

**ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN
HOTEL MENGGUNAKAN METODE BIDIRECTIONAL
ENCODER REPRESENTATIONS
FROM TRANSFORMERS**

SKRIPSI



disusun oleh

Huda Fatkhur Rahman

17.11.1613

PROGRAM SARJANA

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2022

**ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN
HOTEL MENGGUNAKAN METODE BIDIRECTIONAL
ENCODER REPRESENTATIONS
FROM TRANSFORMERS**

SKRIPSI



disusun oleh

Huda Fatkhur Rahman

17.11.1613

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2022

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN HOTEL
MENGUNAKAN METODE BIDIRECTIONAL ENCODER
REPRESENTATIONS FROM TRANSFORMERS**

yang disusun dan diajukan oleh

Huda Fatkhur Rahman

17.11.1613

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 04 Juni 2022

Dosen Pembimbing,

Mardhiya Hayaty, S.T., M.Kom.,

NIK. 190302108

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN HOTEL
MENGGUNAKAN METODE BIDIRECTIONAL ENCODER
REPRESENTATIONS FROM TRANSFORMERS

Huda Fatkhur Rahman

17.11.1613

Telah dipertahankan di depan Dewan Penguji
pada tanggal 27 Agustus 2022
Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Arif Dwi Laksito, M.Kom.

NIK. 190302150

Ika Nur Fajri, M.Kom

NIK. : 190302268

Mardhiya Havaty, S.T., M.Kom.

NIK. 190302108

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 24 September 2022

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta,S.Kom., M.Kom.

NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi ivustakaivan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 10 September 2022



Huda Fatkhur Rahman

NIM. 17.11.1613

MOTTO

“We are responsible for experiences that aren’t our fault all the time.

This is part of life.”

- Mark Manson

“The only true wisdom is in knowing you know nothing.”

- Socrates

"Anyone who has never made a mistake has never tried anything new."

- Albert Einstein

“The desire for more positive experience is itself a negative experience.
And, paradoxically, the acceptance of one’s negative experience is itself a
positive experience.”

- Mark Manson

PERSEMBAHAN

Alhamdulillah dengan kerja keras serta doa, skripsi ini dapat diselesaikan dengan baik. Segala puji dan syukur bagi Allah SWT yang tiada henti memberikan keberkahan. Dengan ini saya mempersembahkan skripsi ini kepada semua pihak yang terlibat secara langsung atau tidak langsung, yaitu untuk :

1. Kedua orang tua, dan kedua kakak saya, yang selalu mendoakan, selalu menyemangati dan memberikan motivasi tiada henti.
2. Dosen pembimbing saya Ibu Mardhiya Hayaty, S.T., M.Kom., yang telah membimbing saya dari awal sampai akhir pembuatan skripsi.
3. Dosen-dosen Universitas AMIKOM Yogyakarta yang telah memberikan banyak ilmu selama kuliah.
4. Teman-teman satu kloter penelitian yang memberikan dukungan dan motivasi, serta menemani saya untuk menyelesaikan skripsi ini.
5. Teman-teman kelas 17-IF-11 yang telah menemani dan selalu memberikan semangat untuk menyelesaikan skripsi ini.
6. Teman-teman Aliansi Kontrakan yang telah menemani dan selalu memberikan semangat untuk menyelesaikan skripsi ini dari awal hingga akhir.
7. Hanif Andri dan Ahfas Reza yang telah Membantu dalam penyusunan dan memberi hiburan selama pengerjaan naskah penelitian.

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat serta hidayah-Nya dan shawalat serta salam juga tidak lupa penulis panjatkan kepada junjungan kita Nabi Muhammad SAW yang telah memberikan teladan mulia dalam menuntun umatnya sehingga penulis dapat menyelesaikan skripsi ini.

Skripsi yang berjudul “**ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN HOTEL MENGGUNAKAN METODE *BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS*** “ ini disusun sebagai salah satu syarat utama untuk menyelesaikan program sarjana pada Universitas AMIKOM Yogyakarta. Penyelesaian skripsi ini juga tidak lepas dari bantuan berbagai pihak, karena itu pada kesempatan ini penulis ingin menyampaikan rasa hormat dan terima kasih kepada :

1. Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Ibu Windha Mega Pradnya D, M.Kom. selaku ketua Program Studi Informatika Universitas AMIKOM Yogyakarta.
4. Ibu Mardhiya Hayaty, S.T., M.Kom. selaku dosen pembimbing yang selalu bijaksana memberikan bimbingan, perhatian, nasehat serta waktunya selama penulisan skripsi ini.
5. Bapak Ika Nur Fajri, M.Kom. dan Bapak Arif Dwi Laksito, M.Kom. selaku dosen penguji. Terimakasih atas saran yang diberikan selama pengujian untuk memperbaiki penelitian menjadi lebih baik lagi.

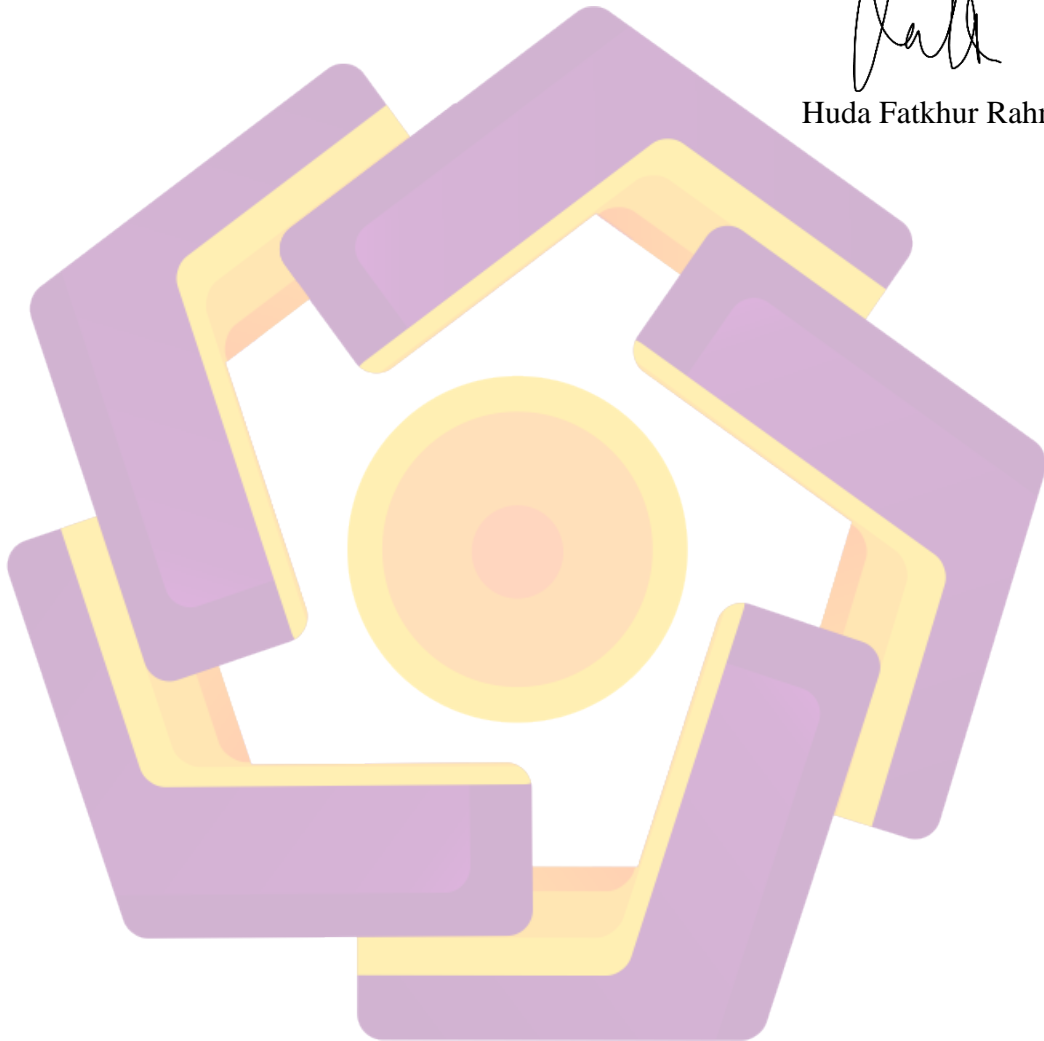
Penulis menyadari skripsi ini masih ada kekurangan. Maka, penulis menerima kritik dan saran yang membangun serta teguran dari semua pihak. Penulis menerima dengan lapang dada untuk kesempurnaan karya selanjutnya. Semoga skripsi yang sederhana ini bisa bermanfaat, khususnya bagi penulis dan pembaca

yang budiman pada umumnya. Apabila terdapat kesalahan semoga Allah SWT melimpahkan magfirah-Nya. Aamiin.

Yogyakarta, 5 September 2022



Huda Fatkhur Rahman



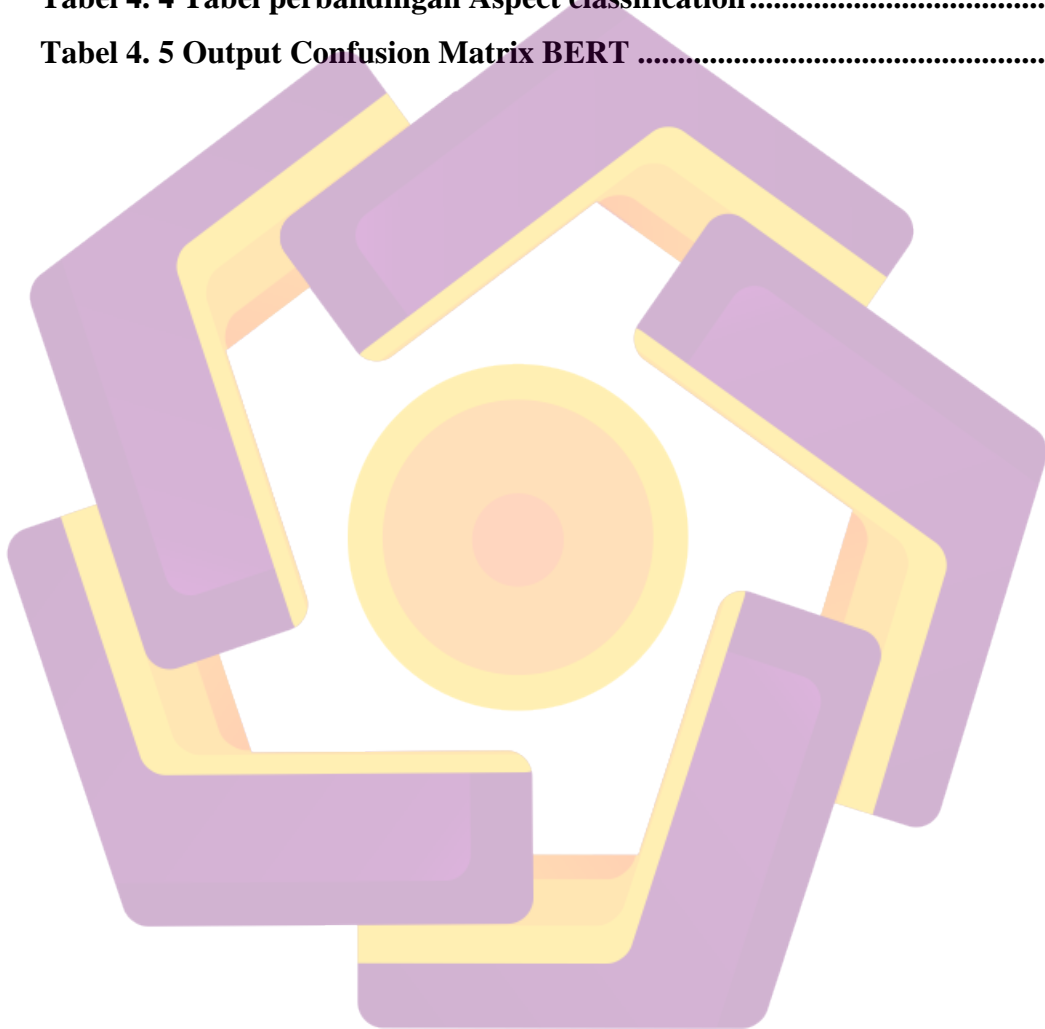
DAFTAR ISI

HALAMAN JUDUL	I
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	IV
MOTTO	V
KATA PENGANTAR	VII
DAFTAR ISI	IX
DAFTAR TABEL	XI
DAFTAR GAMBAR	XII
ABSTRACT	XV
PENDAHULUAN	1
1.1 LATAR BELAKANG	1
1.2 RUMUSAN MASALAH	2
1.3 BATASAN MASALAH	2
1.4 MAKSUD DAN TUJUAN PENELITIAN	3
1.5 MANFAAT PENELITIAN	3
1.6 METODE PENELITIAN	3
1.7 SISTEMATIKA PENULISAN	4
BAB II LANDASAN TEORI	6
2.1 KAJIAN PUSTAKA	6
2.2 DASAR TEORI	8
2.2.1 NATURAL LANGUAGE PROCESSING	8
2.2.2 ANALISIS SENTIMEN	9
2.2.3 ANALISIS SENTIMEN BERBASIS ASPEK.....	10
2.2.4 WORD EMBEDDING	11
2.2.5 VECTOR SPACE MODEL.....	12
2.2.6 BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS (BERT).....	13
2.2.7 LINEAR REGRESSION	18
2.2.8 PREPROCESSING DATA	19
2.2.9 CONFUSION MATRIX	21
BAB III METODE PENELITIAN	23
3.1 ALAT DAN BAHAN PENELITIAN	23

3.1.1	Perangkat Keras (Hardware).....	23
3.1.2	Perangkat Lunak (Software).....	23
3.2	ALUR PENELITIAN	24
3.3	PENGUMPULAN DATA.....	24
3.4	MANUAL LABELING	24
3.5	PREPROCESSING DATA.....	25
3.6	DATA TRAINING, DATA VALID DAN DATA TESTING	25
3.7	EKSTRAKSI FITUR	25
3.8	IMPLEMENTASI ALGORITMA KLASIFIKASI	25
3.9	EVALUASI.....	26
BAB IV	HASIL DAN PEMBAHASAN	27
4.1	PENGUMPULAN DATA.....	27
4.2	DATASET	27
4.3	PERSIAPAN	29
4.4	PERSIAPAN DATA PRE-TRAIN.....	30
4.5	PREPROCESSING DATA.....	30
4.1.1	Case Folding.....	30
4.1.2	Tokenizing.....	31
4.6	PERMODELAN	34
4.7	MELAKUKAN TESTING MODEL.....	37
4.8	PENGUKURAN AKURASI TESTING	38
BAB V	PENUTUP	45
5.1	KESIMPULAN	45
5.2	SARAN	45
	DAFTAR PUSTAKA	46

DAFTAR TABEL

Tabel 2. 1 Tabel Perbandingan	7
Tabel 4. 1 Contoh data untuk klasifikasi sentimen	28
Tabel 4. 2 Hasil Case Folding	31
Tabel 4. 3 Tabel perbandingan Aspect classification	42
Tabel 4. 4 Tabel perbandingan Aspect classification	42
Tabel 4. 5 Output Confusion Matrix BERT	43



DAFTAR GAMBAR

Gambar 2. 1 Representasi Model Ruang Vektor.....	13
Gambar 2. 2 Visualisasi sederhana kinerja encoder dan decoder	14
Gambar 2. 3 Ilustrasi proses fine-tuning pada BERT	16
Gambar 2. 4 Representasi input pada BERT	16
Gambar 2. 5 Visualisasi proses fine-tuning pada BERT	18
Gambar 3. 1 Diagram Alur Tahapan Penelitian	24
Gambar 4. 1 Scraping data menggunakan WebHarvy.....	27
Gambar 4. 2 Dataset.....	28
Gambar 4. 3 Perintah memasang library transformers	29
Gambar 4. 4 Perintah cloning repository IndoNLU	29
Gambar 4. 5 Kode untuk mengimpor library/ package dan function.....	29
Gambar 4. 6 Kode untuk mempersiapkan data <i>pre-train</i>	30
Gambar 4. 7 Kode Case Folding	30
Gambar 4. 8 Kode untuk memeriksa kinerja Tokenizer.....	32
Gambar 4. 9 Hasil keluaran proses Tokenizing	32
Gambar 4. 10 Deklarasi masing-masing dataset	33
Gambar 4. 11 Deklarasi tokenizing dataset	33
Gambar 4. 12 Konfigurasi tambahan pada dataset	34
Gambar 4. 13 Optimizer Adam serta pengiriman model ke GPU.....	34
Gambar 4. 14 Output ditampilkan dari hasil training	35
Gambar 4. 15 Visualisasi proses training dan evaluasi BERT.....	36
Gambar 4. 16 Visualisasi proses training dan evaluasi BERT ratio data 80% train, 10% test , 10% valid	36
Gambar 4. 17 Fungsi <code>forward_sequence_multi_classification</code>	37
Gambar 4. 18 Fungsi logits	38
Gambar 4. 19 Perhitungan f1 Score , Confusion Matrix dan Classification Report	38
Gambar 4. 20 Output perhitungan f1 Score , Confusion Matrix dan Classification Report	39

Gambar 4. 21 Output perhitungan f1 Score , Confusion Matrix dan Classification Report dengan stopwords..... 40

Gambar 4. 22 Kode perbandingan label file asli dengan labeling BERT 41

Gambar 4. 23 Grafik perbandingan labeling..... 41



INTISARI

Bidang usaha perhotelan merupakan usaha bisnis yang bergerak pada bidang pariwisata yang penting baik dari segi fasilitas, pelayanan ataupun jarak tempuh perjalanan wisata. Saat ini sudah banyak website wisata yang menyediakan fasilitas untuk pengguna internet menuliskan opini dan pengalaman pribadinya secara online. Analisis sentimen berbasis aspek atau biasa disingkat ABSA adalah mengidentifikasi aspek dari entitas yang diberikan dan mengidentifikasi sentiment yang diekspresikan pada tiap aspek. Penelitian ini menggunakan data ulasan hotel untuk melakukan analisis sentimen yang didapatkan dari situs Traveloka.

Penelitian ini memanfaatkan sebuah metode *deep learning* yaitu *Bidirectional Encoder Representation from Transformer* atau BERT sebagai metode *word embedding* untuk mempresentasikan kata menjadi vektor. Klasifikasi pada penelitian ini dilakukan dengan menambahkan layer *linear regression* pada layer paling atas di BERT.

Dari percobaan yang dilakukan dengan pembagian data training sebanyak 60%, data eval sebanyak 20%, dan data tes sebanyak 20% dari total 5.000 data, dapat dilihat bahwa metode BERT untuk klasifikasi sentimen memberikan hasil dengan nilai precision sebesar 0.89, recall sebesar 0.90, dan f1-score sebesar 0.90. Sedangkan untuk klasifikasi sentimen memberikan hasil terbaik pada model BERT dengan nilai precision sebesar 0.88, recall sebesar 0.90, dan f1-score sebesar 0.88. Preprocessing yang dilakukan dengan data tanpa stemming memiliki hasil yang hampir mirip bahkan lebih baik dibandingkan menggunakan data dengan stemming.

Kata Kunci: Analisis Sentimen berbasis aspek, *Word Embedding*, *Deep Learning*, BERT

ABSTRACT

The hospitality business is a business that is engaged in the tourism sector which is important in terms of facilities, services, or travel distances. Currently, there are many travel websites that provide facilities for internet users to write their opinions and personal experiences online. Aspect-based sentiment analysis or commonly abbreviated as ABSA is to identify aspects of a given entity and identify the sentiment expressed in each aspect. This study uses hotel review data to analyse sentiments obtained from the Traveloka site.

This research utilizes a deep learning method, namely Bidirectional Encoder Representation from Transformer or BERT as a word embedding method to represent words into vectors. Classification in this study is done by adding a linear regression layer to the top layer in BERT.

From the experiments conducted with the distribution of training data as much as 60%, eval data as much as 20%, and test data as much as 20% of the total 5,000 data, it can be seen that the BERT method for sentiment classification gives results with precision values of 0.89, recall of 0.90, and f1-score of 0.90. Meanwhile, the sentiment classification gives the best results on the BERT model with a precision value of 0.88, a recall of 0.90, and an f1-score of 0.88. Pre-processing carried out with data without stemming has almost the same results and even better than using data with stemming.

Keyword: Aspect Based Sentiment Analysis, Word Embedding, Deep Learning, BERT