

**IMPLEMENTASI NORMALISASI TEKS
INTEGRATIF BERBASIS LEKSIKON DAN WORD
EMBEDDING FASTTEXT PADA PEMODELAN
TOPIK ISU INDONESIA GELAP DI PLATFORM X**



**PROGRAM STUDI SISTEM INFORMASI (SI) (S1)
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KEJURUAN
UNIVERSITAS PENDIDIKAN GANESHA**

SKRIPSI

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



Menyetujui

Pembimbing I	I Nyoman Saputra Wahyu Wijaya, S.Kom., M.Cs. NIP.198910262019031004
Pembimbing II	Ir. I Made Dendi Maysanjaya, S.Pd., M.Eng. NIP.199005152019031008



- 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.

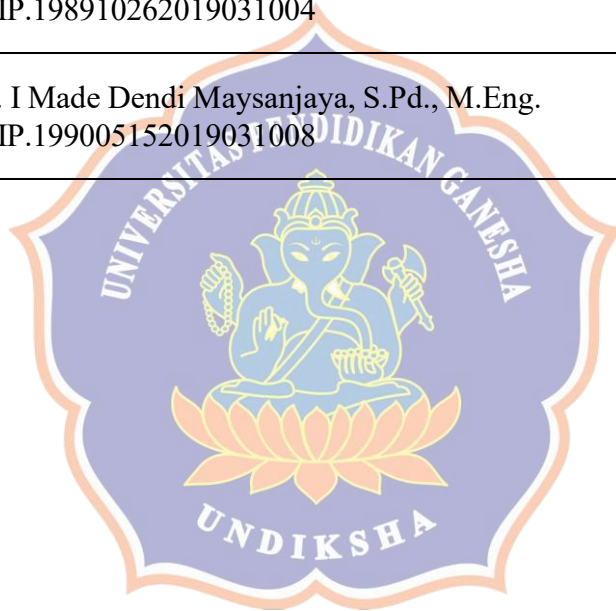


**Balai Besar
Sertifikasi
Elektronik**

Skripsi oleh FRANSESKO INDRAJID ini
telah dipertahankan di depan dewan penguji
Pada tanggal 06 Januari 2026

Dewan Penguji

Ketua	Ir. I Gusti Ayu Agung Diatri Indradewi, S.Kom., M.T. NIP.198907112020122004
Anggota	Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs. NIP.198307252008011008
Anggota	I Nyoman Saputra Wahyu Wijaya, S.Kom., M.Cs. NIP.198910262019031004
Anggota	Ir. I Made Dendi Maysanjaya, S.Pd., M.Eng. NIP.199005152019031008



- 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.

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

Menyetujui

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



- 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.

PERNYATAAN

Dengan ini saya menyatakan bahwa skripsi yang berjudul "*Implementasi Normalisasi Teke Integratif Berbasis Leksikon dan IFord Embedding FastText pada Pemodelan Topik Im ilIndonesiaGelap di Platform X*" beserta seluruh isinya adalah benar-benar karya saya sendiri. Saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini, atau jika ada klaim dari pihak lain terhadap keaslian karya saya ini.

Singaraja, 23 Januari 2026
Yang membuat pernyataan,



The image shows a handwritten signature in blue ink over a red rectangular stamp. The stamp contains the Garuda Pancasila emblem, the text 'METERAL TEMPEL', and the alphanumeric code 'F7B0ANX 19 710102 27'.

Fransesko Indrajid
NIM. 2115091048

PRAKATA

Puji dan syukur Penulis panjatkan ke hadirat Tuhan Yesus Kristus atas segala kasih karunia, anugerah, dan penyertaan-Nya yang tiada henti. Hanya karena kebaikan dan kekuatan dari-Nya, Penulis dimampukan untuk melewati setiap proses perkuliahan hingga akhirnya dapat menyelesaikan penyusunan skripsi yang berjudul ***“Implementasi Normalisasi Teks Integratif Berbasis Leksikon dan Word Embedding FastText pada Pemodelan Topik Isu Indonesia Gelap di Platform X”***. Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan Program Sarjana Komputer pada Jurusan Teknik Informatika, Fakultas Teknik dan Kejuruan, Universitas Pendidikan Ganesha.

Penulis menyadari sepenuhnya bahwa penyusunan skripsi ini tidak akan terwujud tanpa adanya doa, bantuan, bimbingan, dan dukungan dari berbagai pihak. Segala tantangan dan hambatan yang dihadapi dapat teratasi berkat semangat serta motivasi yang diberikan. Oleh karena itu, pada kesempatan ini dengan segala kerendahan hati, Penulis ingin menyampaikan ucapan terima kasih yang setulus-tulusnya kepada:

1. Kedua orang tua hebat yang Tuhan berikan kepada Penulis, Bapak Petrus Matangkin dan Ibu Agustina Liling. Tidak ada deretan kata yang cukup indah dan tidak ada lembaran kertas yang cukup luas untuk melukiskan besarnya pengorbanan, tetesan keringat, dan air mata yang telah Papa dan Mama curahkan. Terima kasih telah menjadi rumah tempat Penulis pulang, yang selalu menyebut nama Penulis paling setia dalam doa, dan menjadi alasan terkuat Penulis untuk tidak menyerah. Skripsi ini adalah persembahan kecil yang tak sebanding dengan kasih sayang tanpa batas yang kalian berikan. Terima kasih telah menjadi orang tua terbaik bagi Penulis.
2. Saudara-saudara kandung Penulis, yaitu kakak Charles yang senantiasa menjadi mentor sekaligus tempat mencari pertolongan, adik Herbert Valem yang selalu hadir sebagai teman berbagi, dan adik kecil cantik Alverinne Anoria yang keceriaannya selalu menjadi sumber semangat bagi Penulis selama menempuh perkuliahan. Semoga Tuhan senantiasa memuluskan jalan pekerjaan Kakak dan menerangi setiap langkah pendidikan adik-adik menuju masa depan yang lebih cerah.

3. Bapak I Nyoman Saputra Wahyu Wijaya, S.Kom., M.Cs., selaku Pembimbing Utama, dan Bapak Ir. I Made Dendi Maysanjaya, S.Pd., M.Eng., selaku Pembimbing Pendamping. Penulis mengucapkan terima kasih yang sebesar-besarnya atas segala waktu, tenaga, dan pikiran yang telah diluangkan untuk membimbing, mengarahkan, serta memberikan masukan-masukan sejak awal penyusunan proposal hingga selesainya skripsi ini.
4. Ibu Ir. I Gusti Ayu Agung Diatri Indradewi, S.Kom., M.T. dan Bapak Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs., selaku Dosen Penguji. Penulis menyampaikan terima kasih yang setulusnya atas ketersediaan waktu untuk menguji, serta atas segala kritik, saran, dan masukan konstruktif yang diberikan. Arahan dari Bapak dan Ibu sangat berperan penting dalam penyempurnaan skripsi ini sehingga menjadi karya ilmiah yang lebih baik.
5. Teman-teman yang sudah Penulis anggap seperti saudara kandung sendiri, teman-teman *Cornerfam*, yaitu Agas, Allan, Nanda, Jovin, Farhil, Augusto, Lanang, Alit, Zidan, Bintang, Turah, Gabriel, Bagas, Jo, Dapa, Jordan, dan Reno. Terima kasih yang tak terhingga karena telah membuka pintu persahabatan selebar-lebarnya bagi Penulis yang datang dari jauh. Terima kasih telah merangkul dan menerima segala kekurangan Penulis, menjadi tempat berbagi tawa dan cerita, serta mengajak Penulis belajar, bermain, dan bertumbuh bersama. Kehadiran kalian adalah alasan mengapa tanah rantau ini terasa seperti rumah sendiri. Sangat bersyukur bisa mengenal dan berbagi masa perkuliahan ini dengan kalian.

Akhir kata, Penulis menyadari sepenuhnya bahwa skripsi ini masih jauh dari kesempurnaan karena keterbatasan pengetahuan dan pengalaman yang Penulis miliki. Oleh karena itu, dengan segala kerendahan hati, Penulis sangat mengharapkan kritik dan saran yang membangun dari semua pihak demi perbaikan di masa mendatang.

Singaraja, 22 Januari 2026

Penulis

DAFTAR ISI

PERNYATAAN.....	i
PRAKATA	ii
ABSTRAK	iv
<i>ABSTRACT</i>	v
DAFTAR ISI	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian.....	4
1.4 Ruang Lingkup Penelitian.....	4
1.5 Manfaat Penelitian.....	5
BAB II KAJIAN PUSTAKA DAN LANDASAN TEORI.....	7
2.1 Kajian Pustaka.....	7
2.2 Landasan Teori	12
2.2.1 <i>Text Preprocessing</i>	12
2.2.2 <i>Slang Lexicon</i>	14
2.2.3 <i>FastText</i>	17
2.2.4 Cosine Similarity	18
2.2.5 Bag Of Words (BoW).....	19
2.2.6 <i>Latent Dirichlet Allocation</i>	19
2.2.7 <i>Topic Coherence</i>	22

2.2.8	<i>Perplexity</i>	23
2.2.9	Intepretasi dan Evaluasi Hasil	24
BAB III METODE PENELITIAN		27
3.1	Alur Penelitian.....	27
BAB IV HASIL DAN PEMBAHASAN		47
4.1	Hasil Penelitian.....	47
4.1.1	<i>Slang Lexicon Construction</i>	47
4.1.2	<i>Data Preparation</i>	51
4.1.3	<i>Topic Modeling</i>	60
4.2	Pembahasan	78
4.2.1	Implementasi Normalisasi Teks Berbasis Leksikon dan <i>FastText</i> Pada Pemodelan Topik	78
4.2.2	Pemodelan Topik Isu IndonesiaGelap Dengan <i>Latent Dirichlet Allocation</i> 83	
BAB V PENUTUP		90
5.1	Kesimpulan.....	90
5.2	Saran.....	91
DAFTAR PUSTAKA.....		92
LAMPIRAN		101

DAFTAR TABEL

Tabel 2. 1 Penelitian Terkait Normalisasi Teks	7
Tabel 2. 2 Kategori Variasi Kata Tidak Baku Indonesia Colloquial Lexicon.....	15
Tabel 2. 3 Kategori Transformasi Kata Tidak Baku IndoCollex	16
Tabel 3. 1 Contoh Hasil Crawling	32
Tabel 3. 2 Contoh Tahapan Case Folding	32
Tabel 3. 3 Contoh Hasil Tahapan Cleaning Data.....	33
Tabel 3. 4 Contoh Hasil Tahapan Tokenization.....	34
Tabel 3. 5 Contoh Hasil Tahapan Other Language Removal	34
Tabel 3. 6 Contoh Hasil Normalisasi Teks	37
Tabel 3. 7 Contoh Hasil Stemming.....	37
Tabel 3. 8 Contoh Hasil Stop Word Removal	38
Tabel 3. 9 Contoh Hasil Tahapan pembobotan Bag of Words	40
Tabel 3. 10 Perhitungan Cosine Similarity.....	41
Tabel 3. 11 Simulasi Perhitungan Perplexity.....	42
Tabel 3. 12 Contoh Hasil Interpretasi Topik	43
Tabel 4. 1 Colloquial Indonesian Lexicon.....	47
Tabel 4. 2 IndoCollex	48
Tabel 4. 3 Merged Lexicon.....	49
Tabel 4. 4 Slang yang Memiliki Padanan Formal Lebih dari Satu.....	51
Tabel 4. 5 Daftar Kata Ambigu.....	51
Tabel 4. 6 Hasil Tahapan Case Folding	52
Tabel 4. 7 Dataset Hasil Data Cleaning.....	53
Tabel 4. 8 Hasil Tahapan Tokenization	54
Tabel 4. 9 Hasil Tahapan Other Language Removal	55
Tabel 4. 10 Hasil Manual Text Normalization.....	55
Tabel 4. 11 20 Kata Teratas Yang Dinormalisasi Dengan Leksikon.....	56
Tabel 4. 12 20 Kata Teratas Yang Dinormalisasi Dengan FastText	57
Tabel 4. 13 Hasil Tahapan Normalization	58
Tabel 4. 14 Hasil Tahapan Stemming	58
Tabel 4. 15 Hasil Tahapan Stop Word Removal.....	59

Tabel 4. 16 Contoh Hasil Pembobotan Bag of Words.....	60
Tabel 4. 17 Coherence score Untuk Setiap Topik Dengan Normalisasi Kamus Indonesian Colloquial Lexicon	62
Tabel 4. 18 Coherence score Untuk Setiap Topik Dengan Normalisasi Kamus IndoCollex	63
Tabel 4. 19 Coherence score Untuk Setiap Topik Pada Normalisasi Leksikon Gabungan.....	65
Tabel 4. 20 Coherence score Untuk Setiap Topik Pada Normalisasi Leksikon.....	66
Tabel 4. 21 Hasil Topic Coherence score (Cv) untuk Setiap K.....	68
Tabel 4. 22 Hasil Evaluasi Perplexity Skenario Leksikon.....	69
Tabel 4. 23 Hasil Evaluasi Perplexity Skenario FastText	69
Tabel 4. 24 Hasil Evaluasi Perplexity Skenario Normalisasi Integratif	70
Tabel 4. 25 Daftar Kata Kunci Teratas untuk Setiap Topik.....	72
Tabel 4. 26 Hasil Interpretasi dan Pelabelan Topik oleh Ahli	77
Tabel 4. 27 Hasil Uji Evaluasi Kualitatif oleh Responden	78
Tabel 4. 28 Perbandingan Skor Coherence (Cv) dengan Penelitian Terdahulu....	87



DAFTAR GAMBAR

Gambar 3. 1 Alur Utama Penelitian	27
Gambar 3. 2 Tahapan Assets Preparation	29
Gambar 3. 3 Alur Data Preprocessing	31
Gambar 3. 4 Tahapan Topic Modeling	39
Gambar 3. 5 Tahapan Word Intrusion.....	45
Gambar 3. 6 Tahapan Topic Intrusion	46
Gambar 4. 1 Coherence Score Pemodelan Topik Dengan Normalisasi Colloquial Indonesian Lexicon	61
Gambar 4. 2 Coherence Score Pemodelan Topik Dengan Normalisasi IndoCollex Lexicon.....	63
Gambar 4. 3 Coherence Score Pemodelan Topik Dengan Normalisasi Merged Lexicon.....	64
Gambar 4. 4 Coherence score (C_v) terhadap Jumlah Topik Pada Normalisasi FastText.....	66
Gambar 4. 5 Coherence score (C_v) terhadap Jumlah Topik Pada Normalisasi Integratif.....	67
Gambar 4. 6 Visualisasi Word Cloud Topik 1	73
Gambar 4. 7 Visualisasi Word Cloud Topik 2	73
Gambar 4. 8 Visualisasi Word Cloud Topik 3	73
Gambar 4. 9 Visualisasi Word Cloud Topik 4	74
Gambar 4. 10 Visualisasi Word Cloud Topik 5	74
Gambar 4. 11 Visualisasi Word Cloud Topik 6	74
Gambar 4. 12 Visualisasi Distribusi Tweet Berdasarkan Topik	75
Gambar 4. 13 Visualisasi Hasil Pemodelan Topik (Intertopic Distance Map)	76

DAFTAR LAMPIRAN

Lampiran 1. Riwayat Hidup	101
Lampiran 2. Surat Keterangan Validasi Ahli	102
Lampiran 3. Daftar Kata Informal Dengan Penulisan Sama Serta Padanan Formal Yang Dipilih	105
Lampiran 4. Daftar Normalisasi Manual.....	116
Lampiran 5. Interpretasi dan Analisis Ahli Terhadap Hasil Penelitian.....	126
Lampiran 6. Rubrik Tes Word dan Topic Intrusion	128
Lampiran 7. Responden Dan Hasil Tes Evaluasi Word & Topic Intrusion	132
Lampiran 8. Kode Tahapan Slang Lexicon Construction.....	133
Lampiran 9. Kode Tahapan Set Manually Informal with Same Formal Words..	134
Lampiran 10. Kode Tahapan Data Crawling	136
Lampiran 11. Kode Tahapan Case Folding, Data Cleaning, Tokenization dan Other Language Removal.....	137
Lampiran 12. Daftar dan Kode Kode Tahapan Manual Normalization.....	139
Lampiran 13. Kode Normalisasi Otomatis.....	141
Lampiran 14. Kode Stop Word Removal, Stemming, dan Data Aggregation	143
Lampiran 15. Kode Tahapan Topic Modeling.....	145

