

**KLASIFIKASI MULTI ASPEK SENTIMEN ANALIS PADA REVIEW
HOTEL MENGGUNAKAN INDOBERT EMBEDDING DAN CNN**

SKRIPSI



Oleh:

MUHAMMAD MAJDI

2010817210007

**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS TEKNIK
UNIVERSITAS LAMBUNG MANGKURAT
BANJARMASIN, DESEMBER 2024**

**KLASIFIKASI MULTI ASPEK SENTIMEN ANALIS PADA REVIEW
HOTEL MENGGUNAKAN INDOBERT EMBEDDING DAN CNN**

SKRIPSI

Diajukan Untuk Memenuhi Salah Satu Syarat

Sarjana Strata-1 Teknologi Informasi



Oleh:

MUHAMMAD MAJDI

2010817210007

PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS TEKNIK

UNIVERSITAS LAMBUNG MANGKURAT

BANJARMASIN, DESEMBER 2024

LEMBAR PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Muhammad Majdi
NIM : 2010817210007
Fakultas : Teknik
Prodi : Teknologi Informasi
Judul Skripsi : Klasifikasi Multi Aspek Sentimen Analisis Pada Review
Hotel Menggunakan IndoBERT Embedding dan CNN
Pembimbing : Andreyan Rizky Baskara, S.Kom., M.Kom.
Utama

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar akademik di suatu perguruan tinggi, dan sepanjang pengetahuan saya, juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali tertulis diacu dalam naskah ini dan disebutkan dalam daftar rujukan.

Banjarmasin, November 2024



Muhammad Majdi

NIM. 2010817210007

LEMBAR PENGESAHAN
SKRIPSI PROGRAM STUDI S-1 TEKNOLOGI INFORMASI


**Klasifikasi Multi Aspek Sentimen Analisis Pada Review Hotel Menggunakan IndoBERT
Embedding dan CNN**

Oleh
Muhammad Majdi (2010817210007)

Telah dipertahankan di depan Tim Penguji pada 10 Desember 2024 dan dinyatakan

LULUS

Komite Penguji :
Ketua : Muti'a Maulida, S.Kom, M.T.I
NIP. 198810272019032013
Anggota 1 : Muhammad Bahit, S.Kom., M.Eng
NIP. 198904162024211002
Anggota 2 : Dr. Ir. Yuslena Sari, S.Kom., M.Kom
NIP. 198411202015042002
**Pembimbing
Utama :** Andreyan Rizky Baskara, S.Kom., M.Kom
NIP. 199307032019031011




.....
.....
.....
.....

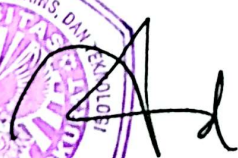
Banjarbaru, ... 27 DEC. 2024


Diketahui dan disahkan oleh:

Wakil Dekan Bidang Akademik
Fakultas Teknik ULM,

Koordinator Program Studi
S-1 Teknologi Informasi,




Dr. Mahmud, S.T., M.T.
NIP. 197401071998021001


Andreyan Rizky Baskara, S.Kom., M.Kom.
NIP. 199307032019031011

PERSETUJUAN SKRIPSI

LEMBAR PERSETUJUAN

KLASIFIKASI MULTI ASPEK SENTIMEN ANALIS PADA REVIEW
MENGUNAKAN INDOBERT EMBEDDING DAN CNN

OLEH
MUHAMMAD MAJDI
NIM. 2010817210007

Telah diperiksa dan terpenuhi semua persyaratan akademik, administrasi
dan disetujui untuk dipertahankan di hadapan dewan penguji

Banjarmasin, 29 November 2024

Pembimbing Utama,



Andreyan Rizky Baskara S.Kom., M.Kom.,
NIP. 199307032019031011

ABSTRAK

Internet menjadi sumber utama informasi bagi banyak orang, termasuk dalam memilih akomodasi hotel. Ulasan pengguna memberikan wawasan penting, namun analisis manual menjadi tidak efisien karena jumlahnya yang besar. Penelitian ini mengembangkan model berbasis *Convolutional Neural Network* (CNN) yang dilengkapi dengan *embedding* IndoBERT untuk mengklasifikasikan sentimen terhadap enam aspek: harga, hotel, kamar, lokasi, pelayanan, dan makanan. CNN dipilih karena kemampuannya dalam mengenali pola lokal dan fitur tekstual, namun memiliki keterbatasan dalam memahami konteks global dan hubungan semantik antar kata. Untuk itu, *embedding* IndoBERT digunakan untuk menghasilkan representasi teks yang kontekstual, membantu CNN memahami hubungan antar kata secara lebih mendalam. Hasil evaluasi menunjukkan bahwa model ini memiliki akurasi keseluruhan 92%, dengan F1-Score untuk sentimen negatif, netral, dan positif masing-masing 87%, 95%, dan 89%. *Precision* dan *recall* untuk sentimen negatif, netral, dan positif juga menunjukkan hasil yang baik. Namun, terdapat tantangan pada aspek hotel, yang mengalami kesalahan prediksi lebih tinggi akibat *overlapping sentiment*, ketidakseimbangan data, dan kompleksitas *multi* aspek. Perbandingan antara klasifikasi *single* aspek dan *multi* aspek menggunakan *K-fold* menunjukkan bahwa klasifikasi *single* aspek memiliki akurasi lebih tinggi sebesar 92,91% dibandingkan dengan *multi* aspek sebesar 90,94% dan F1-Score yang lebih rendah pada sentimen negatif, tetapi lebih tinggi pada sentimen lainnya. Penambahan *embedding* IndoBERT pada CNN terbukti efektif meningkatkan pemahaman konteks, terutama pada analisis sentimen *multi* aspek, dan dapat membantu manajemen hotel dalam meningkatkan kualitas layanan.

Kata kunci: Analisis sentimen, CNN, IndoBERT *embedding*, *multi* aspek, ulasan hotel

ABSTRACT

The internet has become a primary source of information for many, including when selecting hotel accommodations. User reviews provide valuable insights, but manual analysis becomes inefficient due to the large volume of data. This study develops a model based on Convolutional Neural Networks (CNN) enhanced with IndoBERT embeddings to classify sentiment across six aspects: price, hotel, room, location, service, and food. CNN is chosen for its ability to recognize local patterns and textual features, but it has limitations in understanding global context and semantic relationships between words. Therefore, IndoBERT embeddings are used to generate contextual text representations, helping CNN understand word relationships more deeply. Evaluation results show that the model achieves an overall accuracy of 92%, with F1-Scores for negative, neutral, and positive sentiments of 87%, 95%, and 89%, respectively. Precision and recall for each sentiment also show good results. However, challenges arise in the hotel aspect, which experiences higher prediction errors due to overlapping sentiment, data imbalance, and the complexity of multi-aspect contexts. A comparison between single-aspect and multi-aspect classification using K-fold cross-validation shows that single-aspect classification has higher accuracy (92.91%) compared to multi-aspect (90.94%), with a lower F1-Score for negative sentiment but higher for other sentiments. The addition of IndoBERT embeddings to CNN proves effective in enhancing context understanding, especially in multi-aspect sentiment analysis, and can assist hotel management in improving service quality.

Keywords: CNN, hotel reviews, IndoBERT embedding, multi-aspect, sentiment analysis

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur kepada Tuhan Yang Maha Esa, skripsi ini saya persembahkan kepada:

1. Ibu tercinta yang selalu menjadi penyemangat penulis, sebagai sandaran terkuat dari kerasnya dunia, yang tiada hentinya selalu memberikan kasih sayang dan do'a, dengan penuh keikhlasan yang tak terhingga kepada penulis. Dan selalu memberikan semangat serta dukungan secara moral dan materi dalam keberlangsungan penyelesaian skripsi ini.
2. Bapak Andreyan Rizky Baskara S.Kom., M.Kom., selaku Dosen Pembimbing Utama sekaligus Koordinator Program Studi Teknologi Informasi yang telah bersedia meluangkan waktunya untuk memberikan arahan, bimbingan dan dukungan kepada saya dengan penuh kesabaran agar skripsi ini dapat cepat terselesaikan.
3. Bapak Ir. Muhammad Alkaff S.Kom., M.Kom dan Bapak Andreyan Rizky Baskara S.Kom., M.Kom., selaku Dosen Pembimbing Akademik penulis, yang telah bersedia membantu dan memberikan bimbingan terbaik selama masa perkuliahan dengan penuh kesabaran.
4. Seluruh Dosen beserta Staf Administrasi Program Studi Teknologi Informasi yang turut mengarahkan dan membantu selama menyelesaikan skripsi.

KATA PENGANTAR

Puji syukur tiada hentinya penulis panjatkan atas kehadiran Allah Subhanahu Wa Ta'ala dan junjungannya Nabi Besar Muhammad Shallallahu 'Alaihi wa Sallam, yang telah memberikan rahmat serta karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Klasifikasi Multi Aspek Sentimen Analisis Pada Review Hotel Menggunakan IndoBERT Embedding dan CNN”. Pada kesempatan ini, penulis ingin mengucapkan terima kasih sebesar-besarnya kepada semua pihak yang telah memberikan bantuan, bimbingan, saran, serta dorongan yang membuat penulis dapat menyelesaikan laporan ini dengan sebaik-baiknya, terutama kepada:

1. Orang tua paling berharga bagi hidup penulis yang selalu memberikan cinta kasih yang luar biasa. Ibu tercinta, wanita terhebat yang selalu memberikan dukungan dalam segala bentuk, memberikan motivasi, do'a, memahami, serta selalu memberikan tempat untuk bersandar dan pulang. Ayah tercinta, yang juga selalu mendukung, memberikan perhatian, selalu berjuang, dan memberikan sebaik-baiknya apa yang penulis butuhkan.
2. Bapak Andreyan Rizky Baskara S.Kom., M.Kom., selaku Dosen Pembimbing Utama sekaligus Koordinator Program Studi Teknologi Informasi yang telah bersedia meluangkan waktunya untuk memberikan arahan, bimbingan dan dukungan kepada saya dengan penuh kesabaran agar skripsi ini dapat cepat terselesaikan.
3. Bapak Ir. Muhammad Alkaff S.Kom., M.Kom dan Bapak Andreyan Rizky Baskara S.Kom., M.Kom., selaku Dosen Pembimbing Akademik penulis, yang telah bersedia membantu dan memberikan bimbingan terbaik selama masa perkuliahan dengan penuh kesabaran Seluruh Dosen dan Staf Program Studi Teknologi informasi yang telah membantu penulis dalam segala hal selama penulis berkuliah di Program Studi ini.
4. Seluruh teman-teman Angkatan 2020 Program Studi Teknologi Informasi serta kakak tingkat, dan teman-teman HMTI periode 2021 - 2022 yang pernah membantu, memberi semangat, serta berbagi canda tawa, suka maupun duka, selama perkuliahan hingga penyelesaian skripsi ini.

Penulis menyampaikan terima kasih kepada semua pihak yang telah ikut andil dalam membantu menyelesaikan penyusunan laporan skripsi ini. Penyusunan lapor skripsi telah disusun dengan optimal, berkat bantuan banyak pihak, tetapi masih jauh dari kata sempurna. Oleh karena itu, penulis siap menerima saran serta kritikan yang membangun dari semua pihak agar laporan skripsi ini dapat memberikan manfaat, terutama kepada para pembaca.

Banjarmasin, November 2024

A handwritten signature in black ink, appearing to read 'M. Majdi', with a stylized, cursive script.

Muhammad Majdi

NIM. 2010817210007

DAFTAR ISI

HALAMAN SAMPUL LUAR	i
HALAMAN SAMPUL DALAM	ii
LEMBAR PERNYATAAN	iii
LEMBAR PENGESAHAN	iv
PERSETUJUAN SKRIPSI	v
ABSTRAK	vi
ABSTRACT	vii
HALAMAN PERSEMBAHAN.....	viii
KATA PENGANTAR	ix
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvii
DAFTAR RIWAYAT HIDUP.....	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Penelitian Terkait.....	6
2.1.1 <i>Deep Learning For Aspect-Based Sentiment Analysis On Indonesian Hotels Reviews</i>	6
2.1.2 <i>Multi-aspect Sentiment Analysis of Tiktok Application Usage Using FasText Feature Expansion and CNN Method</i>	7
2.1.3 <i>Deteksi Aspek Review E-Commerce Menggunakan IndoBERT Embedding dan CNN</i>	7
2.1.4 <i>Aspec-Based Sentiment Analysis on Application Review Using CNN</i> 8	8
2.1.5 <i>Indonesian multilabel classification using IndoBERT embedding and MBERT classification</i>	8

2.1.6	Model Klasifikasi Berbasis <i>Multiclass Classification</i> dengan Kombinasi IndoBERT <i>Embedding</i> dan <i>Long ShortTerm Memory</i> untuk <i>Tweet</i> Berbahasa Indonesia.....	8
2.2	Landasan Teori	11
2.2.1	Analisis Sentimen	11
2.2.2	<i>Natural Language Processing</i>	11
2.2.3	<i>Multi Aspek Sentimen Analisis</i>	11
2.2.4	<i>Text Pre-Processing</i>	12
2.2.5	<i>Confussion Matrix</i>	13
2.2.6	<i>Convolutional Neural Network (CNN)</i>	15
2.2.7	<i>Word Embedding</i>	16
2.3	Kerangka Pemikiran	21
BAB III METODOLOGI PENELITIAN.....		22
3.1	Alat dan Bahan Penelitian.....	22
3.1.1	Alat Penelitian.....	22
3.1.2	Bahan Penelitian	22
3.2	Alur Penelitian	22
3.2.1	Identifikasi Masalah.....	23
3.2.2	Studi Literatur	23
3.2.3	Pengumpulan Data	23
3.2.4	Melakukan Eksperimen	25
3.2.5	Evaluasi Performa.....	29
3.2.6	Implementasi Sistem.....	29
BAB IV HASIL DAN PEMBAHASAN		30
4.1	Pengumpulan Data.....	30
4.2	Pelabelan Dataset.....	31
4.3	<i>Text Preprocessing</i>	34
4.3.1	Konversi Teks ke Huruf Kecil	34
4.3.2	Penghapusan URL dan Email	35
4.3.3	Penanganan Frasa Negasi dan Intensitas Sentimen	35
4.3.4	Penyesuaian Frasa Khusus di Industri Hotel	35
4.3.5	Normalisasi Bahasa Gaul dan Singkatan	35
4.3.6	Standarisasi Tanda Baca	36
4.3.7	Penghapusan Karakter Khusus dan Simbol Tidak Relevan.....	36

4.3.8	Penghapusan Karakter Berulang.....	36
4.3.9	Penghapusan Spasi Berlebihan	36
4.4	Pembagian <i>Dataset</i>	37
4.5	<i>Word Embedding</i>	37
4.6	Membangun Model.....	38
4.7	<i>Train Model</i>	43
4.7.1	Skenario Pertama <i>Training Model CNN</i>	43
4.7.2	Skenario Kedua <i>Training Model CNN</i>	46
4.8	Evaluasi Model	49
4.8.1	Skenario Pertama Evaluasi Model CNN	49
4.8.2	Skenario Kedua Evaluasi Model CNN	56
4.9	Membandingkan Hasil Klasifikasi <i>Single</i> dengan <i>Multi</i> Aspek	63
4.10	Analisis dan Pembahasan.....	68
4.10.1	Hasil Akurasi dari Evaluasi <i>Overall</i>	68
4.10.2	Bar <i>Chart</i> Hasil <i>Confusion Matrix</i> per Aspek	68
4.11	Implementasi Model	73
BAB V KESIMPULAN DAN SARAN.....		76
5.1.1	Kesimpulan	76
5.1.2	Saran	76
DAFTAR PUSTAKA		78
LAMPIRAN.....		81

DAFTAR TABEL

Tabel 2.1 Confusion Matrix	14
Tabel 2.2 Perbandingan antara tiga metode (CNN, MBERT, dan LSTM).....	9
Tabel 2.3 Ringkasan Penelitian Terkait	18
Tabel 3.1 Alat Penelitian.....	22
Tabel 3.2 Contoh Data Ulasan	24
Tabel 3.3 Atribut Aspek Hotel [29]	26
Tabel 4.1 Kriteria Penilaian Pemberian Sentimen	32
Tabel 4.2 Acuan Pemberian Aspek.....	32
Tabel 4.3 Contoh Ulasan Setelah Pelabelan	33
Tabel 4.4 Hasil Konversi Teks ke Huruf Kecil.....	34
Tabel 4.5 Hasil Penghapusan URL dan Email.....	35
Tabel 4.6 Penanganan Frasa Negasi dan Intensitas Sentimen	35
Tabel 4.7 Penyesuaian Frasa Khusus di Industri Hotel.....	35
Tabel 4.8 Hasil Normalisasi Bahasa Gaul dan Singkatan.....	36
Tabel 4.9 Standarisasi Tanda Baca	36
Tabel 4.10 Hasil Penghapusan Karakter Khusus dan Simbol Tidak Relevan	36
Tabel 4.11 Hasil Penghapusan Karakter Berulang	36
Tabel 4.12 Hasil Penghapusan Karakter Berulang	36
Tabel 4.13 Hasil dari Word Embedding Menggunakan IndoBERT	38
Tabel 4.14 Parameter Arsitektur CNN.....	40
Tabel 4.15 Hasil Skenario Pertama Training Model CNN	44
Tabel 4.16 Hasil Skenario Pertama Training Model CNN	46
Tabel 4.17 Hasil Evaluasi Model Skenario Pertama untuk Setiap Aspek	55
Tabel 4.18 Hasil Evaluasi Model Skenario Kedua untuk Setiap Aspek	62
Tabel 4.19 Hasil Evaluasi Menggunakan K-Fold Pada Single Aspek.....	64
Tabel 4.20 Hasil Evaluasi Menggunakan K-Fold Pada Multi Aspek	64
Tabel 4.21 Jumlah Pembagian Sentimen Positif dan Negatif Antara Single dan Multi Aspek.....	65
Tabel 4.22 Perbandingan Hasil Klasifikasi Antara Multi dan Single Aspek.....	65

DAFTAR GAMBAR

Gambar 2.1 Convolutional Neural Network [12]	15
Gambar 2.2 Kerangka Pemikiran	21
Gambar 3.1 Alur Penelitian.....	23
Gambar 3.2 Tahapan Eksperimen	25
Gambar 4.1 Download Web Scrapper Extension	30
Gambar 4.2 Membuat Sitemap	31
Gambar 4.3 Menentukan Elemen Yang Diambil.....	31
Gambar 4.4 Grafik Penyebaran Sentimen Pada Dataset.....	34
Gambar 4.5 Arsitektur Model CNN.....	42
Gambar 4.6 Skenario Pertama Training Model CNN	44
Gambar 4.7 Skenario Kedua Training Model CNN.....	46
Gambar 4.8 Confussion Matrix Harga Skenario Pertama.....	49
Gambar 4.9 Confussion Matrix Hotel Skenario Pertama.....	50
Gambar 4.10 Confussion Matrix Kamar Skenario Pertama.....	51
Gambar 4.11 Confussion Matrix Lokasi Skenario Pertama.....	52
Gambar 4.12 Confussion Matrix Pelayanan Skenario Pertama	53
Gambar 4.13 Confussion Matrix Makanan Skenario Pertama.....	54
Gambar 4.14 Confussion Matrix Overall Skenario Pertama	55
Gambar 4.15 Confussion Matrix Harga Skenario Kedua	56
Gambar 4.16 Confussion Matrix Hotel Skenario Kedua	57
Gambar 4.17 Confussion Matrix Kamar Skenario Kedua	58
Gambar 4.18 Confussion Matrix Lokasi Skenario Kedua	59
Gambar 4.19 Confussion Matrix Pelayanan Skenario Kedua.....	60
Gambar 4.20 Confussion Matrix Makanan Skenario Kedua	61
Gambar 4.21 Confussion Matrix Overall Skenario Kedua	62
Gambar 4.22 Confussion Matrix Pada Sampel Single Aspek	66
Gambar 4.23 Confussion Matrix Pada Sampel Multi Aspek.....	67
Gambar 4.24 Grafik Jumlah Data Hasil Prediksi Yang Salah	69
Gambar 4.25 Tampilan Interface Web.....	73
Gambar 4.26 Tampilan Interface Saat Berhasil Mendeteksi Aspek	74
Gambar 4.27 Tampilan Interface Saat Gagal Mendeteksi Aspek	74

Gambar 4.28 Tampilan Interface Saat Tidak Mendeteksi Aspek 75

DAFTAR LAMPIRAN

Lampiran 1 Kode Untuk Text Preprocessing.....	81
Lampiran 2 Kode Untuk Pembagian Dataset.....	85
Lampiran 3 Kode Untuk Word Embedding dengan IndoBERT	85
Lampiran 4 Kode Untuk Model CNN.....	86
Lampiran 5 Kode Untuk Train Model	88
Lampiran 6 Kode Untuk Evaluasi Model	90
Lampiran 7 Kode Untuk Web dengan Streamlit.....	93
Lampiran 8 Lembar Konsultasi.....	98