



SKRIPSI

**ANALISIS EMOSI BERBASIS ASPEK PADA
ULASAN APLIKASI IDENTITAS
KEPENDUDUKAN DIGITAL MENGGUNAKAN
INDOBERT**

VIVIANA PURBA

22082010152

DOSEN PEMBIMBING

Eka Dyar Wahyuni, S.Kom., M.Kom.

Tri Luhur Indayanti Sugata, S.ST, M.IM

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SISTEM INFORMASI
SURABAYA
2026**

LEMBAR PENGESAHAN

ANALISIS EMOSI BERBASIS ASPEK PADA ULASAN APLIKASI IDENTITAS KEPENDUDUKAN DIGITAL MENGGUNAKAN INDOBERT

Oleh :

VIVIANA PURBA

NPM. 22082010152

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Sistem Informasi Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 13 Mei 2026.

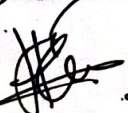
Menyetujui,

Eka Dyar Wahyuni, S.Kom, M.Kom.
NIP. 19841201 2021212 005


.....

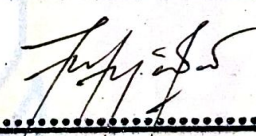
(Pembimbing I)

Tri Luhur Indayanti Sugata, S.ST, M.IM
NIP. 19920616 2024062 001


.....

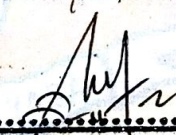
(Pembimbing II)

Seftin Fitri Ana Wati, S.Kom, M.Kom.
NPT. 212199 10 320267


.....

(Ketua Penguji)

Anindo Saka Fitri, S.Kom, M.Kom.
NIP. 19930325 2024062 001


.....

(Penguji II)

Abdul Rezha Efrat Najaf, S.Kom, M.Kom.
NIP. 19940929 2022031 008


.....

(Penguji III)

Mengetahui,
Dekan Fakultas Ilmu Komputer


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

LEMBAR PERSETUJUAN

ANALISIS EMOSI BERBASIS ASPEK PADA ULASAN APLIKASI IDENTITAS KEPENDUDUKAN DIGITAL MENGGUNAKAN INDOBERT


Oleh :

VIVIANA PURBA

NPM. 22082010152



Menyetujui,
Koordinator Program Studi Sistem Informasi
Fakultas Ilmu Komputer


Siti Mukaromah, S.Kom., M.Kom.
NIP. 19810704 2021212 011

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama : Viviana Purba
NPM : 22082010152
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, 26 Mei 2026
Yang Membuat Pernyataan,



VIVIANA PURBA
NPM. 22082010152

ABSTRAK

Nama Mahasiswa / NPM : Viviana Purba / 22082010152
Judul Skripsi : Analisis Emosi Berbasis Aspek Pada Ulasan Aplikasi Identitas Kependudukan Digital Menggunakan IndoBERT
Dosen Pembimbing : 1. Eka Dyar Wahyuni S.Kom, M.Kom
2. Tri Luhur Indayanti Sugata S.ST, M.IM

Aplikasi Identitas Kependudukan Digital (IKD) yang diluncurkan oleh Ditjen Dukcapil Kemendagri telah diunduh lebih dari 10 juta pengguna di Google Play Store dengan rating 3,2 dari sekitar 70,7 ribu ulasan. Ulasan pengguna berperan sebagai indikator signifikan yang mencerminkan mutu layanan serta memengaruhi tingkat kepercayaan dan penerimaan publik terhadap layanan pemerintah digital, sekaligus menjadi sumber penting bagi pengembang dalam mengidentifikasi prioritas perbaikan. Namun, analisis sentimen umum dinilai tidak memadai karena tidak mampu mengidentifikasi aspek dan sumber permasalahan yang dibahas di ulasan secara spesifik. Penelitian ini menerapkan Analisis Emosi Berbasis Aspek dengan klasifikasi multilabel menggunakan model IndoBERT secara *end-to-end* pada ulasan berbahasa Indonesia yang dikumpulkan dari Google Play Store periode Juni 2024 hingga November 2025. Untuk mengidentifikasi aspek layanan yang dominan, dilakukan perbandingan dua metode pemodelan topik, yaitu Latent Dirichlet Allocation (LDA) dan BERTopic. Hasil evaluasi menunjukkan BERTopic lebih unggul dengan coherence score sebesar 0,6196 dan diversity sebesar 0,92 dibandingkan LDA yang menghasilkan coherence score 0,5083 dan diversity 0,9333. BERTopic berhasil mengidentifikasi lima aspek yang lebih spesifik dan granular, yaitu Login & Akses Akun, Scan Barcode ke Dukcapil, Verifikasi Foto Wajah, Dokumen & Layanan Digital, dan Kompatibilitas Perangkat Android. Aspek-aspek tersebut selanjutnya digunakan dalam pelabelan data secara manual oleh tiga anotator manusia dan dua anotator AI (Claude dan ChatGPT), dengan tingkat kesepakatan yang diukur menggunakan Krippendorff's Alpha berkisar antara 0,71 hingga 0,73. Model klasifikasi dikembangkan dengan uji coba delapan skenario yang secara bertahap menggabungkan *preprocessing*, fungsi loss (BCE standar, Weighted BCE dengan *pos_weight*, dan Focal Loss), threshold tuning, serta variasi rasio pembagian data. Model terbaik diperoleh dari kombinasi preprocessing, Focal Loss, threshold 0,4, dan rasio split 60:20:20, menghasilkan F1 Score Macro sebesar 0,3916 (meningkat 24,1% dari *baseline*), F1 Score Micro sebesar 0,9134, Recall sebesar 0,9423, dan Hamming Loss sebesar 0,0308. Sebagai luaran akhir, dibangun sebuah website pengklasifikasi emosi berbasis aspek menggunakan framework Flask yang untuk menyajikan hasil klasifikasi yang kompleks menjadi informasi yang informatif dan lebih mudah dipahami.

Keywords: Analisis Emosi Berbasis Aspek, IndoBERT, Klasifikasi Multilabel, BERTopic, LDA, Identitas Kependudukan Digital, Focal Loss

ABSTRACT

Student Name / NPM : Viviana Purba / 22082010152
Thesis Title : Analisis Emosi Berbasis Aspek Pada Ulasan Aplikasi Identitas Kependudukan Digital Menggunakan Indobert
Advisor : 1. Eka Dyar Wahyuni S.Kom, M.Kom
2. Tri Luhur Indayanti Sugata S.ST, M.IM

ABSTRACT

The Identitas Kependudukan Digital (IKD) application launched by Ditjen Dukcapil Kemendagri has been downloaded by more than 10 million users on Google Play Store with a rating of 3.2 from approximately 70,700 reviews. User reviews serve as a significant indicator reflecting service quality and influencing public trust and acceptance of digital government services, while also serving as an essential resource for developers in identifying improvement priorities. However, general sentiment analysis is considered inadequate as it is unable to specifically identify the aspects and sources of issues discussed in reviews. This study applies Aspect-Based Emotion Analysis (ABEA) with multilabel classification using an end-to-end IndoBERT model on Indonesian-language reviews collected from Google Play Store covering the period of June 2024 to November 2025. To identify the dominant service aspects, two topic modeling methods were compared, namely Latent Dirichlet Allocation (LDA) and BERTopic. Evaluation results show that BERTopic outperformed LDA with a coherence score of 0.6196 and diversity of 0.92, compared to LDA which yielded a coherence score of 0.5083 and diversity of 0.9333. BERTopic successfully identified five more specific and granular aspects: Login & Akses Akun, Scan Barcode ke Dukcapil, Verifikasi Foto Wajah, Dokumen & Layanan Digital, and Kompatibilitas Perangkat Android. These aspects were subsequently used in manual data labeling by three human annotators and two AI annotators (Claude and ChatGPT), with inter-annotator agreement measured using Krippendorff's Alpha ranging from 0.71 to 0.73. The classification model was developed through eight experimental scenarios that progressively combined preprocessing, loss functions (standard BCE, Weighted BCE with pos_weight, and Focal Loss), threshold tuning, and variations in data split ratios. The best-performing model used a combination of preprocessing, Focal Loss, threshold of 0.4, and a 60:20:20 split ratio, achieving an F1 Score Macro of 0.3916 (a 24.1% improvement over baseline), F1 Score Micro of 0.9134, Recall of 0.9423, and Hamming Loss of 0.0308. As a final output, an aspect-based emotion classifier website was developed using the Flask framework to present complex classification results into informative and more easily understood information.

Keywords: Aspect-Based Emotion Analysis, IndoBERT, Multilabel Classification, BERTopic, LDA, Identitas Kependudukan Digital, Focal Loss

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala berkat dan penyertaan-Nya, sehingga skripsi dengan judul “Analisis Emosi Berbasis Aspek pada Ulasan Aplikasi Identitas Digital Menggunakan IndoBERT” ini dapat terselesaikan dengan baik.

Perjalanan dalam menyusun tugas akhir ini bukanlah proses yang mudah bagi penulis. Banyak hal yang dipelajari, dilewati, dan diperjuangkan, mulai dari awal penyusunan hingga penyelesaian skripsi ini. Namun, di balik setiap proses tersebut, selalu ada doa, dukungan, serta kehadiran orang-orang baik yang membuat penulis mampu bertahan dan menyelesaikan skripsi ini hingga akhir. Oleh karena itu, penulis ingin menyampaikan apresiasi dan terima kasih yang tulus kepada:

1. Ibu Eka Dyar Wahyuni, S.Kom., M.Kom. selaku dosen pembimbing utama, atas kesabaran, arahan, dan ilmu yang telah diberikan hingga penulis dapat menyelesaikan skripsi ini.
2. Ibu Tri Luhur Indayanti Sugata S.ST, M.IM selaku dosen pembimbing kedua, atas segala masukan, motivasi, dan bimbingan yang telah membantu penulis menyempurnakan penelitian ini.
3. Bapak Tri Lathif Mardi Suryanto, S.Kom, M.T. selaku dosen wali, yang telah memberikan bimbingan akademik serta dukungan selama masa studi penulis di Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Keluarga besar penulis yang telah memberikan doa dan dukungan sehingga penulis bisa sampai di titik ini. Terlebih untuk Mamake dan Kak Sari, yang selalu menjadi tempat pulang bagi penulis di saat lelah, hilang semangat, dan takut gagal. Terima kasih karena selalu hadir dan tetap percaya, bahkan ketika penulis meragukan diri sendiri.
5. Mifa, Rahayu, dan Mafda, sobat *kuman-kuman*, yang telah memberikan banyak warna indah dalam perjalanan perkuliahan penulis. Terima kasih telah menjadi tempat nyaman untuk berbagi suka duka, serta saling menguatkan dari awal hingga selesainya masa studi ini.
6. Ferinta dan Effa, sobat *kara yara*. Terima kasih telah setia menemani penulis dan membuat perjalanan yang tidak mudah ini terasa lebih ringan dan menyenangkan.

7. Keluarga Opung Teresha. Terima kasih telah bersedia menjadi kerabat terdekat yang memberikan kehangatan dan rasa aman bagi penulis selama berjuang di tanah perantauan.
8. Mba Putri, Putan, Angel, Diana, dan Hilya, yang turut hadir menjadi bagian dari cerita dan perjalanan penulis selama masa kuliah. Terima kasih atas dukungan dan kebersamaan yang diberikan di tengah proses panjang penyusunan skripsi ini.
9. Indah, Kak Opa, Kak Cebe, dan Kak Jose. Terima kasih untuk senantiasa menyemangati penulis dalam menjalani proses perkuliahan hingga penyusunan skripsi ini.
10. Seluruh pihak yang tidak dapat disebutkan satu per satu, yang telah membantu baik secara langsung maupun tidak langsung dalam penyelesaian skripsi ini.
11. Terakhir, diri sendiri. Terima kasih karena tetap melangkah dan mau berproses hingga tiba di garis akhir perjalanan panjang ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi perbaikan di masa mendatang. Akhir kata, semoga karya ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan, khususnya di bidang Sistem Informasi.

Surabaya, 5 Mei 2026

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI	vii
ABSTRAK	ix
ABSTRACT	xi
KATA PENGANTAR	xiii
DAFTAR ISI	xv
DAFTAR GAMBAR	xix
DAFTAR TABEL	xxi
DAFTAR LAMPIRAN	xxiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah	4
1.4. Tujuan Penelitian	4
1.6. Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7
2.1. Dasar Teori	7
2.1.1. Identitas Kependudukan Digital (IKD)	7
2.1.2. Analisis Emosi	8
2.1.3. Analisis Emosi Berbasis Aspek	9
2.1.4. Klasifikasi Multilabel	10
2.1.5. <i>Latent Dirichlet Allocation (LDA)</i>	11
2.1.6. BERTopic	12
2.1.7. IndoBERT	13
2.1.8. Krippendorff's Alpha	14

2.1.9.	<i>Coherence Score</i>	15
2.1.10.	<i>Diversity</i>	16
2.1.11.	Confusion Matrix.....	16
2.1.12.	F1 Score Micro	17
2.1.13.	F1 Score Macro	18
2.1.14.	<i>Hamming Loss</i>	19
2.1.15.	<i>Cost-sensitive learning</i>	19
2.1.16.	Focal Loss.....	20
2.2.	Penelitian Terdahulu.....	21
BAB III METODOLOGI PENELITIAN		25
3.1.	Studi Literatur.....	26
3.2.	Analisis Kebutuhan	26
3.2.1.	Kebutuhan Perangkat Keras	26
3.2.2.	Kebutuhan Perangkat Lunak	26
3.2.3.	Kebutuhan Data	27
3.3.	Pengumpulan Data.....	27
3.4.	Eksplorasi Data.....	27
3.5.	Pembersihan Data	27
3.6.	<i>Filtering Data</i>	28
3.7.	<i>Preprocessing</i>	28
3.8.	Pemodelan Topik.....	29
3.8.1.	Pemodelan Topik Menggunakan LDA.....	30
3.8.2.	Pemodelan Topik Menggunakan BERTopic.....	30
3.9.	Evaluasi dan Interpretasi Hasil Pemodelan Topik.....	31
3.10.	Pelabelan Data	31
3.10.1.	Pelabelan Manual	31

3.10.2.	Transformasi Label	32
3.11.	Pembagian Data	32
3.12.	Pengembangan Model Klasifikasi.....	33
3.13.	Evaluasi Model.....	35
3.14.	Penyimpanan Model.....	36
3.15.	Deployment.....	36
3.16.	Validasi Model di Sistem.....	37
BAB IV HASIL DAN PEMBAHASAN		39
4.1.	Hasil	39
4.1.1.	Studi Literatur	39
4.1.2.	Analisis Kebutuhan	40
4.1.3.	Pengumpulan Data	40
4.1.4.	Eksplorasi Data	42
4.1.5.	Pembersihan Data.....	45
4.1.6.	<i>Filtering</i> Data	46
4.1.7.	<i>Preprocessing</i> Data	48
4.1.8.	Pemodelan Topik	53
4.1.9.	Evaluasi dan Interpretasi Hasil Pemodelan Topik	60
4.1.10.	Pelabelan Data.....	65
4.1.11.	Pembagian Data	69
4.1.12.	Pengembangan Model Klasifikasi.....	69
4.1.13.	Evaluasi Model.....	77
4.1.14.	Penyimpanan Model.....	79
4.1.15.	Deployment.....	80
4.1.16.	Validasi Model di Sistem.....	83
4.2.	Pembahasan.....	85

4.2.1.	Perbandingan Metode Pemodelan Topik LDA dan BERTopic.....	85
4.2.2.	Pengembangan Model Klasifikasi Emosi Berbasis Aspek.....	86
4.2.3.	Implementasi Sistem Berbasis Web	90
BAB V KESIMPULAN		93
5.1.	Kesimpulan.....	93
5.2.	Saran	94
DAFTAR PUSTAKA.....		95
LAMPIRAN		99

DAFTAR GAMBAR

Gambar 2.1.	Aplikasi Identitas Kependudukan Digital	7
Gambar 3.1.	Alur Metodologi Penelitian [7], [31]	25
Gambar 3.2.	Tahapan <i>Preprocessing</i>	28
Gambar 3.3.	<i>End-to-end</i> IndoBERT Multilabel Untuk ABEA [5]	34
Gambar 3.4.	<i>Wireframe</i> Halaman Home	36
Gambar 3.5.	<i>Wireframe</i> Halaman Hasil Analisis	37
Gambar 4.1.	Kode <i>Scrapping</i> Ulasan IKD	41
Gambar 4.2.	Data Ulasan IKD	42
Gambar 4.3.	Kode Analisis Deskriptif Dataset	42
Gambar 4.4.	Distribusi <i>Rating</i>	43
Gambar 4.5.	<i>Word Cloud</i> Ulasan	43
Gambar 4.6.	Distribusi Frekuensi Jumlah Kata Per Ulasan	44
Gambar 4.7.	Rata-rata Jumlah Kata Per Ulasan	44
Gambar 4.8.	Data Ulasan Setelah Pembersihan	45
Gambar 4.9.	Kode <i>Filtering</i> Data	47
Gambar 4.10.	Kode <i>Lowercase, Tokenizing, dan Normalization</i>	48
Gambar 4.11.	Kode <i>Stopword Removal dan Stemming</i>	49
Gambar 4.12.	Kode Implementasi dan Evaluasi LDA	53
Gambar 4.13.	Kode Implementasi dan Evaluasi BERTopic	56
Gambar 4.14.	BERTopic <i>Cosine Similarity Heatmap</i>	58
Gambar 4.15.	BERTopic Hierarchical Clustering	59
Gambar 4.16.	<i>Intertopic Distance Map</i> LDA	62
Gambar 4.17.	<i>Intertopic Distance Map</i> BERTopic	63
Gambar 4.18.	Kode Menghitung Krippendorff's Alpha	66
Gambar 4.19.	Kode Transformasi Label	68

Gambar 4.20.	Pembagian Data	69
Gambar 4.21.	Kode <i>End-to-end</i> IndoBERT	70
Gambar 4.22.	Hasil Implementasi <i>Baseline Model</i>	70
Gambar 4.23.	Kode Preprocess Teks	71
Gambar 4.24.	Kode Implementasi <i>pos_weight</i>	72
Gambar 4.25.	Hasil Implementasi <i>pos_weight</i> dengan Threshold Optimal.....	73
Gambar 4.26.	Kode Implementasi Focal Loss	74
Gambar 4.27.	Hasil Implementasi Focal Loss dengan Threshold Optimal.....	75
Gambar 4.28.	Kode Split Data 60:20:20	76
Gambar 4.29.	Hasil Implementasi Skenario Kombinasi	77
Gambar 4.30.	Kode Menyimpan Model.....	80
Gambar 4.31.	Kode Implementasi Flask dan <i>Load Model</i>	81
Gambar 4.32.	Halaman Beranda	81
Gambar 4.33.	Halaman Hasil Analisis (Diagram).....	82
Gambar 4.34.	Halaman Hasil Analisis (Tabel)	83
Gambar 4.35.	Halaman Panduan	83

DAFTAR TABEL

Tabel 2.1	Penelitian Terdahulu 1	21
Tabel 2.2	Penelitian Terdahulu 2	22
Tabel 2.3	Penelitian Terdahulu 3	22
Tabel 2.4	Penelitian Terdahulu 4	22
Tabel 2.5	Penelitian Terdahulu 5	23
Tabel 3.1	Contoh Transformasi Label.....	32
Tabel 3.1	Skenario Uji	35
Tabel 4.1	Contoh Ulasan yang Tidak Relevan.....	47
Tabel 4.2	Hasil Lowercase	49
Tabel 4.3	Hasil Tokenizing	50
Tabel 4.4	Hasil Normalization	51
Tabel 4.5	Hasil <i>Stopword Removal</i>	52
Tabel 4.6	Hasil <i>Stemming</i>	52
Tabel 4.7	<i>Coherence Score LDA</i>	54
Tabel 4.8	Daftar Kata pada LDA Paling Optimal.....	54
Tabel 4.9	<i>Coherence Score BERTopic</i>	56
Tabel 4.10	Daftar Kata pada BERTopic Paling Optimal.....	57
Tabel 4.11	Daftar Kata pada Topik BERTopic Setelah Merge.....	60
Tabel 4.12	Hasil Perbandingan <i>Coherence Score</i> LDA dan BERTopic.....	61
Tabel 4.13	Interpretasi Topik LDA.....	63
Tabel 4.14	Interpretasi Topik BERTopic.....	64
Tabel 4.15	Hasil Skor Perhitungan Krippendorff's Alpha.....	66
Tabel 4.16	Distribusi Label pada Setiap Aspek	67
Tabel 4.17	Hasil Evaluasi Model pada Seluruh Skenario	78
Tabel 4.18	Hasil Evaluasi Model di Sistem	84

DAFTAR LAMPIRAN

Lampiran 1. Email Permohonan Izin Penelitian	99
Lampiran 2. Surat Permohonan Izin Penelitian	100
Lampiran 3. Hasil Scrapping dan Hasil Pelabelan.....	101
Lampiran 4. Panduan Anatator	101
Lampiran 5. Google Collab Notebook.....	101
Lampiran 6. Github Deployment	101