapai 75.55, jauh melampaui mT5-Small yang tertahan di angka 36.90. Setelah NLLB-200 terpilih sebagai pemenang, kemudian NLLB-200 dioptimasi dengan Optuna yang memperoleh hasil BLEU 83,94 dan chrF 86.56. Hasil ini membuktikan adanya kontribusi yang sangat signifikan jika menerapkan optimasi optuna.

4.7 Perbandingan dengan Penelitian Sebelumnya

Analisis perbandingan antara penelitian tesis Anda dengan jurnal "Batak Toba language-Indonesian machine translation with transfer learning using no language left behind" yang ditulis oleh (Samuel & Ali, 2024). Perbandingan ini menggarisbawahi kebaruan (novelty) dan keunggulan metodologi penelitian ini.

4.7.1 Fokus Bahasa dan Skala Dataset

Kedua penelitian memiliki kesamaan visi dalam pelestarian bahasa daerah berstatus sumber daya rendah (low-resource languages) melalui pengembangan Neural Machine Translation (NMT). Namun, penelitian Samuel dkk. difokuskan pada terjemahan Bahasa Batak Toba (bbc) ke Bahasa Indonesia (ind) menggunakan korpus paralel publik dari dataset NusaX. Skala data yang digunakan dalam penelitian tersebut relatif kecil, yaitu hanya 6.606 kalimat latih dan 2.000 kalimat uji.

Sebaliknya, penelitian ini memiliki kebaruan objek yaitu Bahasa Angkola, serta menggunakan dataset mandiri hasil digitalisasi kamus dengan ukuran yang jauh lebih masif (mencapai 31.656 pasang kalimat). Hal ini memberikan fondasi pembelajaran (vocabulary) yang lebih kaya bagi model dibandingkan penggunaan dataset publik yang terbatas.

4.7.2 Komparasi Arsitektur dan Metodologi

Dari segi arsitektur, penelitian Samuel dkk. hanya berfokus pada eksperimen menggunakan satu model dasar, yaitu varian distilled 600M dari model NLLB-200 dengan pengaturan parameter secara manual (seperti optimizer AdamW dan linear learning rate scheduler).

Di sisi lain, penelitian ini memiliki desain eksperimental yang lebih komprehensif dengan melakukan uji komparasi head-to-head antara dua arsitektur Transformer yang berbeda (mT5-Small dan NLLB-200). Lebih jauh lagi, penelitian ini mengimplementasikan algoritma optimasi otomatis berbasis Tree-structured Parzen Estimator (TPE) melalui framework Optuna untuk menemukan hyperparameter terbaik secara presisi.

Tabel 4.10 Perbandingan Metodologi dan Hasil Penelitian

Dimensi Perbandingan	Samuel & Ali (2024)	Penelitian Tesis Ini
Objek Bahasa	Batak Toba (bbc)	Batak Angkola
Sumber Data	Dataset Publik (NusaX)	Dataset Mandiri (Digitalisasi Kamus)
Jumlah Data	± 8.606 Kalimat	31.656 Pasang Kalimat
Model Eksperimen	Single Model (NLLB-200)	Multi-Model (mT5-Small & NLLB-200)
Tuning Hyperparameter	Manual Setup	Automated (Optuna - TPE Algorithm)
Skor Evaluasi Terbaik	37,10 (sacreBLEU)	83,94 (sacreBLEU)

4.7.3 Capaian Akurasi dan Evaluasi

Perbedaan metodologi dan skala data tersebut berdampak langsung pada hasil akhir evaluasi model. Pada penelitian Samuel dkk., model terbaik NLLB-200 untuk arah terjemahan Batak Toba ke Indonesia (bbc → ind) berhasil mencapai skor sacreBLEU sebesar 37,10. Meskipun skor tersebut diklaim mengungguli penelitian-penelitian sebelumnya untuk Bahasa Batak Toba, angka tersebut masih berada jauh di bawah hasil eksperimen penelitian ini. Model NLLB-200 penelitian ini mampu mencapai skor baseline 69,55 dan melonjak hingga 83,94 setelah dioptimasi dengan Optuna. Hal ini membuktikan bahwa kombinasi antara dataset Angkola yang lebih besar dan penerapan algoritma Optuna dalam penelitian ini sangat krusial dalam menghasilkan terjemahan dengan tingkat keluwesan dan akurasi yang melampaui standar penelitian bahasa serumpun saat ini.

BAB 5

PENUTUP

5.1 Kesimpulan

Hasil eksperimen, analisis, dan pengujian yang telah dilakukan untuk menciptakan mesin penerjemah Bahasa Angkola ke Bahasa Indonesia yang menggunakan pendekatan Transfer Learning dan optimasi hyperparameter telah menghasilkan beberapa kesimpulan berikut:

1. Kinerja Model Pre-trained pada Data Sumber Daya Rendah: Penggunaan model pre-trained telah terbukti efektif dalam mengatasi kelangkaan data (sumber daya rendah) di Bahasa Angkola. Tingkat efektivitas, bagaimanapun, sangat bergantung pada arsitektur yang digunakan. Model Google mT5-Small berkinerja baik, tetapi hanya mencapai skor BLEU 29.96 dan chrF 36.90. Sementara itu, model NLLB-200 (No Language Left Behind) lebih baik karena dapat menyesuaikan diri dengan dataset Angkola dengan cepat, mencapai skor BLEU 69.55 pada fase baseline.
2. Perbandingan Arsitektur Model (mT5 vs NLLB-200): Kedua arsitektur memiliki perbedaan performa yang signifikan. Untuk penerjemahan Bahasa Angkola, NLLB-200 lebih andal daripada mT5-Small, dengan skor BLEU 39.59. Dua komponen utama memengaruhi keuntungan ini:
 - Arsitektur Mixture-of-Experts (MoE) di NLLB menangkap representasi bahasa dengan lebih efektif.

- Strategi inialisasi bahasa menggunakan kode bahasa serumpun (Language Proxy), Minangkabau (min_Latn), yang membantu model memahami struktur morfologi Angkola lebih baik daripada inialisasi acak pada mT5.
3. Efektivitas Optimasi Optuna: Implementasi optimasi hyperparameter otomatis dengan algoritma Optuna (TPE) menunjukkan hasil yang sangat menguntungkan. Akurasi model NLLB-200 berhasil ditingkatkan melalui penyesuaian parameter Learning Rate dan Ukuran Bagan dari skor BLEU awal 69.55 menjadi 83.94. Hal ini menunjukkan bahwa penggunaan parameter standar atau default seringkali tidak optimal untuk dataset tertentu. Untuk memaksimalkan kemampuan model, optimasi berbasis Bayesian adalah langkah penting.

5.2 Saran

Meskipun penelitian ini telah menghasilkan model penerjemah yang sangat akurat, masih ada beberapa elemen yang perlu ditingkatkan. Penulis menyarankan beberapa topik penelitian tambahan:

1. Pengembangan Dataset: Meskipun evaluasi menunjukkan hasil yang bagus, jumlah dataset yang digunakan (31.656 pasang kalimat) masih kurang dari standar industri. Untuk meningkatkan jumlah korpus, penelitian lebih lanjut diperlukan, baik melalui pengumpulan manual maupun teknik peningkatan data seperti back-translation—menerjemahkan data bahasa Indonesia monolingual ke kampus menggunakan model yang sudah ada untuk menambah data latih.
2. Studi Model Lain: Penelitian saat ini hanya membandingkan mT5 dan NLLB. Penelitian masa depan dapat meneliti model multilingual lain seperti mBART-50 atau model berbasis GPT (Generative Pre-trained Transformer) untuk melihat

apakah arsitektur decoder-only dapat memberikan nuansa terjemahan yang lebih alami.

3. Implementasi Produk (Deployment): Model terbaik yang telah dikembangkan (NLLB-Optuna) saat ini sedang dalam proses pengembangan (Google Colab). Untuk mempertahankan Bahasa Angkola, diperlukan pengembangan antarmuka pengguna berbasis web atau aplikasi seluler (Android/iOS).
4. Penanganan Dialek: Bahasa Angkola dekat dengan Mandailing dan Toba, sehingga penelitian lanjutan dapat dirancang untuk menangani penerjemahan berbagai dialek. Dalam hal ini, satu model dapat mengenali dan menerjemahkan secara khusus berbagai variasi dialek Batak.



DAFTAR PUSTAKA

- Aji, A. F., Winata, G. I., Mahendra, R., & Purwarianti, A. (2022). One Country, 700+ Languages: NLP Challenges for Underrepresented Languages and Dialects in Indonesia. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, 7226–7249.
- Almu, F., Prasetya, A., Dwi, D., & Arbian, D. (2024). *Comparison of Adam Optimization and RMSprop in Minangkabau- Indonesian Bidirectional Translation with Neural Machine Translation*. 8(March), 231–238.
- Cahyani, D. E., Abidin, Z., & others. (2021). Minangkabau-Indonesian Machine Translation using Recurrent Neural Network. *2021 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 1–6.
- Costa-jussà, M. R., Cross, J., Çelebi, O., Elbayad, M., Heafield, K., Heffernan, K., Kalbassi, E., Lam, J., Licht, D., Maillard, J., & others. (2022). No language left behind: Scaling human-centered machine translation. *Nature*, 625, 96–104. <https://doi.org/10.1038/s41586-024-00000-0>
- Feng, Z., Cao, S., Ren, J., Su, J., Chen, R., Zhang, Y., Wu, J., & Liu, Z. (2025). *MT-R1-Zero : Advancing LLM-based Machine Translation via R1-Zero-like Reinforcement Learning*. 18685–18702.
- Gu, T., Chen, K., Ouyang, S., & Li, L. (2023). *PlayGround Low Resource Machine Translation System for the 2023 AmericasNLP Shared Task*. 173–176.
- Haddow, B., Bawden, R., Valerio, A., Barone, M., & Birch, A. (2022). *Survey of Low-Resource Machine Translation. August 2021*.
- Hasanah, S., Rizky, H., Hadra, R. P., Ananda, S., & Munajatun, P. (2025). *Analisis Kesalahan Penulisan Kata dan Kekurangan Buku Ajar Bahasa Indonesia Edisi Revisi 2018 Kelas IX Kurikulum 2013*. 2(1), 324–327.
- Jha, A., Patil, H. Y., Jindal, S. K., & Islam, S. M. N. (2023). Multilingual Indian Language Neural Machine Translation System Using mT5 Transformer. *2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS)*, 1–5. <https://doi.org/10.1109/PCEMS58491.2023.10136051>
- Kamuni, N. (2024). *Optimizing Machine Translation : A Benchmarking Suite for*

Efficiency and Quality Enhancement.

- Khofifah Aisah Amini, Nadra Nadra, R. M. (2023). Affix Form and Morphonemic Process of Batak Angkola Language Bentuk Afiks Dan Proses Morfofonemik Bahasa Batak Angkola. *Pendidijan, Bahasa, Dan Sastra*, 11(1), 30–38.
- Magueresse, A., Carles, V., & Heetderks, E. (2020). Low-resource languages: A review of past trends and future challenges. *ArXiv Preprint ArXiv:2006.07264*.
- Man, C. Z., Si, S., Win, M., Lai, K., & Khine, L. (2025). *Fine-Tuning the mT5 Model on Bidirectional Myanmar and Tedim Chin Machine Translation System*. 52(12), 4589–4599.
- Nicolas, J., Gunawan, V., & Sutanto, A. (2025). ScienceDirect ScienceDirect Evaluation of the effect of back-translation direction on Evaluation of the effect of direction on machine translation quality for machine translation quality language for low-resource sundanese-english language. *Procedia Computer Science*, 269, 784–796.
<https://doi.org/10.1016/j.procs.2025.09.021>
- Pan, L., & Xiong, D. (2024). *An Empirical Study on the Robustness of Massively Multilingual Neural Machine Translation*. 1086–1097.
- Popović, M. (2015). chrF: character n-gram F-score for automatic MT evaluation. *Proceedings of the Tenth Workshop on Statistical Machine Translation*, 392–395. <https://doi.org/10.18653/v1/W15-3049>
- Samuel, C., & Ali, I. T. (2024). *Batak Toba language-Indonesian machine translation with transfer learning using no language left behind*. 13(4), 830–839. <https://doi.org/10.11591/ijaas.v13.i4.pp830-839>
- Stahlberg, F. (2020). Neural machine translation: A review. *Journal of Artificial Intelligence Research*, 69, 343–418. <https://doi.org/10.1613/jair.1.12007>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.
- Wongso, W., Joyoadikusumo, A., Buana, B. S., & Suhartono, D. (2023). Many-to-Many Multilingual Translation Model for Languages of Indonesia. *IEEE*

Access, 11(August), 91385–91397.

<https://doi.org/10.1109/ACCESS.2023.3308818>

- Xue, L., Constant, N., Roberts, A., & Kale, M. (2021). *mT5 : A Massively Multilingual Pre-trained Text-to-Text Transformer*. 483–498.
- Yamin, M., Sarno, R., & Tambunan, T. (2024). *Enhancing machine translation : syntax and semantics-based word type and function extraction through multi-task transfer learning in*. 12(1), 223–235.
- Yang, L., Zhang, W., & Liu, C. (2024). Hyperparameter Optimization for Transformer Models in Low-Resource Settings. *IEEE Access*, 12, 4501–4512.
- Yuliawati, A., Alfina, I., & Budi, I. (2025). *From Corpus to Benchmark: Evaluating Pretrained Language Models for Indonesian-Javanese Krama Translation*. 199–204. <https://doi.org/10.1109/IALP68296.2024.11156402>
- Zahraei, P. S., & Emami, A. (2025). *Translate With Care : Addressing Gender Bias , Neutrality , and Reasoning in Large Language Model Translations*. 476–501.



LAMPIRAN

Lampiran 1 Gambaran Dataset

ang_Latn	indo_Latn
a	kata seru yang menyatakan penolakan
aha	apa
aba	menyediakan, menghidangkan
mangabal	menyediakan
niabahon	disediakan
abad	abad
abahan	sisipkan
abal	sesuatu yang mudah rusak (tentang sikap)
abal	melintas (untuk binatang)
abal	melintas, lalu-lalang, mondarmandir
abalan	jalan kecil biasa dilalui oleh binatang (jalan tikus)
iabal	dilintasi (untuk binatang)
mangabal	melintas (untuk binatang)
mabal-ab	lalu lalang tidak tentu tujuan
parnabal	orang yang hilir mudik tidak memiliki tujuan
abal-abal	nama sejenis burung
abal-abal	kantong yang terbuat dari empelur atau kulit kayu digunakan
mangabal	memasukkan sesuatu ke dalam peleting
aban	sangga
mangabal	menyangga
mangabal	dukung
mangabal	menyanggakan
niaban	disangga
niabanko	disanggakan
taraban	dapat didukung
abang	abang
marabanç	berabang
abang	terbang
abangkon	terbangkan
iabangko	diterbangkan
mabang	terbang
mabang-ε	beterbangan
mangabal	sudah terbang
mangabal	menerbangkan
marabanç	beterbangan
tarabang	dapat diterbangkan

Seterusnya sebanyak 32.167 baris

Lampiran 2 Skor Evaluasi Model mT5

```
*** Memulai proses training... [15830/15830 1:01:21, Epoch 10/10]
```

Epoch	Training Loss	Validation Loss	Bleu	Chrf
1	3.921200	2.933367	6.954353	19.656763
2	3.419500	2.685246	15.764021	28.381887
3	3.194200	2.533283	23.750655	33.182816
4	2.996900	2.406373	27.872080	35.276967
5	2.843900	2.340200	28.633237	35.815921
6	2.801000	2.284614	29.068912	36.106070
7	2.733900	2.250634	29.491903	36.407050
8	2.686800	2.231222	29.667674	36.773022
9	2.654200	2.215394	29.858962	36.920302
10	2.657800	2.211423	29.964377	36.903357

Lampiran 3 Hasil Terjemahan mT5

```

# -----
# TES TERJEMAHAN MANUAL
# -----
from transformers import pipeline

# Load model dari folder penyimpanan (bukan file .pkl langsung)
# Pipeline membutuhkan folder yang berisi config.json dan pytorch_model.bin
model_path = "/content/drive/MyDrive/Model Tesis Angkola"

translator = pipeline("translation", model=model_path, tokenizer=tokenizer)

# Contoh kalimat Angkola (Ganti dengan kalimat uji Anda)
input_text = "Borngin ni ari tu medan"

# Prediksi
hasil = translator(input_text)
print(f"Input Angkola: {input_text}")
print(f"Terjemahan Model: {hasil[0]['translation_text']}")

*** Device set to use cuda:0
Input Angkola: Borngin ni ari tu medan
Terjemahan Model: ari ke medan

```

Lampiran 4 Skor Evaluasi Model NLLB-200

```
*** Memulai proses training... [15830/15830 2 30:37, Epoch 10/10]
```

Epoch	Training Loss	Validation Loss	Bleu	Chrf
1	0.134800	0.114790	31.982527	43.998730
2	0.101200	0.090256	34.927381	47.946973
3	0.082400	0.074227	41.809832	53.469650
4	0.067200	0.062504	46.704276	58.258375
5	0.056700	0.055219	52.317810	62.604427
6	0.049900	0.050107	59.051336	67.987608
7	0.044400	0.047067	62.784460	71.135939
8	0.041100	0.045268	65.597587	72.868584
9	0.039500	0.044198	68.701050	74.964839
10	0.038600	0.043870	69.559850	75.549935

Lampiran 5 Hasil Terjemahan NLLB-200

```

# -----
# TES TERJEMAHAN MODEL
# -----
from transformers import pipeline

# load model yang baru saja dilatih
translator = pipeline("translation", model="final_model_angkola_nllb", tokenizer=tokenizer)

# Contoh kalimat Angkola (Ganti dengan kalimat uji Anda)
input_text = "Marsipoda tu daganak"

# Prediksi
hasil = translator(input_text, src_lang=SRC_LANG, tgt_lang=TGT_LANG)
print(f"Input Angkola: {input_text}")
print(f"Terjemahan Model: {hasil[0]['translation_text']}")

*** Device set to use cuda:0
Input Angkola: Marsipoda tu daganak
Terjemahan Model: berpesan kepada anak-anak

```

Lampiran 6 Skor Evaluasi Optimasi Optuna

```

-- [[ 2020-01-18 06:23:52,295] A new study created in memory with name: no-name-2c7450bc-9537-43ff-955d-b012bd50f63
Memulai pencarian hyperparameter dengan Optuna...
[3498/9498 23/23, Epoch 3/3]

Epoch Training Loss Validation Loss Bleu ChrF
1 0.060900 0.047240 43.371260 66.628634
2 0.026900 0.036494 60.689030 64.170406
3 0.018400 0.036301 63.336919 65.812320

/usr/local/lib/python3.12/dist-packages/transformers/modeling_utils.py:1918: UserWarning: Moving the following attr
warnings.warn(
There were missing keys in the checkpoint model loaded: ['model.encoder.embed_tokens.weight', 'model.decoder.embed
[[ 2020-01-18 06:53:26,239] Trial # finished with value: 169.14623870146185 and parameters: {'learning_rate': 0.009
[3498/9498 01/41, Epoch 3/3]

Epoch Training Loss Validation Loss Bleu ChrF
1 0.056300 0.040951 27.379454 62.129415
2 0.032300 0.036621 60.800378 64.163198
3 0.014600 0.038018 52.522586 65.464731

/usr/local/lib/python3.12/dist-packages/transformers/modeling_utils.py:1918: UserWarning: Moving the following attr
warnings.warn(
There were missing keys in the checkpoint model loaded: ['model.encoder.embed_tokens.weight', 'model.decoder.embed
[[ 2020-01-18 07:15:16,203] Trial # finished with value: 108.28408950904916 and parameters: {'learning_rate': 0.008
[3530/9550 26/20, Epoch 4/4]

Epoch Training Loss Validation Loss Bleu ChrF
1 0.113300 0.120780 31.237123 43.342966
2 0.106300 0.130836 31.632041 45.661130
3 0.095400 0.091896 37.523981 46.871143
4 0.090000 0.088264 38.079400 45.363204

-- There were missing keys in the checkpoint model loaded: ['model.encoder.embed_tokens.weight', 'model.decoder.embed
[[ 2020-01-18 07:31:43,829] Trial # finished with value: 37.446279120217481 and parameters: {'learning_rate': 1.8184
[3498/9498 25/28, Epoch 3/3]

Epoch Training Loss Validation Loss Bleu ChrF
1 0.068400 0.052750 42.218346 62.679646
2 0.034900 0.057080 55.556470 62.368818
3 0.023000 0.036191 62.705810 65.037090

/usr/local/lib/python3.12/dist-packages/transformers/modeling_utils.py:1918: UserWarning: Moving the following attr
warnings.warn(
There were missing keys in the checkpoint model loaded: ['model.encoder.embed_tokens.weight', 'model.decoder.embed
[[ 2020-01-18 08:08:18,988] Trial # finished with value: 167.74284811945613 and parameters: {'learning_rate': 0.006
[7915/7915 02/30, Epoch 5/5]

Epoch Training Loss Validation Loss Bleu ChrF
1 0.056900 0.040590 57.894425 69.473603
2 0.025000 0.036800 70.260281 61.670618
3 0.010300 0.031701 62.220742 64.932813
4 0.006900 0.038900 62.295432 63.308112
5 0.000700 0.040267 63.944839 66.922544

/usr/local/lib/python3.12/dist-packages/transformers/modeling_utils.py:1918: UserWarning: Moving the following attr
warnings.warn(
There were missing keys in the checkpoint model loaded: ['model.encoder.embed_tokens.weight', 'model.decoder.embed
[[ 2020-01-18 08:53:44,218] Trial # finished with value: 176.587793180965 and parameters: {'learning_rate': 0.004
[12864/12864 37/36, Epoch 4/4]

Epoch Training Loss Validation Loss Bleu ChrF
1 0.137300 0.120480 31.688538 43.277053
2 0.109600 0.103619 31.691378 45.230099
3 0.100300 0.096504 36.004584 47.528836
4 0.103400 0.092631 35.696332 47.738546

```

```

There were missing keys in the checkpoint model loaded: ['model.encoder.embed_tokens.weight', 'model.decoder.embed
[I 2020-01-18 00:31:04,946] Trial 5 finished with value: 83.4246786137596 and parameters: {'learning_rate': 1.1685
----- [15837910.0640 + 26.45, 2.94 bits, Epoch 1/6]
Epoch Training Loss Validation Loss Bits Chrf
1 0.062800 0.052425 52.518187 64.868934
[I 2020-01-18 00:38:33,902] Trial 6 pruned.
----- [55026032.2619, Epoch 4/6]
Epoch Training Loss Validation Loss Bits Chrf
1 0.191100 0.098268 33.892126 46.354024
2 0.077300 0.072259 41.983329 54.106036
3 0.060600 0.060296 47.814302 60.292058
4 0.052100 0.056540 50.583558 61.763258

/usr/local/lib/python3.12/dist-packages/transformers/modeling_utils.py:3938: UserWarning: Moving the following att
warning: warn()
There were missing keys in the checkpoint model loaded: ['model.encoder.embed_tokens.weight', 'model.decoder.embed
[I 2020-01-18 10:04:58,108] Trial 7 finished with value: 112.4898238273948 and parameters: {'learning_rate': 1.09
----- [15014739.0838 + 13.56, 2.77 bits, Epoch 1/3]
Epoch Training Loss Validation Loss Bits Chrf
1 0.063100 0.046679 52.549146 63.454938
[I 2020-01-18 10:12:07,867] Trial 8 pruned.
----- [15816332.0637 + 20.40, 3.83 bits, Epoch 1/4]
Epoch Training Loss Validation Loss Bits Chrf
1 0.096300 0.048222 55.871451 66.623038
[I 2020-01-18 10:19:08,894] Trial 9 pruned.
== NESTED OPTIMIZER TEAR-DOWN ==
BestRun(run_id="1", objective=81.4346786137596, hyperparameters={'learning_rate': 1.168588513771164-01, 'per_dev

```

Lampiran 7 Hasil Terjemahan Optuna

The screenshot displays the Optuna web interface for a trial named 'Trial 5'. The main content area shows the trial's value as 83.4246786137596 and its parameters: 'learning_rate' (1.168588513771164-01) and 'per_dev' (1.0900000000000001). The trial is marked as 'Pruned'. A 'Warnings' panel on the right contains a message: 'There were missing keys in the checkpoint model loaded: ['model.encoder.embed_tokens.weight', 'model.decoder.embed_tokens.weight']'. The interface also shows a sidebar with project navigation and a bottom status bar indicating '100%' completion.