



SKRIPSI

***ASPECT-BASED SENTIMENT ANALYSIS
PADA RESPONS SURVEI OPEN-ENDED
MENGUNAKAN LDA, BERT, DAN SVM***

DIAN RAHMAWATI

NPM 21082010100

DOSEN PEMBIMBING

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

Dhian Satria Yuda Kartika, S.Kom., M.Kom

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SISTEM INFORMASI
SURABAYA
2025**



SKRIPSI

***ASPECT-BASED SENTIMENT ANALYSIS
PADA RESPONS SURVEI OPEN-ENDED
MENGUNAKAN LDA, BERT, DAN SVM***

DIAN RAHMAWATI
NPM 21082010100

DOSEN PEMBIMBING
Eka Dyar Wahyuni, S.Kom., M.Kom
Dhian Satria Yuda Kartika, S.Kom., M.Kom

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025**

Halaman ini sengaja dikosongkan

LEMBAR PENGESAHAN SKRIPSI

ASPECT-BASED SENTIMENT ANALYSIS PADA RESPONS SURVEI OPEN-ENDED MENGGUNAKAN LDA, BERT, DAN SVM

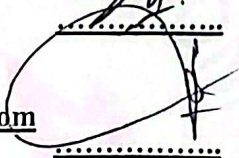
Oleh:
DIAN RAHMAWATI
NPM. 21082010100

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Sistem Informasi Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 07 Maret 2025.

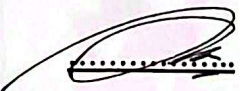
Eka Dyar Wahyuni, S. Kom., M. Kom
NIP. 198412012021212005


..... (Pembimbing I)

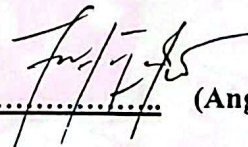
Dhian Satria Yuda Kartika, S.Kom, M.Kom
NIP. 20119860522249


..... (Pembimbing II)

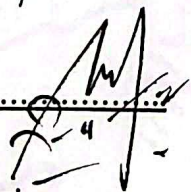
Nur Cahyo Wibowo, S. Kom., M. Kom
NIP. 197903172021211002


..... (Ketua Penguji)

Seftin Fitri Ana Wati, S. Kom., M. Kom.
NPT. 21219910320267


..... (Anggota Penguji II)

Anindo Saka Fitri, S. Kom., M. Kom.
NIP. 199303252024062001


..... (Anggota Penguji III)

Mengetahui,
Dekan Fakultas Ilmu Komputer


Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 196811261994032001

Halaman ini sengaja dikosongkan

LEMBAR PERSETUJUAN

**ASPECT-BASED SENTIMENT ANALYSIS PADA RESPONS SURVEI
OPEN-ENDED MENGGUNAKAN LDA, BERT, DAN SVM**

Oleh:

DIAN RAHMAWATI

NPM. 21082010100



Menyetujui,

**Koordinator Program Studi
Sistem Informasi Fakultas Ilmu Komputer**

A handwritten signature in black ink, appearing to be 'Agung Brastama Putra', is written over the text of the coordinator's name and NIP.

Agung Brastama Putra, S.Kom., M.Kom.
NIP. 198511242021211003

Halaman ini sengaja dikosongkan

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan dibawah ini :

Nama : Dian Rahmawati
NPM : 21082010100
Program : Sarjana (S1)
Program Studi : Sistem Informasi
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, 18 Maret 2025

Yang Membuat pernyataan



Dian Rahmawati

21082010100

Halaman ini sengaja dikosongkan

ABSTRAK

Nama Mahasiswa / NPM : Dian Rahmawati / 21082010100
Judul Skripsi : *Aspect-Based Sentiment Analysis* Pada Respons
Survei *Open-Ended* Menggunakan LDA, BERT,
dan SVM
Dosen Pembimbing : 1. Eka Dyar Wahyuni, S.Kom., M.Kom
2. Dhian Satria Yuda Kartika, S.Kom., M.Kom

Analisis sentimen berbasis aspek pada respons survei *open-ended* berperan penting dalam memahami opini dan pengalaman peserta suatu acara. Penelitian ini bertujuan untuk mengklasifikasikan sentimen dan mengidentifikasi aspek utama dalam respons survei *open-ended* dari acara seminar nasional FKBM-IK 2021 dan 2022. Metode yang digunakan mencakup *Latent Dirichlet Allocation* (LDA) untuk *topic modeling*, serta *Bidirectional Encoder Representations from Transformers* (BERT) untuk *word embedding*. Model klasifikasi sentimen dikembangkan menggunakan *Support Vector Machine* (SVM). Hasil penelitian menunjukkan bahwa model SVM dengan *word embedding* IndoBERT dan teknik *resampling Synthetic Minority Over-sampling Technique* (SMOTE) mencapai akurasi optimal dengan rata-rata skor *accuracy* sebesar 94%, *precision* 89%, *recall* 93%, *f1-score* 91%. Analisis *topic modeling* dengan LDA berhasil mengidentifikasi 7 topik utama yang mencerminkan aspek penting dari respons peserta. Mayoritas sentimen yang ditemukan bersifat positif, terutama dalam topik kesuksesan acara dan harapan acara.

Kata kunci : ABSA, LDA, *Word Embedding*, SVM, survei *open-ended*.

Halaman ini sengaja dikosongkan

ABSTRACT

Student Name / NPM : Dian Rahmawati / 21082010100
Thesis Title : *Aspect-based Sentiment Analysis on Open Survey Responses Using LDA, BERT, and SVM*
Advisor : 1. Eka Dyar Wahyuni, S.Kom, M.Kom.
2. Dhian Satria Yuda Kartika, S.Kom, M.Kom

Aspect-based sentiment analysis on open-ended survey responses is important in understanding the opinions and experiences of participants in an event. This research aims to classify sentiment and identify key aspects in open-ended survey responses from the 2021 and 2022 FKBM-IK national seminar events. The methods used include Latent Dirichlet Allocation (LDA) for topic modeling, as well as Bidirectional Encoder Representations from Transformers (BERT) for word embedding. A sentiment classification model was developed using Support Vector Machine (SVM). The results showed that the SVM model with IndoBERT word embedding and Synthetic Minority Over-sampling Technique (SMOTE) resampling technique achieved optimal accuracy with an average accuracy score of 94%, precision 89%, recall 93%, f1-score 91%. Topic modeling analysis with LDA successfully identified 7 main topics that reflected important aspects of the participants' responses. The majority of sentiments found were positive, especially in the topics of event success and event expectation.

Keywords: ABSA, LDA, Word Embedding, SVM, survei open-ended.

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala rahmat, hidayah dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul **“Aspect-Based Sentiment Analysis Pada Respons Survei Open-Ended Menggunakan LDA, BERT, dan SVM”**. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana pada Program Studi Sistem Informasi Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Proses penyusunan skripsi ini merupakan perjalanan yang penuh dengan pembelajaran, tantangan, serta momen berharga. Keberhasilan dalam menyelesaikan penelitian ini tidak terlepas dari dukungan berbagai pihak. Oleh karena itu, dengan penuh rasa syukur, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua penulis, atas segala kasih sayang, dukungan, kepercayaan, serta doa yang tiada henti, sehingga menjadi kekuatan utama bagi penulis dalam menyelesaikan skripsi.
2. Kakak Perempuan dan Laki-Laki penulis yang senantiasa memberikan doa, semangat, dan dukungan moral tanpa henti selama proses penyusunan skripsi.
3. Ibu Eka Dyar Wahyuni, S.Kom., M.Kom., selaku dosen wali dan pembimbing pertama, yang telah memberikan dukungan penuh dalam penyelesaian skripsi ini. Beliau senantiasa memberikan arahan, saran, serta solusi terhadap berbagai permasalahan yang muncul selama proses penyusunan skripsi.
4. Bapak Dhian Satria Yuda Kartika, S.Kom., M.Kom., selaku dosen pembimbing kedua, yang telah meluangkan waktu, memberikan dukungan, arahan, serta berbagi pengetahuan selama proses penyusunan skripsi, sehingga penulis dapat menyelesaikan skripsi dengan baik.
5. Bapak Agung Brastama Putra, S.Kom., M.Kom., selaku Koordinator Program Studi Sistem Informasi, yang senantiasa memberikan kemudahan dan dukungan bagi para mahasiswa.
6. Seluruh Bapak dan Ibu dosen Program Studi Sistem Informasi UPN “Veteran” Jawa Timur yang telah berbagi ilmu pengetahuan yang berharga selama masa perkuliahan.

7. Seluruh teman-teman Program Studi Sistem Informasi angkatan 2021, terutama kelas C yang telah bersama-sama menjalani perjuangan sejak awal hingga akhirnya mencapai tahap ini.
8. Sahabat-sahabat penulis di berbagai jenjang pendidikan, mulai dari SD, SMP, SMA, hingga perkuliahan yang selalu memberikan dukungan, semangat, dan kebersamaan. Terima kasih atas motivasi dan bantuan yang berarti dalam perjalanan menyelesaikan skripsi.
9. Semua pihak yang tidak dapat disebutkan satu per satu, tetapi telah memberikan bantuan dan dukungan dalam penyelesaian skripsi ini.
10. Terima kasih juga kepada diri sendiri atas perjuangan, ketekunan, dan keteguhan dalam menghadapi berbagai tantangan hingga akhirnya skripsi ini dapat diselesaikan.

Sebagai penutup, penulis menyadari bahwa skripsi ini masih memiliki kekurangan. Oleh karena itu, penulis sangat mengharapkan masukan, kritik, dan saran yang membangun. Semoga skripsi ini dapat memberikan kontribusi nyata bagi pengembangan ilmu pengetahuan, khususnya di bidang sistem informasi.

Surabaya, 10 Maret 2025

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	iii
LEMBAR PERSETUJUAN.....	v
SURAT PERNYATAAN BEBAS PLAGIASI	vii
ABSTRAK	ix
ABSTRACT	xi
KATA PENGANTAR	xiii
DAFTAR ISI.....	xv
DAFTAR GAMBAR	xix
DAFTAR TABEL.....	xxiii
DAFTAR NOTASI	xxv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	5
1.3. Batasan Masalah.....	5
1.4. Tujuan Penelitian.....	6
1.5. Manfaat Penelitian.....	6
BAB II TINJAUAN PUSTAKA.....	7
2.1. Penelitian Terdahulu.....	7
2.2. Landasan Teori	9
2.2.1. Survei <i>Open-Ended</i>	10
2.2.2. Text Mining.....	10
2.2.3. Natural Language Processing (NLP)	11
2.2.4. <i>Aspect-Based Sentiment Analysis</i> (ABSA)	12
2.2.5. Word Embedding	12
2.2.6. Latent Dirichlet Allocation (LDA).....	14

2.2.7.	Support Vector Machine (SVM)	16
2.2.8.	VADER	17
2.2.9.	TextBlob	18
2.2.10.	Confusion Matrix	19
2.2.11.	Flask.....	20
2.2.12.	Fleiss' Kappa	21
BAB III DESAIN DAN IMPLEMENTASI SISTEM.....		23
3.1.	Metode Penelitian	23
3.2.	Studi Literatur	24
3.3.	Analisis Kebutuhan.....	24
3.3.1.	Kebutuhan Data	24
3.3.2.	Kebutuhan Software dan Hardware.....	24
3.4.	Data Collecting	24
3.5.	Exploratory Data Analysis (EDA)	25
3.6.	Filtering Data	25
3.7.	Pelabelan Sentimen.....	25
3.8.	Data Preprocessing	26
3.8.1.	Data Cleaning	27
3.8.2.	Case Folding.....	27
3.8.3.	Tokenization	27
3.8.4.	Normalization	27
3.8.5.	Stopword Removal	28
3.9.	Model Analisis Sentimen Berbasis Aspek (ABSA)	28
3.9.1.	Topic Modeling	29
3.9.2.	Word Embedding.....	30
3.9.3.	Klasifikasi Sentimen Menggunakan <i>Support Vector Machine</i>	30

3.9.4.	Analisis Sentimen Berbasis Aspek (ABSA)	32
3.10.	Visualisasi.....	32
3.11.	Kesimpulan	34
BAB IV PENGUJIAN DAN ANALISA		35
4.1.	Implementasi Kebutuhan.....	35
4.1.1.	Kebutuhan Data.....	35
4.1.2.	Kebutuhan Software dan Hardware	35
4.2.	Data Collecting.....	35
4.3.	Exploratory Data Analysis (EDA)	37
4.4.	Filtering Data.....	38
4.5.	Pelabelan Sentimen	38
4.5.1.	Translate Teks	39
4.5.2.	Skenario Pelabelan Sentimen.....	39
4.5.3.	Sentimen Akhir	46
4.6.	Data Preprocessing	47
4.6.1.	Data Cleaning.....	47
4.6.2.	Case Folding	49
4.6.3.	Tokenization.....	50
4.6.4.	Normalization.....	50
4.6.5.	Stopword Removal.....	53
4.6.6.	Word Cloud.....	55
4.7.	Model Analisis Sentimen Berbasis Aspek (ABSA)	56
4.7.1.	Topic Modeling.....	56
4.7.2.	Word Embedding	64
4.7.3.	Klasifikasi Sentimen Menggunakan SVM.....	65
4.7.4.	Analisis Sentimen Berbasis Aspek (ABSA)	70

4.8. Visualisasi	72
BAB V KESIMPULAN DAN SARAN	75
5.1. Kesimpulan	75
5.2. Saran Pengembangan	76
DAFTAR PUSTAKA.....	77

DAFTAR GAMBAR

Gambar 1.1 Formulir <i>Feedback</i> dan Distribusi Penggunaan Survei Terpopuler di Indonesia [2].....	1
Gambar 1.2 Form <i>Feedback Open-Ended</i> (FKBM-IK, 2022).....	2
Gambar 2.1 Proses Text Mining (Vasili et al., 2018)	11
Gambar 2.3 Teknik Word Embedding BERT [34]	14
Gambar 2.4 Representasi Model LDA [15]	15
Gambar 2.6 Ilustrasi Linear SVM [19]	17
Gambar 3.1 Alur Metode Penelitian	23
Gambar 3.2 Alur Pelabelan Sentimen.....	26
Gambar 3.3 Alur Data Preprocessing.....	27
Gambar 3.4 Alur <i>Topic Modeling</i>	29
Gambar 3.5 Alur Klasifikasi Sentimen	31
Gambar 3.6 Desain Tampilan Website: (a) Opsi Teks Analisis, (b) Opsi Unggah File, (c) Output Hasil Analisis	34
Gambar 4.1. <i>Feedback</i> Seminar Nasional FKBM-IK: (a) 2021, (b) 2022.....	36
Gambar 4.2. Dataset Penelitian.....	37
Gambar 4.3. Word Cloud Dataset Awal	37
Gambar 4.4. <i>Word Cloud</i> Setelah <i>Filtering Data</i>	38
Gambar 4.5 Kode Translate Teks.....	39
Gambar 4.6 Kode Vader Menggunakan Tiga Kelas Sentimen	40
Gambar 4.7 Kode Textblob Menggunakan Tiga Kelas Sentimen	40
Gambar 4.8 Kode Cardiff Menggunakan Tiga Kelas Sentimen	40
Gambar 4.9 Kode Distillbert Menggunakan Tiga Kelas Sentimen.....	41
Gambar 4.10 Percobaan Pertama Fleiss' Kappa Sebelum Penambahan Data	42
Gambar 4.11 Percobaan Kedua Fleiss' Kappa Sesudah Penambahan Data	43

Gambar 4.12 Kode Vader Menggunakan Dua Kelas Sentimen	43
Gambar 4.13 Kode DistilBERT Menggunakan Dua Kelas Sentimen.....	44
Gambar 4.14 Percobaan Ketiga Fleiss' Kappa Sebelum Penambahan Data	45
Gambar 4.15 Percobaan Keempat Fleiss' Kappa Sesudah Penambahan Data.....	45
Gambar 4.16 Word Cloud Data Sebelum <i>Preprocessing</i>	47
Gambar 4.17 Kode untuk Data Cleaning	48
Gambar 4.18 Potongan Kode untuk Case Folding	49
Gambar 4.19 Potongan Kode untuk Tokenization	50
Gambar 4.20 Potongan Kode <i>Normalization</i> dengan Kamus Normalisasi	51
Gambar 4.21 Potongan Kode <i>Normalization</i> dengan Kamus Normalisasi dan Sinonim.....	53
Gambar 4.22 Potongan Kode Daftar <i>Stopwords</i>	54
Gambar 4.23 Potongan Kode untuk Menghapus <i>Stopwords</i>	54
Gambar 4.24 <i>Word Cloud</i> Setelah <i>Data Preprocessing</i>	55
Gambar 4.25 Import library untuk Topic Modeling.....	56
Gambar 4.26 Potongan Kode Mengubah Format Token ke List.....	57
Gambar 4.27 Potongan Kode Membuat Kamus dan Korpus untuk LDA	57
Gambar 4.28 Potongan Kode TF-IDF untuk Vektorisasi.....	57
Gambar 4.29 Potongan Kode Topic Modeling Skenario (a)	58
Gambar 4.30 Hasil Coherence untuk Skenario (a)	58
Gambar 4.31 Visualisasi LDA dengan 10 Topik pada Skenario (a)	59
Gambar 4.32 Potongan Kode Menghitung <i>Jaccard Similarity</i>	60
Gambar 4.33 Potongan Kode Menghitung Stabilitas LDA dan Rata-Rata	60
Gambar 4.34 Potongan Kode Menghitung Perbedaan Koherensi dan Stabilitas ..	60
Gambar 4.35 Metrik Model LDA Berdasarkan Jumlah Topik.....	61
Gambar 4.36 Visualisasi LDA dengan 7 Topik pada Skenario (b).....	62

Gambar 4.37 Potongan Kode untuk Import Library dan Load Model.....	65
Gambar 4.38 Potongan Kode untuk Word Embedding IndoBERT	65
Gambar 4.39 Potong Kode untuk Import Library dalam Pemodelan Klasifikasi SVM	66
Gambar 4.40 Potongan Kode Pembagian Data dengan dan Tanpa SMOTE	66
Gambar 4.41 Potong Kode untuk Melatih Model SVM	67
Gambar 4.42 <i>Confusion Matrix</i> SVM.....	67
Gambar 4.43 <i>Classification Report</i> SVM.....	68
Gambar 4.44 Data Akhir Hasil <i>Topic Modeling</i> dan Klasifikasi Sentimen.....	71
Gambar 4.45 (a) dan (b) Visualisasi Hasil Penerapan Model.....	73

Halaman ini sengaja dikosongkan

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	7
Tabel 2.2 Rentang Compound Score Vader.....	18
Tabel 2.3 Confusion Matrix	19
Tabel 2.4 Skor Fleiss' Kappa [47]	21
Tabel 4.1 Data yang dilakukan Filtering.....	38
Tabel 4.2 Translate Teks dari Indonesia ke Inggris	39
Tabel 4.3 Pelabelan Menggunakan Tiga Kelas Sentimen oleh Lima Pelabel	41
Tabel 4.4 Distribusi Tiga Kelas Sentimen untuk Menghitung Fleiss' Kappa.....	42
Tabel 4.5 Pelabelan Sentimen Dua Kelas oleh Tiga Pelabel	44
Tabel 4.6 Distribusi Dua Kelas Sentimen untuk Menghitung Fleiss' Kappa	44
Tabel 4.7 Perbandingan Nilai Fleiss' Kappa	46
Tabel 4.8 Hasil Akhir dengan Dua Kelas Sentimen	46
Tabel 4.9 Hasil Data Cleaning	49
Tabel 4.10 Hasil Case Folding	49
Tabel 4.11 Hasil Tokenization	50
Tabel 4.12 Kamus Normalisasi	51
Tabel 4.13 Hasil <i>Normalization</i> dengan Kamus Normalisasi.....	52
Tabel 4.14 Kamus Sinonim.....	52
Tabel 4.14 Hasil <i>Normalization</i> dengan Kamus Normalisasi dan Sinonim.....	53
Tabel 4.15 Hasil <i>Stopword Removal</i> dari Normalisasi dengan Satu Kamus	54
Tabel 4.16 Hasil <i>Stopword Removal</i> dari Normalisasi dengan Dua Kamus.....	55
Tabel 4.17 Perbandingan Stabilitas dan Koherensi untuk Skenario (b)	61
Tabel 4.18 Kata Kunci dan Interpretasi Aspek	63
Tabel 4.19 Data Feedback Acara Seminar Setelah Pelabelan Aspek	64
Tabel 4.20 Analisis Skenario SVM.....	69

Tabel 4.21 Data Feedback Acara Seminar Setelah Pelabelan Aspek.....	71
---	----

DAFTAR NOTASI

$f(x)$	:	Fungsi <i>output</i> atau prediksi nilai
w	:	<i>Weight</i> /koefisien
x	:	Vektor input atau data
b	:	Bias
K	:	Koefisien Kappa
p_a	:	<i>Observed agreement</i>
p_ε	:	<i>Expected agreement</i>
n	:	Jumlah data
m	:	Jumlah penilai
x_{ij}	:	Frekuensi kesepakatan antara penilai i dan kategori j
q_j	:	Proporsi rata-rata penilaian dalam kategori j

Halaman ini sengaja dikosongkan