



SKRIPSI

ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN APLIKASI SAPAWARGA JAWA BARAT SUPER APPS MENGGUNAKAN LDA DAN INDOBERT

LAYLA MAZIDATUS SA`ADAH

22082010105

DOSEN PEMBIMBING

Abdul Rezha Efrat Najaf, S.Kom, M.Kom.

Nambi Sembilu, S.Kom, M.Kom.

KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI SISTEM INFORMASI

SURABAYA

2026

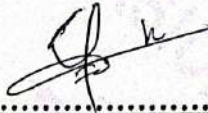
LEMBAR PENGESAHAN

ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN APLIKASI SAPAWARGA JAWA BARAT SUPER APPS MENGGUNAKAN LDA DAN INDOBERT

Oleh :
LAYLA MAZIDATUS SA'ADAH
NPM. 22082010105

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Sistem Informasi Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 29 Juni 2026.

Abdul Rezha Efrat Najaf, S.Kom., M.Kom.
NIP. 199409292022031008



(Pembimbing I)

Nambi Sembilu, S.Kom., M.Kom.
NIP. 199005162024061003



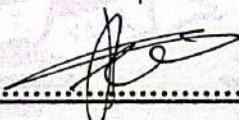
(Pembimbing II)

Reisa Permatasari, S.T., M.Kom.
NIP. 199205142022032007



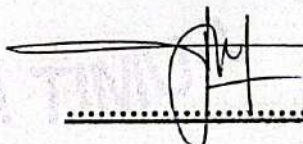
(Ketua Penguji)

Tri Luhur Indavanti Sugata, S.ST., M.IM.
NIP. 199206162024062001



(Anggota Penguji II)

Mohammad Al Hafidz, S.Kom., M.Kom.
NIP. 199109222025061003



(Anggota Penguji III)

Mengetahui,
Dekan Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

**ANALISI SENTIMEN BERBASIS ASPEK PADA ULASAN APLIKASI
SAPAWARGA JAWA BARAT SUPER APPS MENGGUNAKAN LDA
DAN INDOBERT**

Oleh :
LAYLA MAZIDATUS SA'ADAH
NPM. 22082010105

Menyetujui,
Koordinator Program Studi Sistem Informasi
Fakultas Ilmu Komputer



Siti Mukaromah, S.Kom, M.Kom.
NIP. 198107042021212011

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama : LAYLA MAZIDATUS SA'ADAH
NPM : 22082010105
Program : Sarjana (S1)
Program Studi : Sistem Informasi
Fakultas : Fakultas Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi Apabila dikemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.



Surabaya, 15 Juli 2026
Yang Membuat Pernyataan,



LAYLA MAZIDATUS SA'ADAH
NPM. 22082010105

ABSTRAK

Nama Mahasiswa / NPM : Layla Mazidatus Sa`adah / 22082010105
Judul Skripsi : Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Sapawarga Jawa Barat Super Apps Menggunakan LDA dan IndoBERT
Dosen Pembimbing : 1. Abdul Rezha Efrat Najaf, S.Kom., M.Kom.
2. Nambi Sembilu, S.Kom, M.Kom.

Ulasan pengguna pada aplikasi Sapawarga Jawa Barat SuperApps mengandung informasi penting tentang pengalaman serta opini pengguna terhadap berbagai fitur serta layanan yang tersedia dalam aplikasi. Namun, pendekatan analisis sentimen secara umum belum dapat mengaitkan sentimen dengan aspek-aspek spesifik yang dibahas dalam setiap ulasan. Penelitian ini menerapkan *Aspect Based Sentiment Analysis* (ABSA) dengan memanfaatkan LDA untuk mengidentifikasi aspek-aspek utama serta membandingkan kinerja Naïve Bayes, SVM, dan IndoBERT dalam klasifikasi sentimen berbasis aspek. Data dikumpulkan melalui *web scraping* Google Play Store dan App Store sebanyak 6.012 ulasan, yang setelah proses *filtering* menghasilkan 4.386 ulasan valid. Tahapan penelitian meliputi *preprocessing*, pemodelan topik menggunakan LDA, pelabelan data, penanganan ketidakseimbangan kelas, serta pembangunan model menggunakan Naive Bayes, SVM, dan IndoBERT sebagai pembanding. Hasil pemodelan topik LDA menghasilkan tiga aspek utama, yaitu kendala login dan akses akun, kinerja aplikasi, serta fitur pembayaran pajak kendaraan, dengan jumlah topik optimal sebanyak tiga berdasarkan coherence score sebesar 0,4247. Pada model Machine Learning, Naive Bayes secara konsisten unggul dibandingkan SVM, sementara teknik penyeimbangan data ROS dan SMOTE menghasilkan performa yang relatif setara satu sama lain. Model Machine Learning terbaik diperoleh oleh Naive Bayes dengan TF-IDF unigram dan bigram menggunakan ROS, dengan Accuracy 87,88%, Precision 79,15%, Recall 81,38%, dan Macro F1-Score 79,59%. Pada model IndoBERT, Focal Loss secara konsisten menghasilkan performa terbaik dibandingkan Class Weight (CW) maupun kombinasi CW+FL, sedangkan kombinasi CW+FL cenderung menurunkan performa pada beberapa skenario. Performa terbaik dicapai pada konfigurasi dengan *epoch* 10, *batch size* 32, *learning rate* 2×10^{-5} , dan *dropout* 0,3, dengan Accuracy 94,32%, Precision 89,82%, Recall 91,61%, dan Macro F1-Score 90,58%. Hasil penelitian menunjukkan bahwa pendekatan ABSA berbasis LDA dan IndoBERT memberikan performa yang lebih unggul dibandingkan model Machine Learning, sementara Focal Loss menjadi strategi penanganan ketidakseimbangan kelas yang paling efektif untuk proses *fine-tuning* IndoBERT.

Kata kunci : ABSA, LDA, IndoBERT, *Machine Learning*, Sapawarga.

ABSTRACT

Student Name / NPM : Layla Mazidatus Sa`adah / 22082010105
Thesis Title : *Aspect-Based Sentiment Analysis of User Reviews on the Sapawarga Jawa Barat Super Apps Using LDA and IndoBERT*
Advisor : 1. Abdul Rezha Efrat Najaf, S.Kom., M.Kom.
2. Nambi Sembilu, S.Kom, M.Kom.

ABSTRACT

User reviews of the Sapawarga Jawa Barat SuperApps provide valuable insights into users' experiences and opinions regarding the various features and services offered by the application. However, conventional sentiment analysis approaches are unable to associate sentiment with the specific aspects discussed in each review. This study applies Aspect-Based Sentiment Analysis (ABSA) by utilizing Latent Dirichlet Allocation (LDA) to identify the main aspects and compares the performance of Naïve Bayes, Support Vector Machine (SVM), and IndoBERT for aspect-based sentiment classification. The dataset was collected through web scraping from the Google Play Store and Apple App Store, resulting in 6,012 reviews, which were filtered to obtain 4,386 valid reviews. The research process consisted of text preprocessing, topic modeling using LDA, data annotation, class imbalance handling, and model development using Naïve Bayes, SVM, and IndoBERT for comparative evaluation. The LDA topic modeling identified three main aspects, namely login and account access issues, application performance, and vehicle tax payment features. The optimal number of topics was determined to be three based on the highest coherence score of 0.4247. Among the machine learning models, Naïve Bayes consistently outperformed SVM, while the Random Oversampling (ROS) and Synthetic Minority Oversampling Technique (SMOTE) methods achieved relatively comparable performance. The best machine learning model was Naïve Bayes with TF-IDF unigram and bigram features combined with ROS, achieving an Accuracy of 87.88%, Precision of 79.15%, Recall of 81.38%, and a Macro F1-score of 79.59%. For the IndoBERT model, Focal Loss consistently outperformed both Class Weight (CW) and the combined CW+FL approach, whereas the CW+FL combination tended to reduce performance in several experimental scenarios. The best performance was achieved using a configuration of 10 epochs, a batch size of 32, a learning rate of 2×10^{-5} , and a dropout rate of 0.3, resulting in an Accuracy of 94.32%, Precision of 89.82%, Recall of 91.61%, and a Macro F1-score of 90.58%. The results demonstrate that the LDA–IndoBERT-based ABSA approach outperformed the conventional machine learning models, while Focal Loss proved to be the most effective strategy for handling class imbalance during the fine-tuning process of IndoBERT.

Keywords: ABSA, LDA, IndoBERT, Machine Learning, Sapawarga.

KATA PENGANTAR

Segala puji hanya milik Allah SWT, atas segala rahmat, pertolongan, serta kasih sayang-Nya yang senantiasa menyertai setiap langkah penulis. Berkat izin dan kehendak-Nya, skripsi yang berjudul "**Analisis Sentimen Berbasis Aspek pada Ulasan Aplikasi Sapawarga Jawa Barat Super Apps Menggunakan LDA dan IndoBERT**" akhirnya dapat diselesaikan. Semoga shalawat serta salam senantiasa tercurah limpahkan kepada junjungan Nabi Muhammad SAW, teladan yang membawa umat manusia menuju kehidupan yang diterangi iman, akhlak, dan ilmu pengetahuan.

Penyusunan skripsi ini dilakukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sistem Informasi pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, UPN "Veteran" Jawa Timur. Proses penyusunannya menjadi perjalanan yang penuh pembelajaran, tantangan, sekaligus pengalaman berharga. Selama proses tersebut, penulis memperoleh begitu banyak bantuan, arahan, dukungan, doa, dan motivasi dari berbagai pihak. Oleh karena itu, dengan penuh rasa hormat dan penghargaan, penulis menyampaikan ucapan terima kasih kepada:

1. Allah SWT, atas limpahan rahmat, kesehatan, kemudahan, kekuatan, dan kesempatan yang diberikan sehingga setiap tahapan penelitian dapat dilalui dengan baik.
2. Kedua orang tua serta seluruh keluarga besar penulis, termasuk kakek, nenek, tante, dan adik-adik tercinta, yang senantiasa menghadirkan kasih sayang, doa, dukungan moral maupun material, serta menjadi sumber semangat yang tidak pernah berhenti menguatkan penulis.
3. Almarhumah nenek tercinta, Ibu Ainul Faizah, yang semasa hidupnya selalu menghadirkan kasih sayang, perhatian, doa, dan semangat bagi penulis. Semoga Allah SWT melimpahkan rahmat-Nya, menempatkan beliau di tempat terbaik di sisi-Nya, serta menerima seluruh amal kebaikan beliau sebagai amal jariyah yang terus mengalir.
4. Ibu Siti Mukaromah, S.Kom., M.Kom., selaku dosen wali sekaligus Koordinator Program Studi Sistem Informasi UPN "Veteran" Jawa Timur.

5. Bapak Abdul Rezha Efrat Najaf, S.Kom., M.Kom. dan Bapak Nambi Sembilu, S.kom., M.Kom., selaku dosen pembimbing yang dengan penuh kesabaran telah meluangkan waktu, tenaga, dan pemikiran untuk memberikan arahan, masukan, serta bimbingan selama proses penelitian hingga penyusunan skripsi ini.
6. Seluruh dosen Program Studi Sistem Informasi UPN “Veteran” Jawa Timur yang telah membagikan ilmu, pengalaman, dan nilai-nilai pembelajaran selama masa perkuliahan.
7. Sahabat-sahabat terdekat penulis, yang selalu hadir memberikan dukungan, motivasi, dan semangat dalam setiap proses yang dilalui selama masa studi.
8. Teman-teman Program Studi Sistem Informasi angkatan 2022 serta seluruh pihak yang tidak dapat disebutkan satu per satu, atas segala bentuk bantuan, kerja sama, dan doa yang turut mengiringi penyelesaian penelitian ini.
9. Diri penulis sendiri, yang memilih untuk tetap melangkah ketika keadaan terasa berat, bersedia terus belajar di tengah berbagai keterbatasan, serta tidak berhenti berusaha hingga setiap tahapan penelitian ini dapat diselesaikan. Terima kasih telah menjaga keyakinan bahwa setiap proses akan menemukan akhirnya pada waktu yang telah ditetapkan.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan, baik dari sisi pembahasan maupun hasil yang diperoleh. Oleh sebab itu, kritik dan saran yang bersifat membangun akan diterima dengan terbuka sebagai bahan refleksi dan penyempurnaan penelitian di masa mendatang. Besar harapan penulis agar penelitian ini dapat memberikan manfaat, menambah khazanah keilmuan, serta menjadi salah satu referensi bagi pengembangan penelitian di bidang Aspect-Based Sentiment Analysis pada analisis ulasan aplikasi layanan publik.

Surabaya, 29 Juni 2026

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN.....	iii
LEMBAR PERSETUJUAN	v
SURAT PERNYATAAN BEBAS PLAGIASI	vii
ABSTRAK.....	ix
ABSTRACT.....	xi
KATA PENGANTAR.....	xiii
DAFTAR ISI.....	xv
DAFTAR GAMBAR.....	xxi
DAFTAR TABEL.....	xxv
DAFTAR LAMPIRAN.....	xxvii
DAFTAR SINGKATAN, ISTILAH DAN SIMBOL	xxix
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	5
1.3. Batasan Masalah.....	6
1.4. Tujuan Penelitian.....	6
1.5. Manfaat Penelitian.....	7
1.6. Sistematika Penulisan	7
BAB II TINJAUAN PUSTAKA	11
2.1. Dasar Teori.....	11
2.1.1. Aplikasi Sapawarga	11
2.1.2. <i>Text Mining</i>	12
2.1.3. Web Scraping	13
2.1.4. Analisis Sentimen Berbasis Aspek (<i>ABSA</i>)	13

2.1.5. <i>Latent Dirichlet Allocation (LDA)</i>	14
2.1.6. Model Transformer IndoBERT	14
2.1.7. <i>Support Vector Machine</i>	15
2.1.8. Naïve Bayes.....	17
2.1.9. <i>Feature Extraction</i>	18
2.1.10. <i>Handling Imbalance</i>	19
2.1.11. <i>Confusion matrix</i>	23
2.1.12. <i>Fleiss Kappa</i>	25
2.1.13. Flask.....	25
2.2. Penelitian Terdahulu.....	26
BAB III METODOLOGI PENELITIAN.....	33
3.1. Studi Literatur	34
3.2. Analisis Kebutuhan	34
3.2.1. Kebutuhan Data.....	34
3.2.2 Kebutuhan Hardware	35
3.3. Pengumpulan Data	35
3.4. Eksplorasi Data	36
3.5. Data Filtering	37
3.6. Text Preprocessing	37
3.6.1 <i>Text Cleaning</i>	38
3.6.2 Case Folding.....	39
3.6.3 Tokenization.....	39
3.6.4 Normalization.....	40
3.6.5 Stopword Removal	41
3.6.6 Stemming	41
3.7 Pemodelan Topik	42

3.8 Pelabelan Data	42
3.9 Pembagian Data	45
3.10 <i>Handling Imbalance</i>	46
3.11 Pemodelan Klasifikasi	49
3.12 Evaluasi Model	52
3.13 Implementasi Sistem	52
3.14 Penyusunan Laporan	53
BAB IV HASIL DAN PEMBAHASAN.....	55
4.1. Studi Literatur	55
4.2. Analisis Kebutuhan	55
4.2.1 Kebutuhan Data.....	56
4.2.2 Kebutuhan Hardware	56
4.3. Pengumpulan Data	56
4.4. Explanatory Data Analysis (EDA)	61
4.5. Data Filtering	66
4.6. <i>Text Preprocessing</i>	69
4.6.1 Text Cleaning	69
4.6.2 Case Folding	71
4.6.3 Normalisasi Kata	72
4.6.4 Tokenization.....	74
4.6.5 Stopword Removal	75
4.6.6 Stemming	77
4.6.7 Penghapusan Data Kosong Setelah Pre-processing	78
4.7. Pemodelan Topik.....	79
4.8. Interpretasi Topik	83
4.9. <i>Data Labelling</i>	87

4.9.1 Pedoman Pelabelan Data.....	88
4.9.2 Media Pelabelan Data	89
4.9.3 Menghitung Tingkat Kesepakatan Anotator	90
4.9.4 Penentuan Label Akhir dengan Majority Voting	92
4.9.5 Distribusi Label Sentimen.....	93
4.10. <i>Data Splitting</i>	93
4.10.1 Pembagian Data Aspect Detection	93
4.10.2 Pembagian Data Sentiment Classification	95
4.11. <i>Handling Imbalance</i>	97
4.11.1. Random Oversampling (ROS)	97
4.11.2. SMOTE (<i>Synthetic Minority Oversampling Technique</i>)	98
4.11.3. Class Weight	100
4.11.4. Focal Loss	102
4.11.5 Kombinasi <i>Class Weight</i> dan <i>Focal Loss</i>	103
4.12. <i>Data Modelling</i>	104
4.12.1. Pembangunan Model SVM	104
4.12.2. Pembangunan Model Naive Bayes	105
4.12.3. Pembangunan Model IndoBERT	106
4.13. Evaluasi Model.....	107
4.13.1 Hasil Evaluasi Model Machine Learning (Naïve Bayes dan SVM). ..	108
4.13.2 Hasil Evaluasi Model IndoBERT (Skenario S5–S8).....	109
4.14. Penentuan Model Terbaik	110
4.14.1. Analisis Model Terbaik (S8-FL)	111
4.15. Implementasi Website	115
4.15.1 Halaman Dashboard.....	116
4.15.2. Halaman Analisis Kalimat	117

4.15.3 Halaman Analisis File.....	117
BAB V PENUTUP.....	119
5.1. Kesimpulan	119
5.2. Saran.....	120
DAFTAR PUSTAKA.....	123
LAMPIRAN.....	129

DAFTAR GAMBAR

Gambar 2. 1 Gambar Aplikasi Sapawarga	11
Gambar 2. 2 BERT Input Representation	15
Gambar 2. 3 Visualisasi SVM dua Kelas.....	16
Gambar 3.1 Alur Metodologi Penelitian.....	33
Gambar 3. 2 Alur <i>Scraping Data</i>	35
Gambar 3. 3 Alur <i>Text Preprocessing</i>	37
Gambar 3.4 Diagram Alur Pembagian Data	45
Gambar 3.5 Alur Penanganan Ketidakseimbangan Data.....	47
Gambar 3. 6 Alur Implementasi Pemodelan Klasifikasi pada Machine Learning dan IndoBERT	49
Gambar 3.7 Alur Klasifikasi Aspect-Based Sentiment Analysis (Aspect Detection dan Sentiment Classification).....	50
Gambar 4. 1 Implementasi <i>Web Scraping</i> Ulasan Google Play Store.....	57
Gambar 4. 2 Implementasi Web Scrapping Ulasan Apple App Store.....	58
Gambar 4. 3 Implementasi Penggabungan dan Filtering Tanggal Dataset.....	60
Gambar 4. 4 Distribusi Ulasan Berdasarkan Sumber Platform.....	62
Gambar 4. 5 Distribusi Rating Ulasan Pengguna	63
Gambar 4. 6 Distribusi Jumlah Kata pada Ulasan.....	63
Gambar 4. 7 Distribusi Ulasan per Tahun.....	64
Gambar 4. 8 Visualisasi WordCloud Ulasan Pengguna	65
Gambar 4. 9 WordCloud Komentar Pendek (≤ 2 Kata).....	68
Gambar 4. 10 Hasil Penandaan Data Filtering	69
Gambar 4. 11 Implementasi <i>Text Cleaning</i>	70
Gambar 4. 12 Implementasi <i>Case Folding</i>	71
Gambar 4. 13 Implementasi Normalisasi Kata.....	73

Gambar 4. 14 Implementasi <i>Tokenization</i>	74
Gambar 4. 15 Implementasi <i>Stopword Removal</i>	75
Gambar 4. 16 Implementasi <i>Stemming</i>	77
Gambar 4. 17 Kode Penghapusan Data Kosong Setelah Pre-processing	79
Gambar 4. 18 Kode Persiapan Data dan Pembentukan Corpus LDA.....	80
Gambar 4. 19 Kode Evaluasi Nilai Coherence Score	81
Gambar 4. 20 Visualisasi Nilai Coherence Score.....	82
Gambar 4. 21 Kode Implementasi Model LDA dengan Parameter Terbaik.....	83
Gambar 4. 22 Diagram Batang Kata-Kata Dominan pada Topik 0.....	84
Gambar 4. 23 Diagram Batang Kata-Kata Dominan pada Topik 1.....	85
Gambar 4. 24 Diagram Batang Kata-Kata Dominan pada Topik 2.....	86
Gambar 4. 25 Tampilan Pelabelan Data di Google Spreadsheet.....	89
Gambar 4. 26 Contoh Hasil Pelabelan oleh Tiga Anotator.....	90
Gambar 4. 27 Kode Program Perhitungan Fleiss' Kappa	91
Gambar 4. 28 Proses Majority Voting pada Google Spreadsheet	92
Gambar 4. 29 Contoh Hasil Label Final	93
Gambar 4. 30 Visualisasi Distribusi Label Sentimen	93
Gambar 4. 31 Kode Program Pembentukan Label Aspect Detection.....	94
Gambar 4. 32 Kode Program Pembagian Data Aspect Detection	95
Gambar 4. 33 Kode Program Pembagian Data Sentiment Classification	96
Gambar 4. 34 Kode Program Random Oversampling	98
Gambar 4. 35 Kode Program SMOTE.....	99
Gambar 4. 36 Kode Program Perhitungan Class Weight.....	101
Gambar 4. 37 Kode Program Implementasi Focal Loss	102
Gambar 4. 38 Kode Program Implementasi Class Weight dan Focal Loss	103
Gambar 4. 39 Kode Program TF-IDF Vectorizer.....	104

Gambar 4. 40 Kode Program Pelatihan Model SVM	105
Gambar 4. 41 Kode Pelatihan Model Naive Bayes	105
Gambar 4. 42 Kode Load Model IndoBERT	106
Gambar 4. 43 Kode Tokenisasi Data	106
Gambar 4. 44 Kode <i>Fine-Tuning</i> IndoBERT	107
Gambar 4. 45 <i>Confusion Matrix</i> Aspect Detection Login (S8-FL)	112
Gambar 4. 46 <i>Confusion Matrix</i> Sentiment Classification Login (S8-FL)	112
Gambar 4. 47 <i>Confusion Matrix</i> Aspect Detection Kinerja (S8-FL)	113
Gambar 4. 48 <i>Confusion Matrix</i> Sentiment Classification Kinerja (S8-FL)	113
Gambar 4. 49 <i>Confusion Matrix</i> Aspect Detection Pajak (S8-FL)	114
Gambar 4. 50 <i>Confusion Matrix</i> Sentiment Classification Pajak (S8-FL)	114
Gambar 4. 51 Kode Program Integrasi Flask dan Model IndoBERT	115
Gambar 4. 52 Tampilan Halaman Dashboard Sistem ABSA Sapawarga	116
Gambar 4. 53 Tampilan Halaman Analisis Kalimat	117
Gambar 4. 54 Tampilan Halaman Analisis File	118

DAFTAR TABEL

Tabel 2. 1 Komponen <i>Confusion Matrix</i>	24
Tabel 2. 2 Penelitian Terdahulu.....	26
Tabel 3. 1 Tahapan <i>Text Preprocessing</i> pada Setiap Model	38
Tabel 3. 2 Contoh <i>Text Cleaning</i>	39
Tabel 3. 3 Contoh <i>Case Folding</i>	39
Tabel 3. 4 Contoh <i>Tokenization</i>	40
Tabel 3. 5 Contoh <i>Normalization</i>	40
Tabel 3. 6 Contoh Penerapan <i>Stopword Removal</i>	41
Tabel 3. 7 Contoh <i>Stemming</i>	42
Tabel 3.8 Kategori Label Sentimen	43
Tabel 3.9 Decision Rules Pelabelan Data	43
Tabel 3.10 Contoh Output Pelabelan oleh Anotator AI (ChatGPT).....	44
Tabel 3.11 Perbandingan Pendekatan Penanganan Imbalance antara Model ML dan IndoBERT.....	48
Tabel 3. 12 Skenario Pengujian Pembangunan Model Klasifikasi.....	51
Tabel 4. 1 Contoh Hasil Ulasan dari Google Play Store.....	57
Tabel 4. 2 Contoh Hasil Ulasan dari Apple App Store.....	59
Tabel 4. 3 Struktur Dataset Hasil Scraping	60
Tabel 4. 4 Ringkasan Jumlah Data Hasil Scraping.....	60
Tabel 4. 5 Struktur Dataset.....	61
Tabel 4. 6 Ringkasan Diagnosis Kualitas Data	65
Tabel 4. 7 Ringkasan Proses Data Filtering	67
Tabel 4. 8 Contoh Proses Filtering Data	67
Tabel 4. 9 Contoh Hasil Text Cleaning.....	70
Tabel 4. 10 Contoh Hasil Case Folding	71

Tabel 4. 11 Contoh Hasil Normalisasi Kata.....	73
Tabel 4. 12 Contoh Hasil Tokenization	74
Tabel 4. 13 Top Terms Setelah Stopword Removal Pertama	76
Tabel 4. 14 Contoh Hasil Stopword Removal.....	76
Tabel 4. 15 Contoh Hasil Stemming.....	77
Tabel 4. 16 Hasil Pengecekan Data Setelah Pre-processing	79
Tabel 4. 17 Hasil Evaluasi Coherence Score	81
Tabel 4. 18 Distribusi Kata pada Setiap Topik.....	83
Tabel 4. 19 Interpretasi Topik	86
Tabel 4. 20 Kategori Label Sentimen	88
Tabel 4. 21 Decision Rules Pelabelan Data	88
Tabel 4. 22 Skala Penilaian Fleiss' Kappa	90
Tabel 4. 23 Hasil Perhitungan Fleiss' Kappa	91
Tabel 4. 24 Keterangan Label Aspect Detection	94
Tabel 4. 25 Distribusi Pembagian Data Aspect Detection	95
Tabel 4. 26 Distribusi Pembagian Data Sentiment Classification	96
Tabel 4. 27 Distribusi Label Sentimen per Aspek	97
Tabel 4. 28 Distribusi Data Latih Sebelum dan Sesudah ROS	98
Tabel 4. 29 Distribusi Data Latih Sebelum dan Sesudah SMOTE	99
Tabel 4. 30 Bobot Class Weight Aspect Detection	101
Tabel 4. 31 Bobot Class Weight Sentiment Classification	101
Tabel 4. 32 Rekapitulasi Hasil Evaluasi Model <i>Machine Learning</i> (S1–S4)	108
Tabel 4. 33 Rekapitulasi Hasil Evaluasi Model IndoBERT (S5–S8).....	109
Tabel 4. 34 Rekapitulasi Perbandingan Seluruh Skenario Pengujian.....	110
Tabel 4. 35 Hasil Evaluasi Detail Model Terbaik S8-FL.....	111

DAFTAR LAMPIRAN

Lampiran 1. Bukti Pengajuan Izin Penelitian melalui Email	129
Lampiran 2. Persetujuan Penelitian	130
Lampiran 3, Dataset Aplikasi Sapawarga	131
Lampiran 4. Wireframe Sistem	131

DAFTAR SINGKATAN, ISTILAH DAN SIMBOL

SINGKATAN/ISTILAH/ SIMBOL	ARTI DAN KETERANGAN
<i>ABSA</i>	: <i>Aspect-Based Sentiment Analysis</i> , metode analisis sentimen berdasarkan aspek tertentu.
<i>LDA</i>	: <i>Latent Dirichlet Allocation</i> , algoritma pemodelan topik untuk ekstraksi aspek.
<i>IndoBERT</i>	: <i>Model bahasa berbasis BERT yang dilatih menggunakan korpus Bahasa Indonesia.</i>
<i>BERT</i>	: <i>Bidirectional Encoder Representations from Transformers.</i>
<i>NLP</i>	: <i>Natural Language Processing</i> , bidang kecerdasan buatan yang mempelajari pemrosesan bahasa alami.
<i>TF-IDF</i>	: <i>Term Frequency–Inverse Document Frequency</i> , metode pembobotan kata pada dokumen.
<i>SMOTE</i>	: <i>Synthetic Minority Over-sampling Technique</i> , metode penyeimbangan data dengan menghasilkan sampel sintetis.
<i>CW</i>	: <i>Class Weight</i> , teknik pemberian bobot pada kelas minoritas saat pelatihan model.
<i>FL</i>	: <i>Focal Loss</i> , fungsi loss yang berfokus pada sampel yang sulit diklasifikasikan.
<i>F1-Score</i>	: <i>Harmonic mean antara precision dan recall.</i>
<i>Accuracy</i>	: <i>Persentase prediksi yang benar terhadap seluruh data.</i>
<i>Precision</i>	: <i>Tingkat ketepatan prediksi pada suatu kelas.</i>
<i>Recall</i>	: <i>Kemampuan model menemukan seluruh data yang relevan pada suatu kelas.</i>
<i>AI</i>	: <i>Artificial Intelligence</i> atau kecerdasan buatan.
<i>API</i>	: <i>Application Programming Interface.</i>