

**IMPLEMENTASI TRANSFORMER MODEL
UNTUK SISTEM PENERJEMAHAN
LOW RESOURCE LANGUAGE
BAHASA KAWI KE BAHASA INDONESIA**

SKRIPSI

Diajukan Kepada

Universitas Pendidikan Ganesha

Untuk memenuhi salah satu persyaratan dalam Menyelesaikan

Program Sarjana Program Studi Sistem Informasi

Oleh

I PUTU ARSANA PUTRA

NIM. 1815091016

PROGRAM STUDI SISTEM INFORMASI

JURUSAN TEKNIK INFORMATIKA

FAKULTAS TEKNIK DAN KEJURUAN

UNIVERSITAS PENDIDIKAN GANESHA

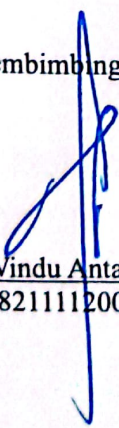
SINGARAJA

2025

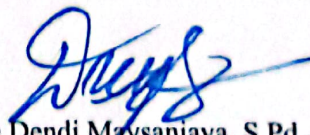
SKRIPSI
DIAJUKAN UNTUK MELENGKAPI TUGAS DAN
MEMENUHI SYARAT-SYARAT UNTUK MENCAPAI
GELAR SARJANA PENDIDIKAN

Menyetujui

Pembimbing I,

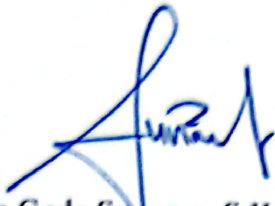

Made Windu Antara Kesiman, S.T., M.Sc., Ph.D.
NIP. 198211112008121001

Pembimbing II,


Ir. I Made Dendi Maysanjaya, S.Pd., M.Eng.
NIP. 199005152019031008

Skripsi oleh I Putu Arsana Putra
telah dipertahankan di depan dewan penguji
pada tanggal 04 Juni 2025

Dewan Penguji,



Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs

NIP. 198307252008011008



Dr. Putu Hendra Suputra, S.Kom., M.Cs.

NIP. 198212222006041001



Made Windu Antara Kesiman, S.T., M.Sc., Ph.D.

NIP. 198211112008121001



Ir. I Made Dendi Maysanjaya, S.Pd., M.Eng.

NIP. 199005152019031008

Diterima oleh Panitia Ujian Fakultas Teknik dan Kejuruan
Universitas Pendidikan Ganesha
guna memenuhi syarat-syarat untuk mencapai Gelar Sarjana Komputer

Pada :

Hari : Selasa

Tanggal : 24 JUN 2025



Mengetahui

Ketua Ujian

Made Windu Antara Kesiman, S.T., M.Sc., Ph.D.
NIP. 198211112008121001

Sekretaris Ujian

Ir. I Made Dendi Maysanjaya, S.Pd., M.Eng.
NIP. 199005152019031008

Mengesahkan,
Dekan Fakultas Teknik dan Kejuruan

Prof. Dr. Kadek Rihendra Dantes, S.T., M.T.
NIP. 197912012006041001

PERNYATAAN KEASLIAN TULISAN

Dengan ini saya menyatakan bahwa skripsi saya dengan judul “Implementasi Transformer Model Untuk Sistem Penerjemahan *Low Resource Language* Bahasa Kawi Ke Bahasa Indonesia” berserta seluruh isinya adalah benar karya saya sendiri, dan saya tidak melakukan penjiplakan dan pengutipan dengan cara-cara tidak sesuai dengan etika yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini, saya siap menanggung risiko atau sanksi yang dijatuhkan kepada saya apabila kemudian ditemukan pelanggaran atas etika keilmuan dalam skripsi saya ini atau ada klaim terhadap keaslian skripsi saya ini.

Singaraja, 04 Juni 2025

Yang membuat pernyataan



I Putu Arsana Putra

NIM 1815091016

PRAKATA

Puji syukur penyusun panjatkan ke hadapan Tuhan Yang Maha Esa karena berkat rahmatnyalah, penulis dapat menyelesaikan skripsi yang berjudul ” **Implementasi Transformer Model Untuk Sistem Penerjemahan *Low Resource Language* Bahasa Kawi Ke Bahasa Indonesia**”. Skripsi ini disusun guna memenuhi persyaratan mencapai gelar sarjana komputer pada Universitas Pendidikan Ganesha.

Dalam menyelesaikan skripsi ini, penulis banyak mendapat bantuan dan dukungan baik berupa moral maupun material dari berbagai pihak. Untuk itu, dalam kesempatan ini, penulis mengucapkan terima kasih kepada:

1. Prof. Dr. Kadek Rihendra Dantes, S.T., M.T. selaku Dekan Fakultas Teknik dan Jurusan atas motivasi dan fasilitas yang diberikan sehingga penulis bisa menyelesaikan studi sesuai dengan rencana.
2. Ir. I Made Ardwi Pradnyana, S.T., M.T. selaku Pembimbing Akademik yang telah memberikan arahan, bimbingan, petunjuk dan motivasi kepada penulis dalam menyelesaikan skripsi.
3. Made Windu Antara Kesiman, S.T., M.Sc., Ph.D. selaku Pembimbing I yang telah memberikan bimbingan, arahan, petunjuk dan motivasi penulis dalam penyusunan skripsi ini.
4. Ir. I Made Dendi Maysanjaya, S.Pd., M.Eng. selaku Pembimbing II yang telah memberikan bimbingan, arahan, petunjuk dan motivasi kepada penulis dalam penyusunan skripsi ini.
5. Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs. selaku Penguji I yang telah memberikan saran, arahan, serta motivasi dalam penyusunan skripsi ini.
6. Dr. Putu Hendra Suputra, S.Kom., M.Cs. selaku Penguji II yang telah memberikan saran, arahan, serta motivasi dalam penyusunan skripsi ini.
7. Grup Riset *Virtual Vision Image and Pattern* (VVIP) yang memberi dukungan, saran dan arahan dalam penyusunan skripsi ini.
8. Keluarga Besar yang selalu memberikan semangat, doa dan motivasi untuk menyelesaikan skripsi ini.

9. Semua pihak yang tidak bisa disebutkan satu per satu yang telah membantu penyusunan skripsi ini.

Penulis menyadari sepenuhnya bahwa isi dari skripsi ini masih jauh dari kata sempurna karena keterbatasan ilmu dan kemampuan yang dimiliki. Untuk itu, kritik dan saran yang bersifat membangun dari berbagai pihak sangat diharapkan. Penulis juga berharap skripsi ini dapat bermanfaat bagi kita semua khususnya dalam pengembangan sistem menuju ke arah digitalisasi.

Singaraja 04 Juni 2025

Penulis



DAFTAR ISI

HALAMAN SAMPUL	i
LEMBAR LOGO	ii
HALAMAN JUDUL.....	iii
LEMBAR PERSETUJUAN PEMBIMBING	iv
LEMBAR PERSETUJUAN DOSEN PENGUJI	v
LEMBAR PERSETUJUAN DAN PENGESAHAN PANITIA UJIAN	vi
PERNYATAAN KEASLIAN TULISAN	vii
PRAKATA.....	viii
ABSTRAK.....	x
DAFTAR ISI	xi
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Tujuan Penelitian	5
1.4. Pembatasan Masalah	5
1.5. Manfaat Penelitian.....	5
BAB II KAJIAN TEORI.....	7
2.1. Studi Sebelumnya.....	7
2.2. Landasan Teori	18
2.2.1. <i>Low Resource Language</i>	18
2.2.2. Bahasa Kawi	18
2.2.3. Dataset Bahasa Kawi ke Bahasa Indonesia	20
2.2.4. Neural Machine Translation.....	22
2.2.5. Pre Processing pada Bahasa.....	23

2.2.6. Transformer.....	25
2.2.7. Evaluasi Model	31
BAB III METODOLOGI PENELITIAN.....	35
3.1. Jenis Penelitian	35
3.2. Alur Penelitian	35
3.2.1. Sumber Data	36
3.2.2. Pre Processing Data	37
3.2.3. Pembangunan Model	40
3.2.4. Proses Training	41
3.2.5. Evaluasi Model	42
BAB IV HASIL DAN PEMBAHASAN.....	51
4.1. Implementasi	51
4.1.1. Perangkat Lunak Pendukung Penelitian	51
4.1.2. Pengumpulan Data.....	51
4.1.3. Proses Pembersihan Data.....	51
4.1.4. Modeling Transformer.....	64
4.1.5. Training Model	70
4.1.6. Evaluasi Model	72
4.2. Hasil dan Pembahasan.....	76
4.2.1. Skema tokenisasi dengan tipe <i>Word Level</i>	76
4.2.2. Skema tokenisasi dengan tipe <i>Subword Level</i>	90
4.2.3. Perbandingan dengan Studi Sebelumnya.....	105
BAB V PENUTUP	108
5.1. Kesimpulan.....	108
5.2. Saran	109
DAFTAR PUSTAKA.....	111
LAMPIRAN.....	117

DAFTAR TABEL

Tabel 1.1 Skor BLEU setiap model dalam penerjemahan bahasa Jerman dan Prancis ke bahasa Inggris.....	3
Tabel 2.1 Skor BLUE tertinggi masing-masing <i>Arabic Vernacular</i> (Baniata dkk., 2021)	8
Tabel 2.2 Skor <i>Human Assesment</i> masing-masing <i>Arabic Vernacular</i> (Baniata dkk., 2021)	8
Tabel 2.3 Skor BLEU setiap model di penelitian (Primandhika dkk., 2021).....	10
Tabel 2.4 Skor BLEU untuk penelitian (Nyoni & Bassett, 2021).....	11
Tabel 2.5 BLEU <i>Score</i> masing-masing model NMT (Budaya dkk., 2022b).	12
Tabel 2.6 Skor Automatic Assesment setiap model pada (Lankford dkk., 2022). 13	
Tabel 2.7 <i>Human Evaluation</i> pada penelitian (Lankford dkk., 2022).....	14
Tabel 2.8 Skor BLEU masing-masing model (Nongbri dkk., 2023)	15
Tabel 2.9 Skor BLEU masing-masing model pada penelitian (Wangchuk dkk., 2023)	16
Tabel 2.10 <i>Automatic Evaluation</i> Model Translasi Bahasa Makassar ke Indonesia	17
Tabel 2.11 <i>Automatic Evaluation</i> Model Translasi Bahasa Indonesia ke Makassar	17
Tabel 2.12 Sampel Data Bahasa Kawi ke Bahasa Indonesia	21
Tabel 2.13 Interpretasi dari skor BLEU (<i>Evaluate Models Cloud Translation</i> , t.t.)	33
Tabel 3.1 Sampel Data Bahasa Kawi ke Bahasa Indonesia	36
Tabel 3.2 Ilustrasi Proses Data Cleaning.....	37
Tabel 3.3 Proses Tokenisasi	37
Tabel 3.4 Kamus Kata masing-masing Bahasa	38
Tabel 3.5 Penambahan <i>Special Token</i>	38
Tabel 3.6 Kamus Kata dengan <i>Special Token</i>	39
Tabel 3.7 Mask Matrix untuk Masked Multi-head Attention.....	41
Tabel 3.8 Representasi Feature (X) dan Target (Y).....	42
Tabel 3.9 Sampel Candidate Language dan Reference Language	43
Tabel 3.10 Proses pembentukan grams	43
Tabel 3.11 Perhitungan Count.....	44
Tabel 3.12 Penjumlahan terhadap gram yang sama pada 1-gram.....	44
Tabel 3.13 Perhitungan Max Ref Count.....	45
Tabel 3.14 Proses perhitungan fungsi <i>Countclip</i>	46
Tabel 3.15 Perhitungan Count pada 2-gram.....	46
Tabel 3.16 Penjumlahan <i>Count</i> pada gram yang sama di 2-gram.....	47
Tabel 3.17 Perhitungan Max Ref Count pada 2-gram.....	47
Tabel 3.18 Clip count pada 2-gram	48
Tabel 3.19 Nilai Precision Score pada setiap gram.....	48
Tabel 3.20 Perhitungan nilai c dan r.....	49

Tabel 3.21 Perhitungan Summation terhadap <i>Log Precision Score</i> dan <i>Weight</i> ..	49
Tabel 4.1 Hasil eksekusi tahap penghapusan elemen tidak benar.....	53
Tabel 4.2 Daftar kalimat dengan kemiripan sejenis	54
Tabel 4.3 Hasil eksekusi dari proses <i>case folding</i>	55
Tabel 4.4 Hasil eksekusi proses <i>remove punctuation</i>	56
Tabel 4.5 Kamus Kata pada Word Level Tokenization.....	59
Tabel 4.6 Proses Tokenisasi pada Word Level Tokenization	59
Tabel 4.7 Kamus kata pada Subword Level Tokenization	61
Tabel 4.8 Proses tokenisasi dari Subword Level Tokenizer	62
Tabel 4.9. Total Token pada <i>Word Level</i> dengan parameter minimum frekuensi.	76
Tabel 4.10. Parameter Model Transformer.....	77
Tabel 4.11. Skor Evaluasi <i>Word Level Greedy Search</i>	77
Tabel 4.12. Skor Evaluasi <i>Word Level Beam Search</i> panjang Beam 2	77
Tabel 4.13. Skor Evaluasi <i>Word Level Beam Search</i> panjang Beam 3	78
Tabel 4.14 Prediksi Model dan Skor BLEU terbaik di setiap <i>record</i>	79
Tabel 4.15. Prediksi Model dan Skor BLEU terkecil di setiap <i>record</i>	86
Tabel 4.16. Total Token pada <i>Subword Level</i> dengan parameter minimum frekuensinya	90
Tabel 4.17. Parameter Model Transformer	90
Tabel 4.18. Skor Evaluasi Greedy Search.....	91
Tabel 4.19. Skor Evaluasi Beam Search Panjang Beam 2	92
Tabel 4.20. Skor Evaluasi Beam Search panjang Beam 3	92
Tabel 4.21. Prediksi Model dan Skor BLEU terbaik di setiap <i>record</i>	93
Tabel 4.22. Prediksi Model dan Skor BLEU terkecil di setiap <i>record</i>	101
Tabel 4.23 Skor BLEU Studi Sebelumnya.....	106
Tabel 4.24 Perbandingan Hasil Terjemahan.....	106



DAFTAR GAMBAR

Gambar 2.1 Proses segmentasi paragraf menjadi kalimat (Budaya dkk., 2022a).	21
Gambar 2.2 Model Encoder dan Decoder (Tan dkk., 2020)	23
Gambar 2.3 Model Transformer (Vaswani dkk., 2017)	26
Gambar 2.4 Scaled Dot Product Attention (Vaswani, et.al 2017)	29
Gambar 3.1 Flowchart Tahapan Penelitian	35
Gambar 4.1 Proses <i>Import</i> data	51
Gambar 4.2 Proses eliminasi data kosong	52
Gambar 4.3 Proses penghapusan elemen tidak benar	52
Gambar 4.4 Proses penghapusan data terduplikasi	53
Gambar 4.5 Proses <i>case folding</i>	54
Gambar 4.6 Proses <i>remove punctuation</i>	55
Gambar 4.7 Proses pembagian data	57
Gambar 4.8 Proses penggabungan dan penyimpanan data train dan data test	57
Gambar 4.9 Kode Word Level Tokenization	58
Gambar 4.10 Kode Subword Level Tokenization	61
Gambar 4.11 Pembuatan DataLoader	63
Gambar 4.12 Definisi DataLoader pada data train dan data test	64
Gambar 4.13 Pembentukan <i>Positional Encoding</i>	65
Gambar 4.14 Pembentukan <i>Positional Embedding</i>	65
Gambar 4.15 Deklarasi <i>Embedding</i>	65
Gambar 4.16 Class Transformer Base	66
Gambar 4.17 Pendefinisian Model Transformer	67
Gambar 4.18 Fungsi Encode pada Transformer Base	67
Gambar 4.19 Fungsi Decode pada Transformer Base	68
Gambar 4.20 <i>Causal mask</i> dengan panjang <i>sequence</i> 6	68
Gambar 4.21 Classifier Transformer	68
Gambar 4.22 Fungsi Forward pada <i>Class Transformer Base</i>	69
Gambar 4.23 Arsitektur Model	69
Gambar 4.24. Transformer Lightning Module	70
Gambar 4.25. Step Train Model	71
Gambar 4.26 Pendefinisian fungsi untuk menghitung nilai <i>loss mean</i>	71
Gambar 4.27. Parameter Training model	72
Gambar 4.28. Properti pada Local Search	72
Gambar 4.29. Fungsi Greedy Decode	73
Gambar 4.30 Fungsi Beam Search	74
Gambar 4.31. Proses Penerjemahan	75
Gambar 4.32. Contoh Hasil Skor BLEU	75

DAFTAR LAMPIRAN

Lampiran 1. Kurva Loss Transformer <i>Word Level Tokenization</i>	118
Lampiran 2. Kurva Loss Transformer <i>Subword Level Tokenization</i>	120

