

TUGAS AKHIR

IMPLEMENTASI BI-LSTM DAN SMOTE UNTUK PENANGANAN

IMBALANCED DATA PADA ANALISIS

SENTIMEN ULASAN MOBILE JKN



Disusun oleh:

Alfredo Mikha Georel L.R

22.18.024

PROGRAM STUDI TEKNIK INFORMATIKA S-1

FAKULTAS TEKNOLOGI INDUSTRI

INSTITUT TEKNOLOGI NASIONAL MALANG

2026

LEMBAR PERSETUJUAN DAN PENGESAHAN
IMPLEMENTASI BI-LSTM DAN SMOTE UNTUK PENANGANAN
IMBALANCED DATA PADA ANALISIS
SENTIMEN ULASAN MOBILE JKN
TUGAS AKHIR

*Disusun dan Diajukan Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik Informatika Strata Satu (S-1)*

Disusun Oleh :

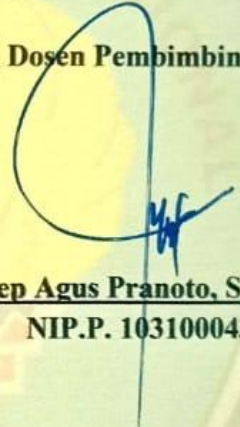
Alfredo Mikha Georel Lie Rumbekwan
22.18.024

Diperiksa dan disetujui oleh :

Dosen Pembimbing I


Eko Heri Susanto S.Kom., M.Kom
NIP.P. 1032400605

Dosen Pembimbing II


Yosep Agus Pranoto, S.T., M.T
NIP.P. 1031000432



Mengetahui,
Wakil Dekan I
Dr. Irmawati Budi Sulistiwati, ST., MT
NIP. 1977061520005012002

PROGRAM STUDI TEKNIK INFORMATIKA S-1
FAKULTAS TEKNOLOGI INDUSTRI
INSTITUT TEKNOLOGI NASIONAL MALANG

2026

LEMBAR PERNYATAAN KEASLIAN TUGAS AKHIR

Sebagai mahasiswa Program Studi Teknik Informatika S-1 Fakultas Teknologi Industri Institut Teknologi Nasional Malang, yang bertanda tangan di bawah ini, saya :

Nama : Alfredo Mikha G.L.R
NIM : 2218024
Program Studi : Teknik Informatika S-1
Fakultas : Fakultas Teknologi Industri

Menyatakan dengan sesungguhnya bahwa skripsi saya dengan judul **“Implementasi Bi-Lstm Dan Smote Untuk Penanganan Imbalanced Data Pada Analisis Sentimen Ulasan Mobile Jkn”** merupakan karya asli dan bukan merupakan duplikat dan mengutip seluruhnya karya orang lain. Apabila di kemudian hari, karya asli saya disinyalir bukan merupakan karya asli saya, maka saya bersedia menerima segala konsekuensi apa pun yang diberikan Program Studi Teknik Informatika S-1 Fakultas Teknologi Industri Institut Teknologi Nasional Malang. Demikian surat pernyataan ini saya buat dengan sebenar-benarnya.

Malang,.....Juli 2026

Yang Membuat Pernyataan



Alfredo Mikha G.L.R

2218024

IMPLEMENTASI BI-LSTM DAN SMOTE UNTUK PENANGANAN IMBALANCED DATA PADA ANALISIS SENTIMEN ULASAN MOBILE JKN

Alfredo Mikha Georel Lie Rumbekwan, Eko Heri Susanto, Yosep Agus Pranoto
Program Studi Teknik Informatika S1, Fakultas Teknologi Industri
Institut Teknologi Nasional Malang, Jalan Raya Karanglo km 2 Malang, Indonesia
2218024@scholar.itn.ac.id

ABSTRAK

Aplikasi Mobile JKN merupakan kanal layanan digital BPJS Kesehatan yang telah menerima lebih dari 954.000 ulasan pengguna di Google Play Store. Volume data yang masif ini menjadikan analisis sentimen otomatis sebagai kebutuhan strategis, namun distribusi ulasan yang tidak seimbang (*imbalanced data*) antara kelas negatif dan positif berpotensi menghasilkan model yang bias terhadap kelas mayoritas. Penelitian ini mengusulkan integrasi arsitektur *Bidirectional Long Short-Term Memory* (Bi-LSTM) dengan prosedur *split-aware* SMOTE untuk menangani permasalahan tersebut secara metodologis. Dataset yang digunakan terdiri dari 20.247 ulasan berbahasa Indonesia hasil *scraping* Google Play Store, dengan distribusi 68,64% negatif dan 31,36% positif (rasio 2,19:1). Representasi fitur dibangun menggunakan Word2Vec *Skip-gram* berdimensi 100 yang dilatih secara mandiri (*trained from scratch*) pada korpus data latih. SMOTE diterapkan secara eksklusif pada data latih setelah pemisahan dataset untuk mencegah kebocoran data (*data leakage*). Validasi statistik dilakukan menggunakan *Wilcoxon Signed-Rank Test* berbasis 10 iterasi *Repeated Holdout*. Model terbaik mencapai akurasi 91,26% dengan *Macro F1-Score* 89,51%, nilai F1 kelas negatif 93,79%, dan F1 kelas positif 85,24%. Hasil uji Wilcoxon menghasilkan nilai *p-value* sebesar 0,1875 ($p > 0,05$), sehingga H_0 gagal ditolak. Temuan ini mengindikasikan bahwa arsitektur Bi-LSTM secara mandiri telah memiliki ketangguhan yang memadai dalam menangani kondisi *imbalanced data* pada rasio 2,19:1.

Kata kunci : Analisis Sentimen, Bi-LSTM, Imbalanced Data, Mobile JKN, SMOTE.

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala Rahmat dan kasih-Nya, sehingga penulis dapat menyelesaikan tugas akhir yang berjudul **“Implementasi Bi-Lstm Dan Smote Untuk Penanganan Imbalanced Data Pada Analisis Sentimen Ulasan Mobile Jkn”** . Tugas akhir ini disusun sebagai salah satu syarat untuk menyelesaikan studi pada Program Sarjana (S-1) Jurusan Teknik Informatika, Fakultas Teknologi Industri, Institut Teknologi Nasional Malang.

Dalam proses penyusunan tugas akhir ini, penulis memperoleh bimbingan, dukungan, serta bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan rasa hormat dan terima kasih kepada:

1. Tuhan Yesus Kristus yang telah memberikan kasih setia, rahmat, dan karuniaNya bagi penyusun sehingga dapat mengerjakan laporan skripsi dengan lancar.
2. Kedua Orang Tua dan Keluarga Besar yang telah memberikan semangat dan dorongan baik secara moral maupun materil untuk menyelesaikan skripsi ini.
3. Bapak Yosep Agus Pranoto, S.T., M.T., selaku Ketua Program Studi Teknik Informatika S-1 Institut Teknologi Nasional Malang.
4. Bapak Eko Heri Susanto S.Kom., M.Kom, selaku Dosen Pembimbing I, atas bimbingan, arahan, serta masukan yang sangat berharga selama proses penelitian dan penyusunan tugas akhir ini.
5. Bapak Yosep Agus Pranoto, S.T., M.T., selaku Dosen Pembimbing II, atas saran, evaluasi, dan dukungan akademik yang membantu penyempurnaan tugas akhir ini.
6. Seluruh dosen Program Studi Teknik Informatika S-1 ITN Malang, atas ilmu, wawasan, dan pengalaman yang telah diberikan selama masa perkuliahan.
7. Seseorang yang istimewa, yang senantiasa memberikan dukungan, semangat, dan motivasi kepada penulis selama proses penyusunan tugas akhir ini.
8. Rekan-rekan dan sahabat, yang telah memberikan dukungan moral serta kebersamaan selama proses penyusunan tugas akhir.
9. Diri sendiri, yang telah berusaha, bertahan, dan menyelesaikan seluruh proses penyusunan tugas akhir ini dengan penuh tanggung jawab.

10. Seluruh pihak yang tidak dapat disebutkan satu per satu, yang telah memberikan bantuan dan dukungan secara langsung maupun tidak langsung dalam penyusunan tugas akhir ini.

Penulis menyadari bahwa tugas akhir ini masih memiliki keterbatasan dan kekurangan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi penyempurnaan penelitian ini di masa mendatang. Semoga tugas akhir ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya di bidang Teknik Informatika.

Malang,... Juli 2026

Penulis

DAFTAR ISI

ABSTRAK	i
KATA PENGANTAR	ii
DAFTAR ISI	iv
DAFTAR GAMBAR	vi
DAFTAR TABEL	viii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	5
1.4 Batasan Masalah	5
1.5 Manfaat Penelitian	7
1.6 Metodologi Penelitian	7
1.7 Sistematika Penulisan	9
BAB II TINJAUAN PUSTAKA	11
2.1 Penelitian Terdahulu	11
2.2 Sentiment Analysis	15
2.3 Natural Language Processing	16
2.4 Prapemrosesan Teks (<i>Text Preprocessing</i>)	18
2.5 <i>Deep Learning</i> dalam Klasifikasi Teks	19
2.6 Long Short-Term Memory (LSTM)	22
2.7 Bidirectional Long Short-Term Memory (Bi-LSTM)	29
2.8 Imbalanced Data dalam Klasifikasi Teks	33
2.9 Synthetic Minority Oversampling Technique (SMOTE)	34

2.10 Word Embedding dan Word2Vec.....	36
2.11 Evaluasi Klasifikasi.....	37
2.12 Wilcoxon Signed-Rank Test	39
BAB III ANALISIS DAN PERANCANGAN	41
3.1 Kebutuhan Fungsional	41
3.2 Kebutuhan Nonfungsional	41
3.3 Arsitektur Sistem.....	42
3.4 Alur Kerja Sistem.....	71
3.5 Desain Prototype Antarmuka Flask	75
BAB IV IMPLEMENTASI DAN PENGUJIAN	77
4.1 Implementasi Sistem Dan Antarmuka	77
4.2 Hasil Prapemrosesan Data	81
4.3 Hasil Penanganan Imbalanced Data.....	87
4.4 Pelatihan Dan Evaluasi Model Bi-Lstm.....	88
4.5 Uji Sensitivitas Hyperparameter	91
4.6 Rangkuman Transformasi Data	93
4.7 Hasil Uji Signifikansi Wilcoxon	98
4.8 Pengujian Penerimaan Pengguna (<i>User Acceptance Test</i>).....	100
4.9 Pengujian Fungsionalitas Sistem (Black Box).....	106
BAB V PENUTUP.....	108
5.1 Kesimpulan	108
5.2 Saran.....	109
DAFTAR PUSTAKA	111

DAFTAR GAMBAR

Gambar 3.1 Arsitektur Sistem	43
Gambar 3.2 Implementasi Kode Program <i>Scraping</i> Ulasan Mobile JKN.....	45
Gambar 3.3 Cuplikan Dataset Hasil <i>Scraping</i>	46
Gambar 3.4 Kode Untuk Program Pembersihan Duplikasi Data	47
Gambar 3.5 Kode Untuk Program Penghapusan Data Tidak Valid	47
Gambar 3.6 Dataset Hasil Pembersihan Duplikasi dan Data Tidak Valid	48
Gambar 3.7 Kode Program Pelabelan Sentimen Otomatis	49
Gambar 3.8 Dataset Final Hasil Tahap Input	49
Gambar 3.9 Kode Program Untuk Text Cleaning	51
Gambar 3.10 Cuplikan Hasil Text Cleaning	52
Gambar 3.11 Implementasi Kode Program Tokenization	53
Gambar 3.12 Cuplikan Hasil Tokenization	53
Gambar 3.13 Implementasi Kode Program Slang Normalization	54
Gambar 3.14 Cuplikan Hasil Slang Normalization	54
Gambar 3.15 Kode Untuk Program Selective Stopword Removal	55
Gambar 3.16 Cuplikan Hasil Selective Stopword Removal	56
Gambar 3.17 Bukit Bahwa Data Test Di Ambil 20 Persen	57
Gambar 3.18 Implementasi Kode Program Pembagian Data Tingkat Pertama	58
Gambar 3.19 Arsitektur Konseptual Word2Vec Skip-gram	59
Gambar 3.20 Transformasi Dimensi Flatten-Reshape pada Pipeline SMOTE	62
Gambar 3.21 Arsitektur Blok Neural Network Bidirectional LSTM.....	66
Gambar 3.22 <i>Flowchart</i> Sistem.....	72
Gambar 3.23 Sub-Flowchart Pelatihan dan Evaluasi Pasangan	74

Gambar 3.24 <i>Low-fi</i> Halaman Input tunggal	76
Gambar 3.25 <i>Low-fi</i> Halaman Klasifikasi <i>Batch</i> CSV	76
Gambar 4.1 Halaman Login	77
Gambar 4.2 Halaman Input Teks Tunggal - Tampilan Formulir	78
Gambar 4.3 Halaman Input Teks Tunggal - Tampilan Hasil Prediksi	79
Gambar 4.4 Halaman Analisis Batch CSV - Tampilan Formulir Unggah	79
Gambar 4.5 Tampilan Hasil Klasifikasi dan Pie Chart	80
Gambar 4.6 Tampilan Fitur Unduh Laporan CSV	80
Gambar 4.7 Halaman Classification Report dan Confusion Matrix.....	81
Gambar 4.8 Confusion Matrix.....	90
Gambar 4.9 Dokumentasi Pengisian Kuesioner oleh Responden	101
Gambar 4.10 Distribusi Usia Responden	101
Gambar 4.11 Distribusi Pekerjaan Responden	101
Gambar 4.12 Evaluasi Kemudahan Penggunaan.....	103
Gambar 4.13 Fitur "Pengujian Teks Tunggal"	104
Gambar 4.14 Fitur penyorotan token membantu.....	104
Gambar 4.15 Fitur "Unggah Data CSV" mempermudah proses	104
Gambar 4.16 Sistem merespons setiap input pengguna dengan responsif.....	104
Gambar 4.17 Sistem berjalan stabil dan tidak mengalami error.....	105
Gambar 4.18 kolom umpan balik	105

DAFTAR TABEL

Tabel 2.1 Ringkasan Penelitian Terdahulu.....	13
Tabel 3.1 Arsitektur BI-LSTM.....	63
Tabel 4.1 Cuplikan Data Mentah Ulasan Mobile JKN.....	82
Tabel 4.2 Hasil Text Cleaning dan Case Folding.....	83
Tabel 4.3 Hasil Normalisasi Slang dan Selective Stopword Removal.....	84
Tabel 4.4 Statistik Distribusi Panjang Token	86
Tabel 4.5 Distribusi Data Latih Sebelum dan Sesudah SMOTE.....	88
Tabel 4.6 Classification Report Bi-LSTM dengan SMOTE (Iterasi Terbaik)	89
Tabel 4.7 Confusion Matrix Model Bi-LSTM dengan SMOTE	89
Tabel 4.8 Hasil Pengujian Sensitivitas Dropout Rate	91
Tabel 4.9 Hasil Pengujian Sensitivitas Learning Rate	93
Tabel 4.10 Rangkuman transformasi data	94
Tabel 4.11 Perbandingan Macro F1-Score pada 10 Iterasi Pengujian	99
Tabel 4.12 Hasil Kuantitatif User Acceptance Test	102
Tabel 4.13 Pengujian Fungsionalitas Login dan Logout.....	106