



SKRIPSI

**ANALISIS SENTIMEN BERBASIS ASPEK PADA
ULASAN TEMPAT WISATA POPULER DI JAWA
TIMUR MENGGUNAKAN CNN**

SITI OKTAVIA EKA PUTRI

NPM 21082010081

DOSEN PEMBIMBING

Amalia Anjani Arifiyanti, S.Kom., M.Kom.

Abdul Rezha Efrat Najaf, 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

ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN TEMPAT WISATA POPULER DI JAWA TIMUR MENGGUNAKAN CNN

SITI OKTAVIA EKA PUTRI

NPM 21082010081

DOSEN PEMBIMBING

Amalia Anjani Arifiyanti, S.Kom., M.Kom.

Abdul Rezha Efrat Najaf, 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**

Halaman ini sengaja dikosongkan

LEMBAR PENGESAHAN

ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN TEMPAT WISATA POPULER DI JAWA TIMUR MENGGUNAKAN CNN

Oleh:

SITI OKTAVIA EKA PUTRI
NPM. 21082010081

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Sistem Informasi Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 10 Juni 2025.

Menyetujui

Amalia Anjani Arifiyanti, S.Kom., M.Kom. (Pembimbing I)
NIP. 19920812 201803 2 001

Abdul Rezha Efrat Najaf, S.Kom., M.Kom. (Pembimbing II)
NIP. 19940929 202203 1 008

Eka Dyar Wahyuni, S.Kom., M.Kom. (Ketua Penguji)
NIP. 19841201 202121 2 005

Reisa Permatasari, S.T., M.Kom. (Penguji II)
NIP. 19920514 202203 2 007

Tri Luhur Indayanti Sugata, M.IM. (Penguji III)
NIP. 19920616 202406 2 001

Mengetahui,
Dekan Fakultas Ilmu Komputer

Prof. Dr. Ir. Novina Hendrasarie, MT.
NIP. 19681126 1994032 001

Halaman ini sengaja dikosongkan

LEMBAR PERSETUJUAN

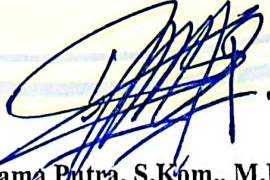
**ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN TEMPAT
WISATA POPULER DI JAWA TIMUR MENGGUNAKAN CNN**

Oleh:

**SITI OKTAVIA EKA PUTRI
NPM. 21082010081**

Menyetujui,

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



Agung Brastama Putra, S.Kom., M.Kom.
NIP. 19851124 202121 1 003

Halaman ini sengaja dikosongkan

SURAT PERNYATAAN BEBAS PLAGIASI

Yang bertanda tangan di bawah ini:

Nama : Siti Oktavia Eka Putri
NPM : 21082010081
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, 22 Juni 2025
Yang membuat pernyataan,



Siti Oktavia Eka Putri
NPM. 21082010081

Halaman ini sengaja dikosongkan

ABSTRAK

Nama Mahasiswa / NPM : Siti Oktavia Eka Putri/ 21082010081
Judul Skripsi : Analisis Sentimen Berbasis Aspek Pada Ulasan Tempat Wisata Populer Di Jawa Timur Menggunakan CNN
Dosen Pembimbing : 1. Amalia Anjani Arifiyanti, S.Kom., M.Kom.
2. Abdul Rezha Efrat Najaf, S.Kom., M.Kom.

Industri pariwisata merupakan salah satu sektor penting dalam perekonomian Indonesia yang menunjukkan tren pemulihan positif pasca pandemi COVID-19. Jawa Timur sebagai provinsi dengan tingkat kunjungan wisatawan tertinggi di Indonesia memiliki potensi besar untuk dianalisis lebih lanjut guna mendukung upaya pemulihan dan pengembangan pariwisata yang berkelanjutan. Penelitian ini bertujuan untuk mengimplementasikan deep learning dengan algoritma berbasis CNN untuk prediksi sentimen berbasis aspek pada ulasan tempat wisata populer di Jawa Timur serta mengetahui skenario seperti apa yang dapat memberikan kinerja terbaik. Aspek yang digunakan pada penelitian ini adalah attraction, amenities, access, dan price dengan data ulasan yang bersumber dari platform digital TripAdvisor dan Google Maps. Metode yang digunakan adalah Aspect-Based Sentiment Analysis (ABSA) dengan pendekatan deep learning menggunakan algoritma Convolutional Neural Network (CNN) yang juga menggunakan word embedding berbeda, yaitu Word2Vec dan BERT serta IndoBERT. Model yang dibangun berjumlah 5 model yang terdiri dari 1 model klasifikasi aspek (multilabel) dan 4 model sentimen untuk masing-masing aspek. Hasil penelitian menunjukkan bahwa model dengan embedding IndoBERT memberikan performa terbaik dengan F1-score dan AUC score berturut-turut: 1) Model klasifikasi aspek menghasilkan 0.81 dan 0.91, 2) Model klasifikasi sentimen Attraction menghasilkan 0.71 dan 0.94, 3) Model klasifikasi sentimen Amenities menghasilkan 0.85 dan 0.93, 4) Model klasifikasi sentimen Access menghasilkan 0.76 dan 0.83, 5) Model klasifikasi sentimen Price menghasilkan 0.81 dan 0.89. Setelah model terbaik dipilih, selanjutnya diaplikasikan ke aplikasi berbasis web untuk melakukan prediksi sentimen berbasis aspek terhadap ulasan wisata yang dapat menerima input single review serta batch review berupa file CSV. Hasil prediksi akan menampilkan sentimen pada masing-masing aspek yang terdeteksi, dan untuk batch review juga ditampilkan pie chart agar memudahkan penarikan kesimpulan soal data yang diprediksi.

Kata kunci: Analisis Sentimen Berbasis Aspek, CNN, Word2Vec, BERT, IndoBERT, Pariwisata

Halaman ini sengaja dikosongkan

ABSTRACT

Student Name / NPM : Siti Oktavia Eka Putri/ 21082010081
Thesis Title : Aspect-Based Sentiment Analysis on Popular Tourism Review of East Java Using CNN
Advisor : 1. Amalia Anjani Arifiyanti, S.Kom., M.Kom.
2. Abdul Rezha Efrat Najaf, S.Kom., M.Kom.

The tourism industry is one of the key sectors in Indonesia's economy, showing a positive recovery trend in the post-COVID-19 period. East Java, as the province with the highest number of tourist visits in Indonesia, holds great potential for further analysis to support efforts toward sustainable tourism recovery and development. This study aims to implement deep learning using a Convolutional Neural Network (CNN)-based algorithm to perform aspect-based sentiment prediction on reviews of popular tourist destinations in East Java and to identify the scenario that yields the best performance. The aspects used in this study are attraction, amenities, access, and price, with review data collected from digital platforms such as TripAdvisor and Google Maps. The method applied is Aspect-Based Sentiment Analysis (ABSA) using a deep learning approach with CNN, integrated with different word embeddings, namely Word2Vec, BERT, and IndoBERT. A total of five models were built: one multilabel aspect classification model and four sentiment classification models for each aspect. The results show that the model using IndoBERT embedding provided the best performance with F1-score and AUC as follows: 1) The aspect classification model achieved 0.81 and 0.91, 2) The sentiment classification model for Attraction achieved 0.71 and 0.94, 3) For Amenities, 0.85 and 0.93, 4) For Access, 0.76 and 0.83, and 5) For Price, 0.81 and 0.89. Once the best-performing model was selected, it was implemented into a web-based application to perform aspect-based sentiment predictions on tourism reviews, capable of accepting both single review input and batch reviews via CSV files. The prediction results display sentiment for each detected aspect, and for batch reviews, a pie chart is also provided to facilitate the interpretation of the predicted data.

Keywords: Aspect-based Sentiment Analysis, CNN, Word2Vec, BERT, IndoBERT, Tourism Review

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Segala puji bagi Tuhan Yang Maha Esa atas kesempatan dan kekuatan yang telah diberikan hingga skripsi yang berjudul “Analisis Sentimen Berbasis Aspek pada Ulasan Tempat Wisata Populer di Jawa Timur Menggunakan CNN” ini dapat diselesaikan dengan baik.

Penulis mengucapkan terima kasih banyak kepada Ibu Amalia Anjani Arifiyanti, S.Kom., M.Kom. selaku Dosen Pembimbing 1 serta Bapak Abdul Rezha Efrat Najaf, S.Kom., M.Kom. selaku Dosen Pembimbing 2 yang telah meluangkan waktunya untuk memberikan bimbingan, masukan, serta nasehat kepada penulis. Selain itu, penulis juga banyak menerima bantuan dari berbagai pihak:

1. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT., selaku Dekan Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur.
2. Bapak Agung Brastama Putra, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi, Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur serta selaku dosen wali yang selalu memberi arahan selama masa perkuliahan.
3. Bapak/Ibu dosen program studi Sistem Informasi UPN “Veteran” Jawa Timur yang telah memberikan ilmu yang sangat berharga sehingga penulis dapat menerapkannya pada penulisan skripsi ini serta praktik.
4. Orang tua penulis, yaitu Bapak Mustofa dan Ibu Siti Rohani, serta segenap keluarga yang selalu memberikan dukungan dan selalu mendoakan demi kelancaran skripsi penulis.
5. Teman-teman Sistem Informasi angkatan 2021 serta teman-teman seperjuangan penulis — Lusi, Sintia, Nasywa, Cendana, Izra, Aca, Nabila — yang selalu mendukung, memberi motivasi, dan pengalaman yang berharga selama penyusunan skripsi ini.
6. Teman seperjuangan penulis di kos — Windy, Iza, Devi, Annisa, Alvi — yang kebersamaan penulis di masa penyusunan skripsi.
7. Sahabat dekat penulis — Ilfiana, Adiesty, Manda, Adhiska, Rista, Vida, Dilla — yang selalu menemani di kala suka dan duka, mendukung, dan memberi semangat selama penyusunan skripsi.

Penulis menyadari bahwa penyusunan skripsi ini masih terdapat banyak kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhir kata, dengan segala keterbatasan ini semoga laporan skripsi ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, 23 Juni 2025

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.....	xix
DAFTAR TABEL	xxiii
DAFTAR NOTASI.....	xxv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	7
1.3 Batasan Masalah.....	7
1.4 Tujuan Penelitian	8
BAB II TINJAUAN PUSTAKA.....	9
2.1 Dasar Teori.....	9
2.1.1 Industri Pariwisata.....	9
2.1.2 Tripadvisor	10
2.1.3 Google Maps	11
2.1.4 Web-scraping	11
2.1.5 Analisis Sentimen Berbasis Aspek.....	12
2.1.6 Deep Learning.....	12
2.1.7 Convolutional Neural Network (CNN)	13
2.1.8 Fleiss' Kappa.....	16
2.1.9 Random Oversampling.....	17
2.1.10 Word Embedding.....	17
2.1.11 Word2Vec	18
2.1.12 BERT	18
2.1.13 IndoBERT	20

2.1.14	<i>Hyperparameter</i>	20
2.1.15	Bayesian Optimization	22
2.1.16	<i>K-fold Cross Validation</i>	22
2.1.17	ROC AUC	23
2.1.18	Confusion Matrix.....	23
2.1.19	Flask	24
2.2	Penelitian Terdahulu.....	24
BAB III METODE PENELITIAN.....		31
3.1	Studi Literatur.....	31
3.2	Analisis Kebutuhan	31
3.2.1	Kebutuhan Data	31
3.2.2	Kebutuhan <i>Software</i> dan <i>Hardware</i>	32
3.3	Pengumpulan Data.....	32
3.4	Pelabelan Data	34
3.5	<i>Exploratory Data Analysis</i> (EDA)	36
3.6	<i>Preprocessing</i> Data	36
3.6.1	Data Cleaning	37
3.6.2	Case Folding.....	37
3.6.3	Normalisasi.....	37
3.6.4	Stopwords Removal.....	38
3.6.5	<i>Stemming</i>	38
3.7	Pembagian Data.....	38
3.8	Pembuatan Model	38
3.8.1	Oversampling	41
3.8.2	Word Embedding.....	41
3.8.3	Arsitektur Model CNN	42
3.8.4	<i>Training</i> dan <i>Validation</i>	42
3.8.5	<i>Retraining</i> dan <i>Hyperparameter Tuning</i>	42
3.9	Pengujian Model.....	43
3.10	<i>Deployment</i>	43
BAB IV HASIL DAN PEMBAHASAN.....		45
4.1	Implementasi Kebutuhan.....	45

4.1.1	Kebutuhan Data.....	45
4.1.2	Kebutuhan <i>Software</i> dan <i>Hardware</i>	45
4.2	Pengumpulan Data	45
4.3	Pelabelan Data.....	47
4.4	<i>Exploratory Data Analysis</i>	54
4.4.1	Info Dataset	54
4.4.2	Distribusi Sentimen per Aspek	55
4.4.3	Distribusi Panjang Teks	57
4.4.4	Distribusi Lokasi yang Diulas.....	58
4.4.5	Wordcloud.....	59
4.5	<i>Preprocessing Data</i>	62
4.5.1	<i>Case Folding</i>	62
4.5.2	<i>Data Cleaning</i>	63
4.5.3	Normalisasi	65
4.5.4	Stopwords Removal	67
4.5.5	<i>Stemming</i>	68
4.5.6	Encoding	69
4.6	Pembagian Data	71
4.7	Pembuatan Model.....	72
4.7.1	Oversampling	73
4.7.2	Word Embedding	75
4.7.3	Arsitektur CNN	78
4.7.4	Training dan Validation	80
4.7.5	Retraining dan Tuning Hyperparameter.....	88
4.8	Pengujian Model	90
4.8.1.	Hasil Uji Model Aspek Multilabel.....	90
4.8.2.	Hasil Uji Model Sentimen Attraction	91
4.8.3.	Hasil Uji Model Sentimen Amenities	93
4.8.4.	Hasil Uji Model Sentimen Access	95
4.8.5.	Hasil Uji Model Sentimen Price	97
4.9	Deployment	99
4.10	Interpretasi Hasil	102

BAB V PENUTUP	109
5.1. Kesimpulan.....	109
5.2. Saran	110
DAFTAR PUSTAKA	111
LAMPIRAN	119

DAFTAR GAMBAR

Gambar 1. 1 Jumlah Kunjungan Wisatawan Mancanegara 2019-2024.....	2
Gambar 1. 2 Jumlah Kunjungan Wisata Nasional 2019-2024	2
Gambar 2. 1 Arsitektur CNN	13
Gambar 2. 2 Arsitektur CNN	15
Gambar 2. 3 Arsitektur BERT	19
Gambar 2. 4 Confussion Matrix.....	23
Gambar 3. 1 Diagram Alur Penelitian.....	31
Gambar 3. 2 Wireframe aplikasi ABSA – Single Review	44
Gambar 3. 3 Wireframe aplikasi ABSA – Batch Review	44
Gambar 4. 1 Contoh Data Hasil Scrapping.....	46
Gambar 4. 2 Code Translating Text.....	47
Gambar 4. 3 Contoh hasil textTranslated.....	47
Gambar 4. 4 Langkah-Langkah Anotasi	49
Gambar 4. 5 Contoh Pelabelan	51
Gambar 4. 6 <i>Dataset</i> dengan Label Akhir	52
Gambar 4. 7 Code Perhitungan Fleiss' Kappa.....	53
Gambar 4. 8 Hasil Perhitungan Fleiss' Kappa Aspek Attraction	53
Gambar 4. 9 Hasil Perhitungan Fleiss' Kappa Aspek Amenities	53
Gambar 4. 10 Hasil Perhitungan Fleiss' Kappa Aspek Access	54
Gambar 4. 11 Hasil Perhitungan Fleiss' Kappa Aspek Price	54
Gambar 4. 12 Output Info Dataset	54
Gambar 4. 13 Distribusi Sentimen Attraction.....	55
Gambar 4. 14 Distribusi Sentimen Amenities	56
Gambar 4. 15 Distribusi Sentimen Access.....	56
Gambar 4. 16 Distribusi Sentimen Price.....	57
Gambar 4.17 Distribusi Panjang Teks	58
Gambar 4. 18 Distribusi Lokasi	58
Gambar 4. 19 Wordcloud.....	59
Gambar 4. 20 Wordcloud Aspek Attraction	60
Gambar 4. 21 Wordcloud Aspek Amenities	60

Gambar 4. 22 Wordcloud Aspek Access.....	61
Gambar 4. 23 Wordcloud Aspek Price.....	62
Gambar 4. 24 Code Case Folding.....	63
Gambar 4. 25 Code <i>Data Cleaning</i>	64
Gambar 4. 26 Code Normalisasi	66
Gambar 4. 27 Code Stopwords Removal	67
Gambar 4. 28 Code Stemming	68
Gambar 4. 29 Code Encoding Aspek	69
Gambar 4. 30 Code Encoding Sentimen	70
Gambar 4. 31 Code Install Library.....	73
Gambar 4. 32 Code Random Oversampling Multilabel	73
Gambar 4. 33 Perbandingan Data Sebelum dan Setelah Oversampling.....	74
Gambar 4. 34 Code Random Oversampling Sentimen.....	75
Gambar 4. 35 Code Embedding Word2Vec.....	76
Gambar 4. 36 Output Embedding Word2Vec	76
Gambar 4. 37 Code Embedding BERT/IndoBERT.....	77
Gambar 4. 38 Output Embedding BERT	77
Gambar 4. 39 Output Embedding IndoBERT	78
Gambar 4. 40 Arsitektur CNN.....	78
Gambar 4. 41 Pseudocode Training Model	81
Gambar 4. 42 Classification Report Model Terbaik Klasifikasi Aspek	90
Gambar 4. 43 ROC AUC per Label	91
Gambar 4. 44 Classification Report Model Terbaik Klasifikasi Attraction	92
Gambar 4. 45 Confusion Matrix Model Terbaik Klasifikasi Attraction	92
Gambar 4. 46 ROC AUC Curve Model Terbaik Klasifikasi Attraction	93
Gambar 4. 47 Classification Report Model Terbaik Klasifikasi Amenities.....	93
Gambar 4. 48 ROC AUC Curve Model Terbaik Klasifikasi Amenities	94
Gambar 4. 49 Confusion Matrix Model Terbaik Klasifikasi Amenities	95
Gambar 4. 50 Classification Report Model Terbaik Klasifikasi Access.....	95
Gambar 4. 51 ROC AUC Curve Model Terbaik Klasifikasi Access	96
Gambar 4. 52 Confusion Matrix Model Terbaik Klasifikasi Access	96
Gambar 4. 53 Classification Report Model Terbaik Klasifikasi Price	97

Gambar 4. 54 Confusion Matrix Model Terbaik Klasifikasi Price	97
Gambar 4. 55 ROC AUC Curve Model Terbaik Klasifikasi Price	98
Gambar 4. 56 Tampilan Input Review ABSA	99
Gambar 4. 57 Hasil Prediksi Single Input.....	100
Gambar 4. 58 Hasil Prediksi Single Input dengan Bahasa Asing	100
Gambar 4. 59 Hasil Prediksi Batch Input dalam Tabel.....	100
Gambar 4. 60 Pie Chart Hasil Prediksi Batch Input.....	101
Gambar 4. 61 Output Kata Kunci Dominan.....	102

Halaman ini sengaja dikosongkan

DAFTAR TABEL

Tabel 3. 1 Jumlah Ulasan Wisata Jawa Timur pada Tripadvisor.....	32
Tabel 3. 2 Interpretasi Nilai Fleiss' Kappa	36
Tabel 3. 3 Skenario Penelitian	39
Tabel 4. 1 Penjelasan Masing-Masing Aspek	48
Tabel 4. 2 Data Case Folding	63
Tabel 4. 3 Data Cleaning.....	64
Tabel 4. 4 Data Normalisasi.....	66
Tabel 4. 5 Data Stopwords Removal.....	67
Tabel 4. 6 Data Stemming.....	69
Tabel 4. 7 Jumlah Data Aspek	70
Tabel 4. 8 Jumlah Data Sentimen	71
Tabel 4. 9 Pembagian Data Aspek	71
Tabel 4. 10 Pembagian Data Sentimen	72
Tabel 4. 11 Hasil Training dan Validasi Klasifikasi Aspek.....	82
Tabel 4. 12 Hasil Training dan Validasi Klasifikasi Sentimen Attraction.....	83
Tabel 4. 13 Hasil Training dan Validasi Klasifikasi Sentimen Amenities	84
Tabel 4. 14 Hasil Training dan Validasi Klasifikasi Sentimen Access.....	86
Tabel 4. 15 Hasil Training dan Validasi Klasifikasi Sentimen Price.....	87
Tabel 4. 16 Hasil Hyperparameter Tuning.....	89
Tabel 4. 17 Perbandingan Sebelum dan Sesudah Hyperparameter Tuning	105

Halaman ini sengaja dikosongkan

DAFTAR NOTASI

k	: Nilai Fleiss' Kappa (ukuran reliabilitas interrater)
p_i	: Tingkat kesepakatan antar-penilai untuk item ke- i
$\bar{P}$	: Rata-rata dari semua p_i (kesepakatan aktual)
$\bar{P}_e$	: Tingkat kesepakatan yang diharapkan

Halaman ini sengaja dikosongkan