

**PERANCANGAN *NEURAL MACHINE TRANSLATION* BAHASA
INDONESIA-BAHASA JAWA MENGGUNAKAN PENDEKATAN
*SEQUENCE-TO-SEQUENCE***

Skripsi



oleh:

ENDRICH ABEDNEGO
71200542

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2025

**PERANCANGAN NEURAL MACHINE TRANSLATION BAHASA
INDONESIA-BAHASA JAWA MENGGUNAKAN PENDEKATAN
SEQUENCE-TO-SEQUENCE**

Skripsi



Diajukan kepada Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

ENDRICH ABEDNEGO

71200542

**PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA**

2025

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa tugas akhir dengan judul:
**PERANCANGAN NEURAL MACHINE TRANSLATION BAHASA
INDONESIA-BAHASA JAWA MENGGUNAKAN PENDEKATAN
SEQUENCE-TO-SEQUENCE**

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari tugas akhir kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil tugas akhir ini adalah hasil plagiasi atau tiruan dari tugas akhir lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 15 Januari 2025



ENDRICH O ABEDNEGO
71200542

HALAMAN PERSETUJUAN

Judul Tugas Akhir : PERANCANGAN NEURAL MACHINE
TRANSLATION BAHASA INDONESIA-
BAHASA JAWA MENGGUNAKAN
PENDEKATAN SEQUENCE-TO-SEQUENCE

Nama Mahasiswa : ENDRICHO ABEDNEGO

NIM : 71200542

Mata Kuliah : Tugas Akhir

Kode : TI0366

Semester : Ganjil

Tahun Akademik : 2024/2025

Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 15 Januari 2025

Dosen Pembimbing I

Dosen Pembimbing II



Dr. Antonius Rachmat C., S.Kom., M.Cs.



Dr. Phil. Lucia Dwi Krisnawati, S.S.,
M.A.

PERNYATAAN PENYERAHAN KARYA ILMIAH

Saya yang bertanda tangan di bawah ini :

Nama : Endricho Abednego
NIM/NIP/NIDN : 71200542
Program Studi : Informatika
Judul Karya Ilmiah : PERANCANGAN *NEURAL MACHINE TRANSLATION* BAHASA
INDONESIA-BAHASA JAWA MENGGUNAKAN PENDEKATAN
SEQUENCE-TO-SEQUENCE

dengan ini menyatakan:

- a. bahwa karya yang saya serahkan ini merupakan revisi terakhir yang telah disetujui pembimbing/promotor/reviewer.
- b. bahwa karya saya dengan judul di atas adalah asli dan belum pernah diajukan oleh siapa pun untuk mendapatkan gelar akademik baik di Universitas Kristen Duta Wacana maupun di universitas/institusi lain.
- c. bahwa karya saya dengan judul di atas sepenuhnya adalah hasil karya tulis saya sendiri dan bebas dari plagiasi. Karya atau pendapat pihak lain yang digunakan sebagai rujukan dalam naskah ini telah dikutip sesuai dengan kaidah penulisan ilmiah yang berlaku.
- d. bahwa saya bersedia bertanggung jawab dan menerima sanksi sesuai dengan aturan yang berlaku berupa pencabutan gelar akademik jika di kemudian hari didapati bahwa saya melakukan tindakan plagiasi dalam karya saya ini.
- e. bahwa Universitas Kristen Duta Wacana tidak dapat diberi sanksi atau tuntutan hukum atas pelanggaran hak kekayaan intelektual atau jika terjadi pelanggaran lain dalam karya saya ini. Segala tuntutan hukum yang timbul atas pelanggaran dalam karya saya ini akan menjadi tanggung jawab saya pribadi, tanpa melibatkan pihak Universitas Kristen Duta Wacana.
- f. menyerahkan hak bebas royalti noneksklusif kepada Universitas Kristen Duta Wacana, untuk menyimpan, melestarikan, mengalihkan dalam media/format lain, mengelolanya dalam bentuk pangkalan data (*database*), dan mengunggahnya di Repositori UKDW tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis dan pemilik hak cipta atas karya saya di atas, untuk kepentingan akademis dan pengembangan ilmu pengetahuan.
- g. bahwa saya bertanggung jawab menyampaikan secara tertulis kepada Universitas Kristen Duta Wacana jika di kemudian hari terdapat perubahan hak cipta atas karya saya ini.

h. bahwa meskipun telah dilakukan pelestarian sebaik-baiknya, Universitas Kristen Duta Wacana tidak bertanggung jawab atas kehilangan atau kerusakan karya atau metadata selama disimpan di Repositori UKDW.

i. mengajukan agar karya saya ini: *(pilih salah satu)*

- ☐ Dapat diakses tanpa embargo.
- ☒ Dapat diakses setelah 2 tahun.*
- ☐ Embargo permanen.*

Embargo: penutupan sementara akses karya ilmiah.

*Halaman judul, abstrak, dan daftar pustaka tetap wajib dibuka.

Alasan embargo *(bisa lebih dari satu)*:

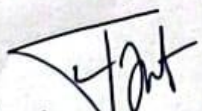
- ☐ dalam proses pengajuan paten.
- ☐ akan dipresentasikan sebagai makalah dalam seminar nasional/internasional.**
- ☒ akan diterbitkan dalam jurnal nasional/internasional.**
- ☐ telah dipresentasikan sebagai makalah dalam seminar nasional/internasional ... dan diterbitkan dalam prosiding pada bulan ... tahun ... dengan DOI/URL ... ***
- ☐ telah diterbitkan dalam jurnal ... dengan DOI/URL artikel ... atau vol./no. ... ***
- ☐ berisi topik sensitif, data perusahaan/pribadi atau informasi yang membahayakan keamanan nasional.
- ☐ berisi materi yang mengandung hak cipta atau hak kekayaan intelektual pihak lain.
- ☐ terikat perjanjian kerahasiaan dengan perusahaan/organisasi lain di luar Universitas Kristen Duta Wacana selama periode tertentu.
- ☐ Lainnya (mohon dijelaskan)

**Setelah diterbitkan, mohon informasikan keterangan publikasinya ke repository@staff.ukdw.ac.id.

***Tuliskan informasi kegiatan atau publikasinya dengan lengkap.

Yogyakarta, 16 Januari 2025

Mengetahui,


Dr. Antonius R.C.

Tanda tangan & nama terang pembimbing
NIDN/NIDK 0523128101

Yang menyatakan,



Endricho Abednego

Tanda tangan & nama terang pemilik karya/penulis
NIM 71200542.

HALAMAN PENGESAHAN

PERANCANGAN NEURAL MACHINE TRANSLATION BAHASA INDONESIA-BAHASA JAWA MENGGUNAKAN PENDEKATAN SEQUENCE-TO-SEQUENCE

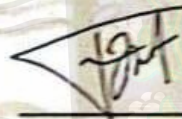
Oleh: ENDRICHO ABEDNEGO / 71200542

Dipertahankan di depan Dewan Penguji Tugas Akhir
Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 16 Desember 2024.

Yogyakarta, 15 Januari 2025
Mengesahkan,

Dewan Penguji:

1. Dr. Antonius Rachmat C., S.Kom.,
M.Cs.



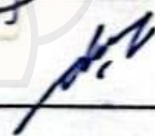
2. Dr. Phil. Lucia Dwi Krisnawati, S.S.,
M.A.



3. Maria Nila Anggia Rini S.T., M.T.I



4. Danny Sebastian, S.Kom., M.M., M.T



(Restyandito, S.Kom., MSIS., Ph.D.)

Ketua Program Studi


(Joko Purwadi, S.Kom., M.Kom.)



Karya sederhana ini dipersembahkan
kepada Tuhan, Keluarga Tercinta,
dan Kedua Orang Tua



“Sebab itu janganlah kamu kuatir akan hari besok, karena hari besok mempunyai kesusahannya sendiri. Kesusahan sehari cukuplah untuk sehari”

(Matius 6:34)

KATA PENGANTAR

Segala puji dan syukur kepada Tuhan yang maha kasih, karena atas segala rahmat, bimbingan, dan bantuan-Nya maka akhirnya Tugas Akhir Skripsi dengan judul PERANCANGAN NEURAL MACHINE TRANSLATION BAHASA INDONESIA-BAHASA JAWA MENGGUNAKAN PENDEKATAN SEQUENCE-TO-SEQUENCE ini telah selesai disusun.

Penulis memperoleh banyak bantuan dari kerja sama baik secara moral maupun spiritual dalam penulisan Tugas Akhir ini, untuk itu tak lupa penulis ucapkan terima kasih yang sebesar-besarnya kepada:

1. Tuhan Yesus Kristus.
2. Orang tua, Basuki dan Dyah Retnani Dwi Arini yang selama ini telah sabar membimbing dan mendoakan penulis tanpa kenal untuk selama-lamanya,
3. Restyandito, S.Kom., MSIS., Ph.D. selaku Dekan FTI, yang saya hormati.
4. Joko Purwadi, S.Kom., M.Kom. selaku Kaprodi Informatika, yang saya hormati
5. Dr. Antonius Rachmat C., S.Kom., M.Cs. selaku Dosen Pembimbing 1, yang telah memberikan ilmunya dan dengan penuh kesabaran membimbing penulis,
6. Dr. Phil. Lucia Dwi Krisnawati, S.S., M.A. selaku Dosen Pembimbing 2 yang telah memberikan ilmu dan kesabaran dalam membimbing penulis,
7. Keluarga tercinta: yang saya kasihi
8. Teman-teman Informatika UKDW 2020 yang telah memberikan dukungan moral selama pengerjaan skripsi
9. Immanuel Joy Perkasa dan Edwin Mahendra yang senantiasa mendampingi dan membantu penulis jika terdapat kesulitan dalam proses pengerjaan skripsi
10. Santi Desanda A. yang selalu mendampingi penulis dalam suka dan duka proses pengerjaan skripsi

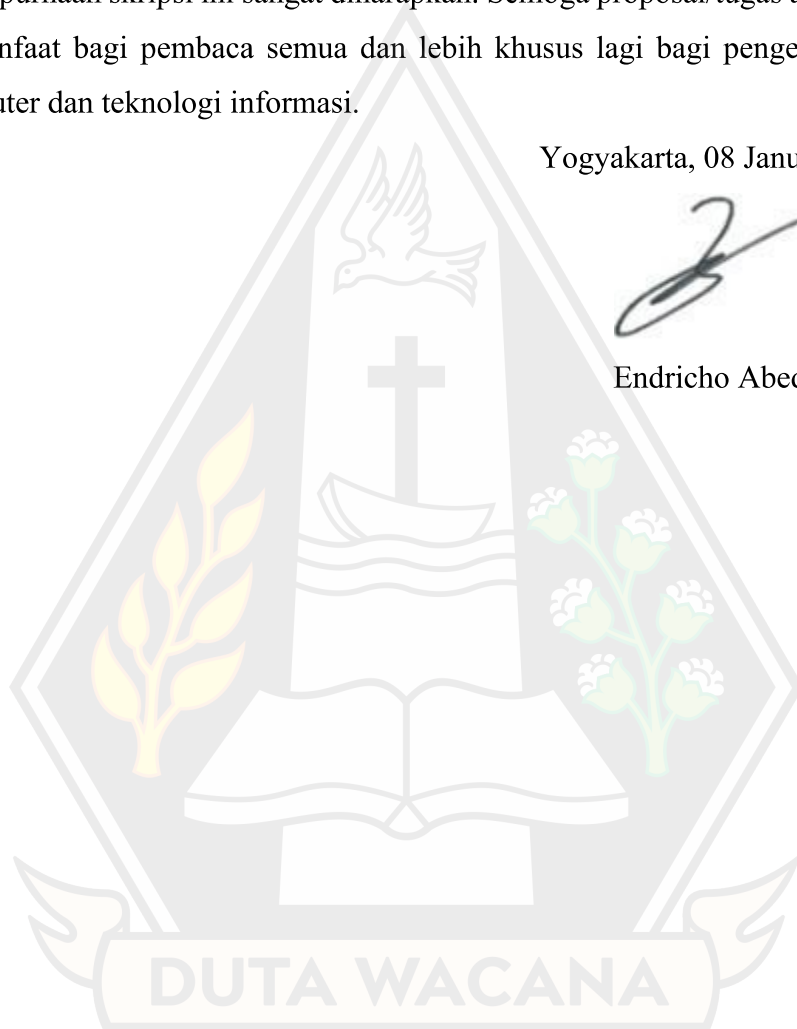
11. Teman-teman GKJ Condongcatur yang telah selalu ada bagi penulis untuk memberikan semangat, hiburan, dan dukungan ketika penulis membutuhkan.

Laporan proposal/tugas akhir ini tentunya tidak lepas dari segala kekurangan dan kelemahan, untuk itu segala kritikan dan saran yang bersifat membangun guna kesempurnaan skripsi ini sangat diharapkan. Semoga proposal/tugas akhir ini dapat bermanfaat bagi pembaca semua dan lebih khusus lagi bagi pengembangan ilmu komputer dan teknologi informasi.

Yogyakarta, 08 Januari 2025



Endricho Abednego



DAFTAR ISI

HALAMAN JUDUL.....	ii
PERNYATAAN KEASLIAN TUGAS AKHIR	iii
HALAMAN PERSETUJUAN	iv
HALAMAN PENGESAHAN.....	v
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR.....	xiv
INTISARI	xv
ABSTRACT	xvi
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah.....	1
1.2. Perumusan Masalah	3
1.3. Batasan Masalah.....	3
1.4. Tujuan Penelitian	4
1.5. Manfaat Penelitian	4
1.6. Metodologi Penelitian	4
1.6.1. Pengumpulan Data	4
1.6.2. Pengembangan Sistem.....	5
1.6.3. Pengujian dan Evaluasi	5
1.7. Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI	7
2.1. Tinjauan Pustaka	7
2.2. Landasan Teori.....	11
2.2.1. <i>Neural Machine Translation</i>	11
2.2.2. <i>Universal Sentence Encoder</i>	12
2.2.3. <i>Sequence-to-sequence</i>	13
2.2.4. <i>Sequence-to-Sequence</i> dengan LSTM.....	14

2.2.5.	<i>Attention Mechanism</i>	16
2.2.6.	Metriks Evaluasi	18
BAB III METODOLOGI PENELITIAN		21
3.1.	Analisis Kebutuhan Sistem	21
3.2.1.	Perangkat Keras	21
3.2.2.	Perangkat Lunak	21
3.2.3.	Pustaka	21
3.2.	Perancangan Penelitian	22
3.2.1.	Studi Literatur	22
3.2.2.	Persiapan Dataset	22
3.2.3.	Pengembangan Sistem	26
3.2.4.	Evaluasi	27
3.2.5.	Arsitektur Model	30
BAB IV IMPLEMENTASI DAN PEMBAHASAN		33
4.1.	Implementasi Awal	33
4.1.2.	Pengumpulan Data	33
4.1.3.	Persiapan Data	34
4.1.5.	Implementasi Word <i>Embedding</i>	40
4.2.	Implementasi Sistem	42
4.2.1.	Konfigurasi 1	42
4.2.2.	Konfigurasi 2	43
4.2.3.	Konfigurasi 3	46
4.2.5.	Inisialisasi model	50
4.2.6.	Pelatihan Model	52
4.2.7.	Pengujian Model	59
BAB V KESIMPULAN DAN SARAN		64

5.1. Kesimpulan	64
5.2. Saran.....	64
DAFTAR PUSTAKA.....	66
LAMPIRAN A.....	68
KODE PROGRAM.....	68
LAMPIRAN B	86
KARTU KONSULTASI	86



DAFTAR TABEL

Tabel 2. 1 Contoh Data Pengujian BLEU <i>Score</i>	18
Tabel 3. 1Contoh hasil Segmentasi Data quad-grams Bahasa Indonesia	23
Tabel 3. 2Contoh hasil segmentasi data 4-grams dalam Bahasa Jawa	23
Tabel 3. 3 Tahapan Persiapan Data.....	25
Tabel 3. 4 Skenario Pengujian Penelitian	28
Tabel 4. 1 Frekuensi munculnya Kata (bahasa Jawa) dalam Dataset	39
Tabel 4. 2 Frekuensi munculnya Kata (bahasa Indonesia) dalam Dataset.....	40
Tabel 4. 3 <i>Encoder</i> model Konfigurasi 1	42
Tabel 4. 4 <i>Decoder</i> model Konfigurasi 1	43
Tabel 4. 5 <i>Decoder</i> Model Konfigurasi 2	45
Tabel 4. 6 Encoder model Konfigurasi 3	46
Tabel 4. 7 Decoder model Konfigurasi 3	47
Tabel 4. 8 <i>Encoder</i> model Konfigurasi 4	48
Tabel 4. 9 <i>Decoder</i> model Konfigurasi 4.....	49
Tabel 4. 10 Pengujian Penerjemahan	62
Tabel 4. 11 Hasil perhitungan Skor BLEU	63

DAFTAR GAMBAR

Gambar 2. 1 <i>Heat Map Skor Sentence Similarity dengan Embedding USE</i>	13
Gambar 2. 2 Arsitektur ConvS2S (Gehring et al., 2017)	14
Gambar 3. 1 Rancangan Arsitektur Model	30
Gambar 4. 1 Contoh Isi dari Buku Materi Bulan Keluarga 2023 (Bahasa Indonesia)	33
Gambar 4. 2 Contoh Isi dari Buku Materi Wulan Brayat 2023 (Bahasa Jawa)	34
Gambar 4. 3 Contoh hasil pengolahan kalimat menjadi <i>quad-grams</i>	38
Gambar 4. 4 Hasil <i>training</i> model pada konfigurasi 1	56
Gambar 4. 5 Hasil <i>training</i> model pada Konfigurasi 2	56
Gambar 4. 6 Hasil <i>training</i> model pada Konfigurasi 3	57
Gambar 4. 7 Hasil <i>training</i> model Konfigurasi 4	58

INTISARI
PERANCANGAN NEURAL MACHINE TRANSLATION BAHASA
INDONESIA-BAHASA JAWA MENGGUNAKAN PENDEKATAN
SEQUENCE-TO-SEQUENCE

Oleh

ENDRICH O ABEDNEGO
71200542

Penelitian ini berfokus pada pengembangan sistem Neural Machine Translation (NMT) untuk penerjemahan dari Bahasa Indonesia ke Bahasa Jawa menggunakan pendekatan sequence-to-sequence berbasis Long Short-Term Memory (LSTM) dan mekanisme attention. Bahasa Jawa, sebagai bahasa dengan sumber daya terbatas, menghadirkan tantangan dalam hal preservasi makna dan kontekstualitas. Penelitian ini menggunakan Universal Sentence Encoder (USE) sebagai model word embedding untuk menangkap makna semantik dari bahasa sumber dan target.

Dataset penelitian diambil dari materi kotbah dan renungan Gereja Kristen Jawa, yang diproses menjadi pasangan bahasa (language pairs) untuk melatih model. Model NMT yang dikembangkan menggunakan arsitektur sequence-to-sequence, dengan LSTM sebagai encoder dan decoder serta mekanisme attention untuk meningkatkan akurasi terjemahan. Evaluasi performa model dilakukan menggunakan metrik BLEU Score untuk mengukur kualitas terjemahan.

Hasil penelitian menunjukkan bahwa model NMT berbasis LSTM dengan USE, berhasil diimplementasikan namun hasil terjemahan masih memiliki perbedaan arti yang cukup signifikan. Skor BLEU yang dicapai menunjukkan potensi pengembangan lebih lanjut, terutama melalui peningkatan kualitas dataset dan penyesuaian arsitektur model untuk bahasa yang memiliki data terbatas.

Kata-kata kunci : *Neural Machine Translation, Sequence-to-Sequence, LSTM, Universal Sentence Encoder, BLEU Score.*

ABSTRACT

DESIGNING AN INDONESIAN-JAVANESE NEURAL MACHINE TRANSLATION USING A SEQUENCE-TO-SEQUENCE APPROACH

By

ENDRICHO ABEDNEGO
71200542

This research focuses on the development of a Neural Machine Translation (NMT) system for translating from Indonesian to Javanese using a sequence-to-sequence approach based on Long Short-Term Memory (LSTM) and an attention mechanism. Javanese, as a low-resource language, presents challenges in preserving meaning and context. This study employs the Universal Sentence Encoder (USE) as a word embedding model to capture the semantic meaning of the source and target languages.

The research dataset was sourced from sermons and devotionals of the Javanese Christian Church and processed into language pairs to train the model. The NMT model utilizes a sequence-to-sequence architecture, with LSTM as the encoder and decoder, along with an attention mechanism to improve translation accuracy. The model's performance was evaluated using the BLEU Score metric to measure translation quality.

The results indicate that while the LSTM-based NMT model with USE was successfully implemented, the translations still exhibit significant differences in meaning. The BLEU Score highlights the potential for further improvements, particularly by enhancing the quality of the dataset and refining the model architecture to better handle low-resource languages.

Keywords : Neural Machine Translation, Sequence-to-Sequence, LSTM, Universal Sentence Encoder, BLEU Score

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Menurut Kamus Besar Bahasa Indonesia (KBBI). (2023) bahasa berarti sistem lambang bunyi yang arbitrer, yang digunakan oleh semua orang atau anggota masyarakat untuk bekerja sama, berinteraksi, dan mengidentifikasi diri dalam bentuk percakapan yang baik, tingkah laku yang baik, sopan santun yang baik. Bahasa juga bisa menjadi bentuk komunikasi untuk menyampaikan suatu tujuan atau maksud tertentu. Salah satu tujuan yang bisa didapatkan melalui bahasa yaitu dalam hal keagamaan, dalam kasus penelitian ini akan berfokus pada aktifitas keagamaan dalam gereja (Kristen). Terdapat berbagai macam jenis gereja yang ada di Indonesia, dan salah satunya adalah gereja etnis atau daerah. Terkhusus di Jawa Tengah dan Daerah Istimewa Yogyakarta, salah satu gereja etnis yang masih sering dijumpai adalah Gereja Kristen Jawa (GKJ).

Salah satu hal yang unik dari kegiatan di GKJ adalah pada setiap minggu tertentu, tata ibadah dari gereja tersebut akan dibawakan dalam Bahasa Jawa. Sebagai contoh, pada GKJ Condongcatur yang berada di Sleman, Yogyakarta, mengadakan Ibadah menggunakan bahasa jawa setiap Ibadah pukul 08.00 WIB setiap minggu ke-3. Mulai dari kotbah, doa, nyanyian, renungan, hingga liturgi akan dibawakan secara penuh menggunakan Bahasa Jawa. Namun, tidak dapat dipungkiri eksistensi Bahasa Jawa di kalangan muda perlahan semakin menghilang. Sebagian besar pemuda hanya mengenali Bahasa Jawa ngoko tanpa mengetahui *undha usuk* (sebuah tingkatan bahasa sebagai bentuk kesopanan terhadap lawan bicara yang diterapkan dalam Bahasa Jawa (Nuryantiningsih & Pandanwangi (2018)). Bahkan, tidak sedikit pemuda yang tidak bisa Bahasa Jawa sama sekali walaupun lahir dan besar di pulau Jawa. Terutama di kota Yogyakarta, yang terkenal dengan istilah kota pelajar dikarenakan banyaknya pendatang dari luar Jawa yang datang ke Yogyakarta untuk menempuh pendidikan. Hal ini tentu saja berdampak pada melunturnya budaya asli Yogyakarta dikarenakan banyaknya budaya luar Yogyakarta yang masuk. Sehingga, masyarakat mencampur -

campurkan budaya yang ada. Menurut Wahyuningsih & Surbakti (2018). Saat ini sering kita dengar istilah ”*wong Jawa ilang Jawane*”, istilah tersebut disematkan kepada masyarakat Jawa yang tidak lagi mengenal atau menggunakan budaya Jawa. Sementara itu, sejalan dengan perkembangan pesat teknologi, kemudahan dalam mengakses informasi semakin meningkat. Teknologi saat ini yang dapat menjembatani komunikasi antar bahasa yang berbeda adalah mesin translasi, dalam hal ini Bahasa Indonesia-Bahasa Jawa.

Perkembangan terbaru dalam teknologi mesin translasi saat ini adalah penggunaan *Neural Machine Translation* (NMT), tetapi pemanfaatan metode ini masih terbilang cukup sedikit dalam penerjemahan bahasa daerah seperti Bahasa Jawa. NMT sendiri merupakan metode penerjemahan otomatis yang memanfaatkan arsitektur *Recurrent Neural Network*. Arsitektur yang paling umum digunakan yaitu *sequence-to-sequence*. Menurut Vaswani dkk. (2018). Terdapat dua komponen utama dalam arsitektur ini, yaitu *encoder* dan *decoder*. *Encoder* berfungsi untuk membaca kalimat sumber secara bertahap kemudian mengubahnya dalam bentuk vektor dengan ukuran yang tetap. Sedangkan untuk *decoder* bertugas untuk mengubah vektor tersebut menjadi kalimat target satu kata pada satu waktu.

Pada penelitian yang dilakukan oleh Jassem & Dwojak (2019) menemukan bahwa evaluator manusia lebih memilih penggunaan NMT dibandingkan *Statistical Machine Translation* (SMT) dikarenakan kelancaran hasil output translasi. Maka dari itu, penelitian ini akan berfokus pada pembuatan sistem *Neural Machine Translation* dengan arsitektur *sequence-to-sequence*. Hasil dari terjemahan model mesin translasi akan dievaluasi menggunakan metrik evaluasi *Bilingual Evaluation Understudy* (BLEU), Papineni dkk. (2002) menjelaskan BLEU pada dasarnya bekerja dengan cara membandingkan terjemahan yang dihasilkan oleh mesin dengan terjemahan yang dihasilkan oleh manusia yang ahli dalam bahasa tersebut.

Dengan demikian, sebuah *machine translation* Bahasa Indonesia ke Bahasa Jawa dapat membantu mempermudah menyelesaikan masalah tersebut. Penelitian ini akan berfokus pada translasi dokumen dengan kategori materi kotbah dan renungan. Mesin akan dilatih menggunakan dataset berupa language pairs yang

didapatkan melalui website resmi Lembaga Pembinaan dan Pengaderan Sinode GKJ & GKI Jateng (LPPS) yaitu <https://lpps.or.id/wp/>.

1.2.Perumusan Masalah

Dengan mempertimbangkan latar belakang yang telah disajikan, masalah penelitian dapat dirumuskan menjadi

1. Bagaimana performa sistem Neural Machine Translation (NMT) dengan pendekatan arsitektur sequence-to-sequence berbasis *Long Short-Term Memory* (LSTM) untuk penerjemahan dari bahasa Indonesia ke bahasa Jawa, berdasarkan evaluasi menggunakan BLEU score sebagai metrik kinerja model?
2. Bagaimana pengaruh penggunaan Universal Sentence Encoder sebagai pengganti word embedding tradisional terhadap performa NMT dengan arsitektur sequence-to-sequence berbasis LSTM?
3. Bagaimana efektivitas penerapan attention mechanism dalam meningkatkan performa sistem NMT berbasis LSTM untuk penerjemahan bahasa Indonesia ke bahasa Jawa?

1.3.Batasan Masalah

Untuk menjaga fokus dalam penelitian ini, beberapa batasan diterapkan dalam pembahasan. Berikut batasan-batasan masalah dari penelitian ini:

1. Penelitian ini merupakan pengembangan mesin translasi, sehingga translasi yang dihasilkan terbatas sesuai pada kategori dataset yang diberikan yaitu materi kotbah gereja. Hasil terjemahan mesin juga akan berlandaskan pada *undha usuk* (ragam bahasa dalam bahasa Jawa) yang ada dalam dataset, yaitu *krama madya*.
2. Proses penerjemahan dilakukan secara satu arah, yaitu dari Bahasa Indonesia ke Bahasa Jawa.
3. Proses penerjemahan dilakukan dalam bentuk frasa dengan penggunaan *quad-gram* sebagai unit analisis bahasa.
4. Perhitungan metrik evaluasi *BLEU* menggunakan perhitungan *quad-gram*.

1.4. Tujuan Penelitian

Berdasarkan masalah yang telah dirumuskan, tujuan penelitian ini adalah

1. Menganalisis performa sistem Neural Machine Translation (NMT) dengan pendekatan arsitektur sequence-to-sequence berbasis LSTM untuk penerjemahan dari bahasa Indonesia ke bahasa Jawa, berdasarkan evaluasi menggunakan BLEU score sebagai metrik kinerja model.
2. Mengkaji pengaruh penggunaan Universal Sentence Encoder sebagai pengganti word embedding tradisional terhadap performa NMT dengan arsitektur sequence-to-sequence berbasis LSTM.
3. Mengevaluasi efektivitas penerapan attention mechanism dalam meningkatkan performa sistem NMT berbasis LSTM untuk penerjemahan bahasa Indonesia ke bahasa Jawa.

1.5. Manfaat Penelitian

Manfaat dari penelitian ini akan berdampak pada GKJ setempat, dengan adanya *machine translation* pembuatan bahan renungan, materi kotbah, ataupun liturgi dalam Bahasa Jawa dapat terlaksana secara lebih cepat. Penelitian ini juga akan bermanfaat bagi jemaat muda untuk memperkenalkan kembali Bahasa Jawa dalam kegiatan gereja.

1.6. Metodologi Penelitian

Penelitian dibagi menjadi 3 bagian, yaitu bagian pengumpulan data, bagian pengembangan, lalu bagian pengujian dan evaluasi.

1.6.1. Pengumpulan Data

Penelitian ini menggunakan dataset dari website resmi Lembaga Pembinaan & Pengaderan Sinode GKJ & GKI Wilayah Jawa Tengah, yaitu <https://lpps.or.id/wp/materi/>. Fokus penelitian adalah penerjemahan materi kotbah dan liturgi gereja dengan 2 dokumen, masing-masing dalam Bahasa Jawa dan Bahasa Indonesia. Dokumen tersebut memiliki 2131 kalimat tanpa preprocessing. *Preprocessing* mencakup *cleaning* dan *filtering* untuk menghapus duplikat, data

kosong, dan karakter tidak dicetak. Normalisasi teks dilakukan untuk menghilangkan tanda baca tidak perlu, spasi ekstra, dan karakter non-alfanumerik. Analisis manual pada tahap *Language Pairs* memastikan jumlah kalimat dalam kedua bahasa sama, dengan dokumen direpresentasikan dalam bentuk file .xlsx. Terakhir, *Word Embedding* digunakan untuk mengonversi teks menjadi vektor numerik dalam ruang dimensi rendah, memungkinkan komputer memahami makna dan hubungan antar kata dalam konteks lebih luas.

1.6.2. Pengembangan Sistem

Penelitian ini mengembangkan model Neural Machine Translation dengan tiga komponen utama: *encoder*, attention mechanism, dan *decoder*. *Encoder* memproses bahasa sumber menggunakan Long Short-Term Memory (LSTM) untuk menghasilkan representasi tersembunyi yang menangkap makna semantik dan konteks dari urutan input. Representasi ini kemudian diteruskan ke attention mechanism, yang bekerja bersama dengan hidden state dari *decoder*. Attention mechanism menghitung attention weight, memungkinkan model untuk fokus pada bagian tertentu dari urutan sumber selama proses decoding. Hasil dari attention mechanism, berupa conditional inputs, digunakan oleh *decoder*, yang juga berbasis LSTM, bersama dengan hidden state-nya untuk menghasilkan urutan terjemahan dalam bahasa target. Teknik LSTM digunakan pada *encoder* dan *decoder* untuk menangani urutan input dan output, dengan tetap mempertahankan mekanisme perhatian sebagai komponen penting.

1.6.3. Pengujian dan Evaluasi

Dalam penelitian ini, evaluasi kinerja model menggunakan metrik *BLEU Score* dengan perhitungan untuk 4 N-grams. Pengujian dilakukan dengan menerjemahkan satu kalimat yang memiliki minimal 4 kata dan tidak terdapat dalam dataset. Kalimat pengujian diambil dari liturgi Gereja Kristen Jawa Condongcatur dalam Bahasa Indonesia, beserta terjemahan Bahasa Jawa yang telah dibuat oleh manusia. Penggunaan kalimat pengujian dari sumber eksternal diperlukan untuk mengukur BLEU Score secara akurat.

1.7.Sistematika Penulisan

Laporan atau proposal skripsi ini disusun dengan sistematika berikut ini.

Bab 1 Pendahuluan yang merupakan pengantar penelitian dimana didalamnya berisi tentang latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, metodologi penelitian serta sistematika penulisan.

Bab 2 Tinjauan Pustaka dan Landasan Teori. Bab ini berisi tinjauan pustaka yang relevan dengan penelitian, mencakup kajian literatur yang telah dilakukan di bidang *Neural Machine Translation* terkhusus pada bagian *Sequence-to-Sequence* sebagai referensi. Selanjutnya, bab ini juga menguraikan landasan teori yang digunakan sebagai dasar penelitian.

BAB 3 adalah Metodologi Penelitian. Bab ini berisi metode penelitian yang digunakan, termasuk teknik pengumpulan data, tahapan pra-pemrosesan data, metode pengembangan model, dan pendekatan evaluasi yang diterapkan. Bab ini memberikan detail mengenai bagaimana penelitian ini dilaksanakan dari awal hingga akhir.

BAB 4 Implementasi dan Pembahasan. Bab ini berisi implementasi model yang telah dikembangkan dan pembahasan hasil yang diperoleh. Bab ini menjelaskan bagaimana model *sequence-to-sequence* diimplementasikan dan diintegrasikan ke dalam mesin penerjemah Bahasa Indonesia-Bahasa Jawa.

BAB 5 Kesimpulan dan Saran. Bab ini berisi ringkasan dari keseluruhan penelitian, termasuk kesimpulan utama yang dapat diambil dari hasil penelitian. Selain itu, akan disampaikan juga saran untuk penelitian selanjutnya atau pengembangan lebih lanjut dari sistem yang telah dibuat, terutama terkait dengan peningkatan kinerja model.

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Penelitian ini telah berhasil mengimplementasikan sistem *neural machine translation* (NMT) untuk penerjemahan Bahasa Indonesia-Bahasa Jawa melalui penerapan LSTM dengan berbagai macam konfigurasi *sequence-to-sequence*. Pendekatan yang diterapkan dalam penelitian ini secara khusus berfokus pada pemanfaatan quadgram sebagai unit pelatihan. Berdasarkan perhitungan skor BLEU yang digunakan sebagai indikator kinerja model, keseluruhan konfigurasi menunjukkan nilai yang belum memuaskan. Hal ini disebabkan oleh keterbatasan dataset yang hanya menggunakan quadgram sebagai unit pelatihan. Selain itu, faktor lain yang turut memengaruhi adalah evaluasi validasi oleh manusia yang hanya dilakukan pada level kalimat, sementara pada tingkat quadgram tidak dilakukan validasi manual untuk memastikan kesesuaian makna setiap pasangan quadgram. Model NMT dengan pendekatan *sequence-to-sequence* berbasis LSTM yang memanfaatkan *traditional embedding* dan dilengkapi dengan komponen *attention mechanism* memberikan hasil terbaik dibandingkan model lain, dengan skor BLEU sebesar 0,0522.

5.2. Saran

Untuk tahap selanjutnya, saran dari penelitian ini adalah sebagai berikut :

1. Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi kombinasi berbagai n-gram, mulai dari *unigram* hingga *quadgram* untuk menganalisis pengaruhnya terhadap kualitas hasil terjemahan.
2. Pengumpulan dataset yang lebih kompleks, beragam, dan tervalidasi oleh evaluasi manusia sangat direkomendasikan, termasuk variasi dialek serta tema teks yang lebih bervariasi.

3. Melakukan eksperimen dengan arsitektur model yang berbeda dan membandingkan performa hasilnya juga menjadi langkah penting untuk pengembangan lebih lanjut.



DAFTAR PUSTAKA

- Abidin, Z., Sucipto, A., Budiman, A., Teknik, F., & Komputer, D. I. (2018). PENERJEMAHAN KALIMAT BAHASA LAMPUNG-INDONESIA DENGAN PENDEKATAN NEURAL MACHINE TRANSLATION BERBASIS ATTENTION TRANSLATION OF SENTENCE LAMPUNG-INDONESIAN LANGUAGES WITH NEURAL MACHINE TRANSLATION ATTENTION BASED APPROACH. *Jurnal KELITBANGAN*, 6(2), 191–205. <https://translate.google.com>.
- Asudani, D. S., Nagwani, N. K., & Singh, P. (2023). Impact of word embedding models on text analytics in deep learning environment: a review. *Artificial Intelligence Review*, 56(9), 10345–10425. <https://doi.org/10.1007/s10462-023-10419-1>
- Bahdanau, D., Cho, K., & Bengio, Y. (2015). *Neural Machine Translation by Jointly Learning to Align and Translate*. <http://arxiv.org/abs/1409.0473>
- Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., John, R. St., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y.-H., Strophe, B., & Kurzweil, R. (2018). *Universal Sentence Encoder*. <http://arxiv.org/abs/1803.11175>
- Fauziyah, Y., Ilyas, R., Kasyidi, F., Achmad Yani, J., Sains dan Informatika, F., Jenderal Achmad Yani, U., & Terusan Sudirman, J. (2022). *MESIN PENTERJEMAH BAHASA INDONESIA-BAHASA SUNDA MENGGUNAKAN RECURRENT NEURAL NETWORKS* (Vol. 16, Issue 2). <https://ejurnal.teknokrat.ac.id/index.php/teknoinfo/index>
- Jassem, K., & Dwojak, T. (2019). Statistical versus neural machine translation - a case study for a medium size domain-specific bilingual corpus. *Poznan Studies in Contemporary Linguistics*, 55(2), 491–515. <https://doi.org/10.1515/psicl-2019-0018>
- Kamus Besar Bahasa Indonesia (KBBI)*. (2023). Badan Pengembangan dan Pembinaan Bahasa, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia. <https://kbbi.kemdikbud.go.id/entri/bahasa>

- Niu, Z., Zhong, G., & Yu, H. (2021). A review on the attention mechanism of deep learning. *Neurocomputing*, 452, 48–62.
<https://doi.org/10.1016/j.neucom.2021.03.091>
- Nuryantiningih, F., & Dyah Pandanwangi, W. (2018). *Politeness and Impoliteness in Javanese Speech Levels*.
- Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). *BLEU: a Method for Automatic Evaluation of Machine Translation*.
- Puspitaningrum, D. (2021). A Study of English-Indonesian Neural Machine Translation with Attention (Seq2Seq, ConvSeq2Seq, RNN, and MHA): A Comparative Study of NMT on English-Indonesian. *ACM International Conference Proceeding Series*, 271–280.
<https://doi.org/10.1145/3479645.3479703>
- Reiter, E. (2018). *A Structured Review of the Validity of BLEU*.
<https://doi.org/10.1162/COLI>
- Sakinah, M. A., Ramadhan, T. I., & Hartono, R. (2024). Neural Machine Translation Untuk Bahasa Sunda Loma-Sunda Halus Menggunakan Long Short Term Memory. *Jurnal Komputer Antartika*, 2, 26–34.
<https://ejournal.mediaantartika.id/index.php/jka>
- Sutskever, I., Vinyals, O., & Le, Q. V. (2014a). *Sequence to Sequence Learning with Neural Networks*.
- Sutskever, I., Vinyals, O., & Le, Q. V. (2014b). *Sequence to Sequence Learning with Neural Networks*.
- Vaswani, A., Bengio, S., Brevdo, E., Chollet, F., Gomez, A. N., Gouws, S., Jones, L., Kaiser, Ł., Kalchbrenner, N., Parmar, N., Sepassi, R., Shazeer, N., & Uszkoreit, J. (2018). *Tensor2Tensor for Neural Machine Translation*.
<http://arxiv.org/abs/1803.07416>
- Wahyuningsih, S. K., & Surbakti, A. H. (2018). PERUBAHAN BAHASA DAN BUDAYA JAWA DI YOGYAKARTA PADA ERA GLOBALISASI. In *Jurnal As-Salam* (Vol. 2, Issue 1).