

EVALUASI MULTI-DIMENSI MODEL TRANSFORMER UNTUK ANALISIS SENTIMEN BAHASA BALI: PERFORMA, TOKENISASI, DAN INTERPRETABILITAS

TESIS



**PROGRAM STUDI ILMU KOMPUTER (S2)
PROGRAM PASCASARJANA
UNIVERSITAS PENDIDIKAN GANESHA**



- UU ITE No. 11 Tahun 2008 Pasal 5 Ayat 1 "Informasi Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"
- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSrE - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSrE
- Cetak dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.



- UU ITE No. 11 Tahun 2008 Pasal 5 Ayat 1 "Informasi Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"
- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSrE - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSrE
- Cetak dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.



Balai
Sertifikasi
Elektronik

TESIS

DIAJUKAN UNTUK MELENGKAPI TUGAS DAN MEMENUHI SYARAT-SYARAT UNTUK MENCAPAI GELAR MAGISTER KOMPUTER

Menyetujui

Pembimbing I	Dr. I Gede Aris Gunadi, S.Si. M.Kom. NIP.197703182008121004
Pembimbing II	Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs. NIP.198307252008011008



- UU ITE No. 11 Tahun 2008 Pasal 5 Ayat 1 "Informasi Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"
- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSrE - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSrE
- Cetakn dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.

Tesis oleh I Komang Deny Supanji telah dipertahankan di depan tim penguji dan dinyatakan diterima sebagai salah satu persyaratan untuk memperoleh gelar Magister Komputer di Ilmu Komputer (S2), Program Pascasarjana, Universitas Pendidikan Ganesha
Pada tanggal 03 Agustus 2026

Dewan Penguji

Ketua	Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs. NIP.198307252008011008
Anggota	Dr. I Gede Aris Gunadi, S.Si. M.Kom. NIP.197703182008121004
Anggota	Prof. Drs. Sariyasa, M.Sc., Ph.D. NIP.196406151989021001
Anggota	Prof. Dr. Gede Suweken, M.Sc. NIP.196111111987021001

Mengetahui Direktur Program Pascasarjana
Universitas Pendidikan Ganesha,



Prof. Dr. I Nyoman Jampel, M.Pd.

NIP.195910101986031003



- UU ITE No. 11 Tahun 2008 Pasal 5 Ayat 1 "Informasi Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"
- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSrE - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSrE
- Cetakan dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.

LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya bahwa tesis yang saya susun sebagai syarat untuk memperoleh gelar Magister Ilmu Komputer dari Program Pascasarjana Universitas Pendidikan Ganesha seluruhnya merupakan hasil karya saya sendiri. Bagian-bagian tertentu dalam penulisan tesis yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas dan sesuai dengan norma, kaidah, serta etika akademis.

Apabila dikemudian hari ditemukan seluruh atau sebagian tesis ini bukan hasil karya saya sendiri atau adanya plagiat dalam bagian-bagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sandang dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku di wilayah Negara Kesatuan Republik Indonesia.

Singaraja, 3 Agustus 2026

Yang memberi pernyataan,



I Komang Deny Supanji

PRAKATA

Puji dan syukur Penulis panjatkan ke hadapan Tuhan Yang Maha Esa atas segala anugerah-Nya, sehingga tesis yang berjudul "Evaluasi Multi-Dimensi Model Transformer untuk Analisis Sentimen Bahasa Bali: Performa, Tokenisasi, dan Interpretabilitas" dapat diselesaikan dengan baik.

Dalam penyusunan tesis ini, Penulis menemui berbagai hambatan dan kesulitan. Namun, berkat dukungan dan bantuan dari berbagai pihak, hambatan tersebut dapat diatasi. Oleh karena itu, pada kesempatan ini Penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Dr. I Gede Aris Gunadi, S.Si., M.Kom., selaku Dosen Pembimbing I yang dengan sabar membimbing, mengarahkan, dan memberikan motivasi yang bermakna, sehingga Penulis mampu melewati berbagai halangan dalam perjalanan studi dan penyelesaian tesis ini.
2. Dr. I Made Gede Sunarya, S.Kom., M.Cs., selaku Koordinator Program Studi Ilmu Komputer Universitas Pendidikan Ganesha sekaligus Dosen Pembimbing II yang dengan sabar membimbing, mengarahkan, dan memberikan motivasi yang bermakna, sehingga Penulis mampu menyelesaikan tesis ini.
3. Prof. Drs. Sariyasa, M.Sc., Ph.D., dan Prof. Dr. Gede Suweken, M.Sc., selaku penguji yang telah memberikan banyak masukan bermanfaat untuk penyempurnaan tesis ini.

4. Direktur Program Pascasarjana Universitas Pendidikan Ganesha, yang telah memberikan pengarahan serta petunjuk hingga terselesaikannya tesis ini.
5. Rektor Universitas Pendidikan Ganesha, yang telah memberikan dukungan moril dan memfasilitasi berbagai kepentingan studi selama Penulis menempuh pendidikan di Program Pascasarjana Universitas Pendidikan Ganesha.
6. Bapak dan Ibu dosen beserta staf, yang telah membagikan ilmu kepada Penulis selama mengikuti pendidikan di Program Pascasarjana Ilmu Komputer Universitas Pendidikan Ganesha.
7. Istri tercinta, Ida Ayu Gede Agus Mahadewi, S.Ag., M.Pd., serta Ibu, Bapak, saudara, dan teman-teman yang telah memberikan semangat, motivasi, serta dukungan moril dan materiil dalam penulisan tesis ini.

Akhir kata, Penulis mengucapkan terima kasih kepada semua pihak yang telah membantu penyusunan tesis ini. Penulis berharap semoga tesis ini dapat bermanfaat bagi kita semua dan menjadi bahan masukan bagi dunia pendidikan ke depannya.

Singaraja, 3 Agustus 2026

Penulis

DAFTAR ISI

PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN.....	xiv
ABSTRAK.....	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah.....	9
1.3 Pembatasan Masalah.....	10
1.4 Rumusan Masalah.....	11
1.5 Tujuan Penelitian.....	11
1.6 Manfaat Penelitian.....	12
1.6.1 Manfaat teoritis.....	12
1.6.2 Manfaat Praktis.....	13
BAB II LANDASAN TEORI.....	14
2.1 Kajian Teori.....	14
2.1.1 Analisis Sentimen.....	14
2.1.2 Pemrosesan Bahasa Alami (Natural Language Processing).....	15
2.1.3 Arsitektur Transformer.....	15
2.1.4 BERT (Bidirectional Encoder Representations from Transformers).....	17
2.1.5 IndoBERT.....	18
2.1.6 Multilingual BERT (mBERT).....	19
2.1.7 XLM-RoBERTa.....	19
2.1.8 Tokenisasi pada Model <i>Pre-trained</i>	20
2.1.9 Explainable Artificial Intelligence (XAI) dan SHAP.....	24
2.1.10 Metrik Evaluasi Klasifikasi.....	25
2.1.11 <i>McNemar's Test</i>	26

2.1.12	Bahasa Bali sebagai <i>Low-Resource Language</i>	27
2.1.13	Dataset NusaX.....	28
2.2	Kajian Penelitian yang Relevan.....	32
2.3	Kerangka Berpikir	41
2.4	Hipotesis dan Proposi Penelitian	45
BAB III METODE PENELITIAN.....		47
3.1	Metode Penelitian	47
3.1.1	Persiapan Penelitian	49
3.1.2	Pengumpulan Data	51
3.1.3	Preprocessing Data.....	52
3.1.4	<i>Training</i>	63
3.1.5	Evaluasi	68
3.1.6	Interpretasi Model.....	75
3.1.7	Integrasi Hasil dan Analisis	78
3.1.8	Tahap Kesimpulan dan Saran.....	79
3.2	Lingkungan Eksperimen.....	80
BAB IV HASIL DAN PEMBAHASAN.....		82
4.1	Hasil Prapemrosesan Data.....	82
4.1.1	Prapemrosesan <i>Dataset</i> Sentimen.....	82
4.1.2	Prapemrosesan Leksikon Bahasa Bali.....	83
4.2	Hasil Pelatihan Model (<i>Training</i>).....	83
4.2.1	Konvergensi Pelatihan IndoBERT	84
4.2.2	Konvergensi Pelatihan mBERT	86
4.2.3	Konvergensi Pelatihan XLM-RoBERTa.....	88
4.3	Dimensi 1: Hasil Evaluasi Performa Klasifikasi	90
4.3.1	Perbandingan Performa Antar Model	91
4.3.2	Analisis <i>Classification Report</i> per Kelas	92
4.3.3	Hasil Uji Signifikansi McNemar.....	96
4.4	Dimensi 2: Hasil Audit <i>Tokenizer</i> Berbasis Leksikon.....	98
4.4.1	Ringkasan Metrik Audit <i>Tokenizer</i>	98
4.4.2	Analisis Fertility rate.....	101

4.4.3 Analisis Perfect match rate dan OOV rate	102
4.4.4 Analisis Contoh Tokenisasi Kata Bahasa Bali	103
4.5 Dimensi 3: Hasil Interpretasi Model dengan SHAP	107
4.5.1 SHAP Global: <i>Mean importance</i> per Model	107
4.5.2 SHAP Global: <i>Mean importance</i> per Kelas	109
4.5.3 SHAP Global: Analisis Top Token per Model	110
4.5.4 SHAP Lokal: Analisis Berdasarkan Pola Kesepakatan Antar Model	112
4.6 Integrasi Multi-Dimensi dan Analisis Korelasi	127
4.6.1 Tabel Integrasi Multi-Dimensi	127
4.6.2 Analisis Korelasi Pearson Antar Metrik	128
4.6.3 Profil Komparatif Holistik Ketiga Model	132
4.7 Pembahasan Umum	133
BAB V PENUTUP	136
5.1 Rangkuman	136
5.2 Kesimpulan	138
5.2.1 Prosedur Analisis Sentimen Bahasa Bali Menggunakan Model Transformer	139
5.2.2 Perbandingan Performa dan Hubungan Tokenisasi dengan Performa Klasifikasi	140
5.2.3 Interpretasi SHAP dan Kerangka Evaluasi Multi-Dimensi	141
5.3 Implikasi Penelitian	142
5.3.1 Implikasi Teoritis	142
5.3.2 Implikasi Praktis	144
5.4 Keterbatasan Penelitian	145
5.5 Saran	147
5.5.1 Saran untuk Penelitian Lanjutan	147
5.5.2 Saran untuk Pengembangan Teknologi dan Kebijakan	149
5.6 Penutup	150
DAFTAR PUSTAKA	152
LAMPIRAN	158

DAFTAR TABEL

Tabel 2.1 Tabel Kontingensi 2x2	27
Tabel 2.2 Sampel Kalimat dan label Sentimen pada Dataset NusaX Bahasa Bali 30	
Tabel 2.3 Sampel Kata Leksikon Bahasa Bali	31
Tabel 2.4 Penelitian yang Relevan	33
Tabel 3.1 Distribusi Label pada masing-masing Subset	51
Tabel 3.2 Hasil Proses Case folding	53
Tabel 3.3 Hasil Proses Cleaning	55
Tabel 3.4 Hasil Proses Encoding	59
Tabel 3.5 Hasil Prapemrosesan Leksikon Bahasa Bali	62
Tabel 3.6 Konfigurasi Hyperparameter <i>Fine-tuning</i>	65
Tabel 3.7 Struktur <i>Confusion Matrix</i> Tiga Kelas Sentimen	69
Tabel 3.8 Contoh Perhitungan <i>Fertility rate</i> (Hipotetis)	72
Tabel 3.9 Spesifikasi Lingkungan Eksperimen	80
Tabel 4.1 Riwayat Pelatihan IndoBERT per <i>Epoch</i>	85
Tabel 4.2 Riwayat Pelatihan mBERT per <i>Epoch</i>	87
Tabel 4.3 Riwayat Pelatihan XLM-RoBERTa per <i>Epoch</i>	89
Tabel 4.4 Ringkasan Performa Ketiga Model pada Test Set	91
Tabel 4.5 <i>Classification Report</i> Lengkap Ketiga Model	92
Tabel 4.6 Hasil Uji McNemar Antar Pasangan Model	96
Tabel 4.7 Ringkasan Metrik Audit <i>Tokenizer</i> Berbasis Leksikon Bahasa Bali (N=830 Kata Unik)	98
Tabel 4.8 Contoh Hasil Tokenisasi Kata Bahasa Bali dengan Fragmentasi Tertinggi	105
Tabel 4.9 Statistik Mean <i>Absolute SHAP</i> Global per Model (N=90 Sampel)	107
Tabel 4. 10 Mean Absolute <i>SHAP</i> per Kelas Sentimen (30 sampel per kelas) ...	109
Tabel 4.11 Tabel Integrasi Multi-Dimensi: Performa, Tokenisasi, dan Interpretabilitas	127

DAFTAR GAMBAR

Gambar 2.1 Arsitektur Transformer.....	16
Gambar 2.2 Arsitektur BERT: Prapelatihan dan Fine-tuning.....	17
Gambar 2.3 Perbandingan Tokenisasi WordPiece vs SentencePiece (Diadaptasi dari Kudo & Richardson, (2018); Rust dkk., (2021)).	22
Gambar 2.4 Kerangka Berpikir	42
Gambar 3.1 Metode Penelitian.....	48
Gambar 3.2 Distribusi Label Dataset NusaX Bahasa Bali per Subset (Train, Valid, Test).....	52
Gambar 3.3 Distribusi Panjang Token Dataset NusaX Bahasa Bali (XLM-R Baseline).....	61
Gambar 4.1 <i>Loss Curve</i> Pelatihan IndoBERT: <i>Training Loss</i> (dicatat per interval <i>logging</i>) dan <i>Validation Loss</i> (per <i>epoch</i>).....	85
Gambar 4.2 <i>Loss Curve</i> Pelatihan mBERT: <i>Training Loss</i> (dicatat per interval <i>logging</i>) dan <i>Validation Loss</i> (per <i>epoch</i>).....	87
Gambar 4.3 <i>Loss Curve</i> Pelatihan XLM-RoBERTa: <i>Training Loss</i> (dicatat per interval <i>logging</i>) dan <i>Validation Loss</i> (per <i>epoch</i>)	89
Gambar 4.4 Confusion Matrix IndoBERT pada Test Set (N=400)	93
Gambar 4.5 Confusion Matrix mBERT pada Test Set (N=400).....	94
Gambar 4.6 Confusion Matrix XLM-RoBERTa pada Test Set (N=400)	95
Gambar 4.7 Perbandingan Tiga Metrik Audit <i>Tokenizer</i> : <i>Perfect match rate</i> , <i>OOV rate</i> , dan <i>Fertility rate</i>	99
Gambar 4.8 Contoh Kata Bahasa Bali dengan Fragmentasi Tertinggi pada Ketiga <i>Tokenizer</i>	104
Gambar 4.9 Distribusi Mean <i>Absolute SHAP</i> per Sampel Ketiga Model (N=90 Sampel <i>Stratified</i>)	108
Gambar 4.10 Mean <i>Absolute SHAP</i> per Kelas Sentimen Ketiga Model (30 Sampel per Kelas).....	110
Gambar 4.11 Waterfall Plot SHAP IndoBERT, Tipe 1 <i>All Correct</i> , Kelas <i>Positive</i>	114
Gambar 4.12 Waterfall Plot SHAP IndoBERT, Tipe 2 <i>Indobert Correct Only</i> , Kelas <i>Neutral</i>	116
Gambar 4.13 Waterfall Plot SHAP mBERT, Tipe 2 <i>Indobert Correct Only</i> , Kelas <i>Neutral</i>	117
Gambar 4.14 Waterfall Plot SHAP XLM-RoBERTa, Tipe 2 <i>Indobert Correct Only</i> , Kelas <i>Neutral</i>	119
Gambar 4.15 Waterfall Plot SHAP IndoBERT, Tipe 3 <i>Indobert Incorrect</i> , Kelas <i>Neutral</i>	121
Gambar 4.16 Waterfall Plot SHAP IndoBERT, Tipe 4 <i>All Incorrect</i> , Kelas <i>Positive</i>	123
Gambar 4.17 Waterfall Plot SHAP mBERT, Tipe 4 <i>All Incorrect</i> , Kelas <i>Positive</i>	124
Gambar 4.18 Waterfall Plot SHAP XLM-RoBERTa, Tipe 4 <i>All Incorrect</i> , Kelas <i>Positive</i>	126

Gambar 4.19 Heatmap Korelasi Pearson: Performa vs Tokenisasi vs XAI (F1-Macro, Fertility rate, Perfect Match %, OOV rate %, Mean SHAP Importance)	129
Gambar 4.20 Radar Chart Profil Komparatif Multi-Dimensi Ketiga Model (F1, Perfect match rate, dan XAI Importance)	130



DAFTAR LAMPIRAN

Lampiran 1. Hasil Tokenisasi Lengkap 830 Kata Unik Leksikon Bahasa Bali pada Ketiga Model.....	158
Lampiran 2. <i>Waterfall Plot</i> SHAP Lokal	173
Lampiran 3. Riwayat Hidup Penulis	222

