

**MULTI-ASPECT SENTIMENT ANALYSIS PADA REVIEW FILM  
MENGUNAKAN METODE BIDIRECTIONAL ENCODER  
REPRESENTATIONS FROM TRANSFORMER (BERT)**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**NUR KARIMAH  
21.21.1524**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2023**

**MULTI-ASPECT SENTIMENT ANALYSIS PADA REVIEW FILM  
MENGUNAKAN METODE BIDIRECTIONAL ENCODER  
REPRESENTATION FROM TRANSFORMERS (BERT)**

**HALAMAN JUDUL**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**NUR KARIMAH  
21.21.1524**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2023**

HALAMAN PERSETUJUAN

SKRIPSI

**MULTI-ASPECT SENTIMENT ANALYSIS PADA REVIEW FILM  
MENGUNAKAN METODE BIDIRECTIONAL ENCODER  
REPRESENTATION FROM TRANSFORMERS (BERT)**

yang disusun dan diajukan oleh

**NUR KARIMAH**

**21.21.1524**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 27 Juli 2023

Dosen Pembimbing,



**Anna Baita, M.Kom**

**NIK. 190302290**

HALAMAN PENGESAHAN

SKRIPSI

**MULTI-ASPECT SENTIMENT ANALYSIS PADA REVIEW FILM  
MENGUNAKAN METODE BIDIRECTIONAL ENCODER  
REPRESENTATION FROM TRANSFORMERS (BERT)**

yang disusun dan diajukan oleh

**NUR KARIMAH**  
21.21.1524

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 27 Juli 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Anggit Dwi Hartanto, M.Kom  
NIK. 190302163

Arif Akbarul Huda, S.Si, M.Eng  
NIK. 190302287

Anna Baita, M.Kom  
NIK. 190302290



Skrripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 27 Juli 2023

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.  
NIK. 190302096

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Nur Karimah  
NIM : 21.21.1524

Menyatakan bahwa Skripsi dengan judul berikut:

**Multi-Aspect Sentiment Analysis Pada Review Film Menggunakan Metode Bidirectional Encoder Representation From Transformers (BERT)**

Dosen Pembimbing: Anna Baita, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 27 Juli 2023

Yang Menyatakan,



Nur Karimah

## HALAMAN PERSEMBAHAN

Dengan mengucapkan syukur kepada Tuhan Yang Maha Esa atas berkat, rahmat serta karunia-Nya, Penulis dapat menyelesaikan Skripsi yang berjudul **MULTI-ASPECT SENTIMENT ANALYSIS PADA REVIEW FILM MENGGUNAKAN METODE BIDIRECTIONAL ENCODER REPRESENTATION FROM TRANSFORMERS (BERT)**.

Penulis selalu mendapatkan bimbingan, dorongan, serta semangat dari banyak pihak. Oleh karena itu Penulis ingin mempersembahkan Skripsi ini kepada:

1. Bapak Ruslan Kasirun dan Ibu Suyatin yang sangat saya sayangi, terima kasih banyak karena telah memberikan begitu banyak dorongan, dukungan, dan doa yang begitu besar. Doa dan dukunganmu selalu menyertai langkahku.
2. Kepada seseorang yang sangat memberikan banyak *support* kepada saya, Aditya Khoirul Anam. Terima kasih telah memberikan segala perhatian dan waktu untuk membantu dan membimbing saya hingga akhirnya saya dapat menyelesaikan skripsi ini dengan nilai yang sangat memuaskan.
3. Kepada teman-teman seperjuangan saya selama menempuh jenjang S1, Nabilla, Lia, dan Zahra. Terimakasih telah kebersamai dan berjuang hingga kita dapat menyelesaikan jenjang pendidikan stars sarjana ini walaupun sangat berat tantangannya.

## KATA PENGANTAR

Dengan mengucapkan syukur kepada Tuhan Yang Maha Esa atas berkat, rahmat serta karunia-Nya, Penulis dapat menyelesaikan Skripsi yang berjudul **MULTI-ASPECT SENTIMENT ANALYSIS PADA REVIEW FILM MENGGUNAKAN METODE BIDIRECTIONAL ENCODER REPRESENTATION FROM TRANSFORMERS (BERT)**.

Skripsi ini ditujukan untuk memenuhi salah satu persyaratan ujian guna memperoleh gelar derajat sarjana komputer (S.Kom) pada Jurusan Informatika Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta. Penulis menyadari dalam penulisan Skripsi ini masih jauh dari sempurna, dan banyak kekurangan baik dalam metode penulisan maupun dalam pembahasan materi. Hal tersebut dikarenakan keterbatasan kemampuan Penulis. Sehingga Penulis mengharapkan saran dan kritik yang bersifat membangun sehingga dapat memperbaiki segala kekurangannya di kemudian hari. Dalam penulisan skripsi ini, Penulis selalu mendapatkan bimbingan, dorongan, serta semangat dari banyak pihak. Oleh karena itu Penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Pembimbing yang terhormat, yakni Yth. Ibu Anna Baita, M.Kom selaku Dosen Pembimbing, yang telah meluangkan waktunya, tenaga dan pikirannya untuk membimbing Penulis dalam penulisan skripsi ini.
2. Tim Dosen Penguji, yakni Yth Bapak Anggit Dwi Hartanto, M.Kom selaku penguji satu dan Bapak Arif Akbarul Huda, S.Si, M.Eng selaku penguji dua, yang telah memberikan masukan terhadap penulisan skripsi ini.
3. Seluruh jajaran Staff Dosen dan Karyawan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Kedua orang tua yang telah memberikan dorongan dan doa sehingga Penulis dapat menyelesaikan skripsi ini.

Akhirnya, Penulis mengucapkan rasa terima kasih kepada semua pihak dan apabila ada yang tidak disebutkan Penulis mohon maaf, dengan besar harapan semoga skripsi yang ditulis oleh Penulis ini dapat bermanfaat khususnya bagi Penulis sendiri dan umumnya bagi pembaca. Bagi para pihak yang telah membantu dalam penulisan skripsi ini semoga segala amal dan kebaikannya mendapatkan balasan yang berlimpah dari Tuhan YME, Amiiin.

Yogyakarta, 24 September 2023

Penulis



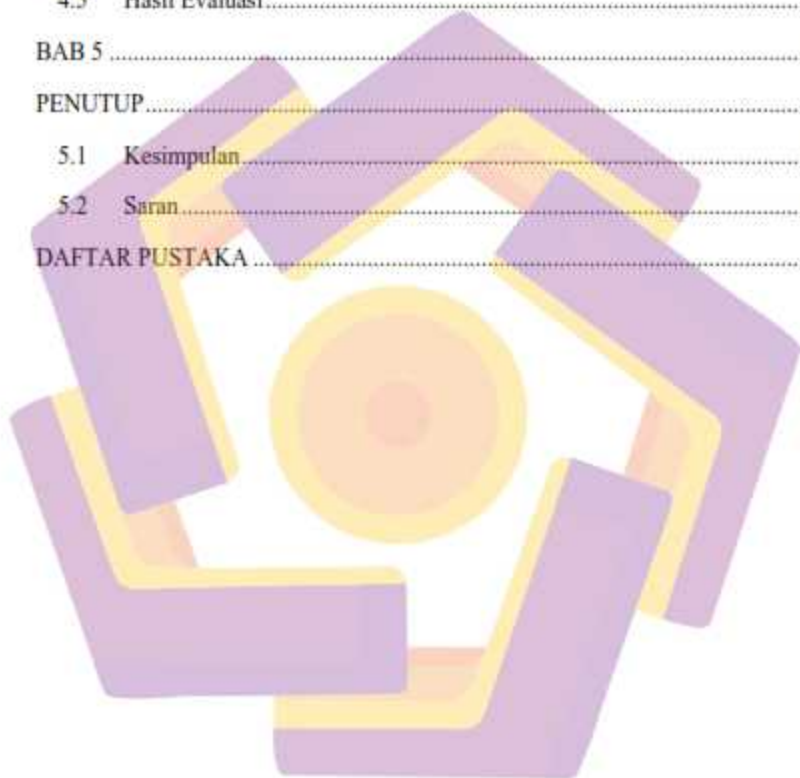


## DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	v
HALAMAN PERSEMBAHAN.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
INTISARI.....	xv
ABSTRACT.....	xvi
BAB 1.....	1
PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	2
1.4 Tujuan.....	3
1.5 Manfaat.....	3
1.6 Sistematika Penulisan.....	4
BAB 2.....	5
TINJAUAN PUSTAKA.....	5
2.1 Studi Literatur.....	5
2.2 Dasar Teori.....	9
2.2.1 <i>Review Film</i> .....	9

2.2.2	<i>Web Scraping</i> .....	9
2.2.3	<i>Natural Language Processing (NLP)</i> .....	10
2.2.4	Analisis Sentimen .....	10
2.2.5	<i>Multi-Aspect Sentiment Analysis</i> .....	11
2.2.6	<i>Attention</i> .....	11
2.2.7	<i>Transformer</i> .....	13
2.2.8	<i>BERT</i> .....	15
BAB 3 .....		19
METODE PENELITIAN .....		19
3.1	Objek Penelitian .....	19
3.2	Alur Penelitian .....	19
3.3	Data Penelitian .....	22
3.3.1	Pengumpulan Data .....	22
3.3.2	Pelabelan Data .....	22
3.3.3	<i>Cleaning Data</i> .....	23
3.3.4	<i>Implementasi BERT</i> .....	27
3.3.5	Evaluasi .....	38
3.4	Alat dan Bahan .....	40
3.4.1	Spesifikasi Perangkat Keras .....	40
3.4.2	Spesifikasi Perangkat Lunak .....	40
3.4.3	Bahan .....	40
BAB 4 .....		41
HASIL DAN PEMBAHASAN .....		41
4.1	Hasil Pengumpulan Data .....	41
4.2	Pelabelan Data .....	42

4.3	Implementasi BERT .....	45
4.3.1	<i>BERT Tokenizing</i> .....	45
4.3.2	Model .....	51
4.4	Hasil Klasifikasi Sentimen .....	53
4.5	Hasil Evaluasi .....	55
BAB 5	.....	57
PENUTUP	.....	57
5.1	Kesimpulan .....	57
5.2	Saran .....	58
DAFTAR PUSTAKA	.....	59



## DAFTAR TABEL

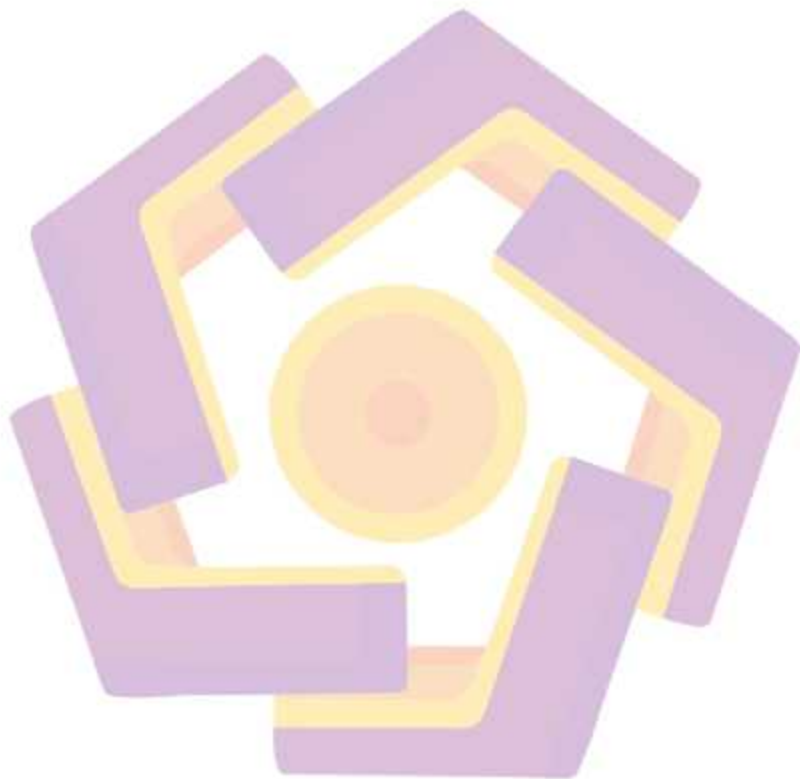
Tabel 2.1 Perbandingan Penelitian.....	7
Tabel 3.1 Labeling Data.....	23
Tabel 3.2 Menghapus Tag HTML.....	25
Tabel 3.3 Case Folding.....	26
Tabel 3.4 Menghapus Tanda baca, Angka, dan Karakter.....	27
Tabel 3.5 Input pada BERT.....	37
Tabel 3.6 Confusion Matrix.....	39
Tabel 4.1 Kumpulan Dataset.....	41
Tabel 4.2 Nilai <i>Train Loss</i> , <i>Val Loss</i> , dan <i>Test Loss</i> .....	54
Tabel 4.3 Nilai <i>Output</i> Probabilitas pada Setiap Percobaan.....	55
Tabel 4.4 Nilai <i>Accuracy</i> pada Setiap Percobaan.....	55
Tabel 4.5 Nilai <i>Confusion Matrix</i> pada Masing-Masing Percobaan.....	56



## DAFTAR GAMBAR

Gambar 2.1 Web Scraping .....	10
Gambar 2.2 Attention pada seluruh kalimat .....	12
Gambar 2.3 Input Kalimat .....	12
Gambar 2.4 Output Representasi Kalimat .....	13
Gambar 2.5 Arsitektur Transformer [29] .....	14
Gambar 2.6 Procedure pre-training dan fine-tuning BERT[30] .....	16
Gambar 2.7 Ilustrasi dari BERT Embedding[30] .....	18
Gambar 3.1 Alur Penelitian.....	20
Gambar 3.2 Alur Pre-Processing Data .....	24
Gambar 3.3 Proses Tokenisasi dengan WordPiece .....	29
Gambar 3.4 Proses Token Embedding .....	29
Gambar 3.5 Proses Pemberian Token Padding .....	30
Gambar 3.6 Tahap Substitusi Input IDs .....	31
Gambar 3.7 Tahap Sentence Embedding .....	31
Gambar 3.8 Tahap Positional Embedding .....	31
Gambar 3.9 Gambaran Keseluruhan Proses Tokenisasi .....	32
Gambar 3.10 Representasi Input pada BERT[30].....	32
Gambar 3.11 Input dan Output dalam BERT.....	33
Gambar 3.12 Ilustrasi Layer untuk Sentimen Analisis .....	34
Gambar 3.13 Ilustrasi Proses Klasifikasi menggunakan BERT .....	36
Gambar 4.1 Hasil Scraping Data.....	41
Gambar 4.2 <i>Wordcloud</i> Ulasan Film .....	42
Gambar 4.3 Labeling Dataset.....	43
Gambar 4.4 Visualisasi <i>Dataset</i> Berdasarkan Jumlah Aspek dan Sentimen .....	43
Gambar 4.5 Visualisasi <i>Dataset</i> Berdasarkan Jumlah Sentimen .....	44
Gambar 4.6 Pemisahan Label Berdasarkan Aspek .....	45
Gambar 4.7 Kode Program Model BERT Tokenizer.....	46
Gambar 4.8 Output BERT Tokenizer .....	47
Gambar 4.9 Grafik Panjang Token .....	47
Gambar 4.10 Inisiasi Panjang Token BERT .....	47

Gambar 4.11 Konversi Token Menjadi Tensor.....	48
Gambar 4.12 Kode Program Dataset Menggunakan TorchDataset .....	49
Gambar 4.13 Kode Program Dataset Menggunakan LightningDataModule .....	50
Gambar 4.14 Kode Program Implementasi Model BERT .....	53



## INTISARI

*Review* film merupakan salah satu hal yang mempengaruhi minat seseorang dalam melihat sebuah film. Beberapa platform yang memberikan informasi terkait *review* film adalah *IMDB*, *Rotten tomatoes*, dan *metacritic*. Penelitian ini dilakukan untuk mengaplikasikan metode *Bidirectional Encoder Representation from Transformer (BERT)* pada *multi-aspect sentiment analysis* terhadap ulasan film. Data *review* diambil dengan menggunakan metode *scrapping*. Data yang digunakan berjumlah 1899 dengan jumlah data yang memiliki sentimen positif sebanyak 3245, sentimen netral sebanyak 4825, dan sentimen negatif sebanyak 1424. Pendekatan yang diusulkan mencakup aspek-aspek seperti *acting*, *plot*, *cast*, *animation*, dan *music*. Aspek yang memiliki sentimen positif terbanyak terdapat pada aspek *music* dengan total 631 data, sentimen netral berada pada aspek *animation* dengan total 1146, dan sentimen negatif berada pada aspek *plot* dengan total 362. *Dataset* yang digunakan melalui tahap *cleaning* yang terdiri dari *case folding* dan penghapusan *Tag HTML*, tanda baca, angka dan karakter. Penelitian ini menggunakan model *BERT<sub>BASE-UNCASE</sub>* dengan empat kali percobaan menggunakan *hyperparameter* *max\_epoch* 10, *batch size* 16, dan *learning rate*  $1e^{-4}$ ,  $5e^{-5}$ ,  $3e^{-5}$ , dan  $2e^{-5}$ . Hasil penelitian menunjukkan bahwa, dari semua percobaan didapatkan bahwa nilai *accuracy* terbaik berada pada percobaan ketiga dengan menggunakan *learning rate*  $3e^{-5}$  sebesar 82,32%. Sedangkan nilai *precision*, *recall*, dan *f1-score* terbaik terletak pada aspek *animation* sebesar 86%, 85%, dan 85%.

**Kata kunci:** *Review film; Multi-aspect; Analisis sentimen; Bidirectional Encoder Representations from Transformer (BERT); BERT Base-Uncase.*

## ABSTRACT

*A film review is one of the things that influences someone's interest in watching a movie. Several platforms that provide information related to film reviews are IMDB, Rotten Tomatoes, and Metacritic. This research was conducted to apply the Bidirectional Encoder Representation from Transformer (BERT) method to multi-aspect sentiment analysis of film reviews. The review data was obtained using the scraping method. The dataset used consists of 1899 data to 3245 data having a positive sentiment, 4825 data with a neutral sentiment, and 1424 data with a negative sentiment. The proposed approach includes the aspects such as acting, plot, cast, animation, and music. The aspect with the most positive sentiment is music with a total of 631 data, the neutral sentiment is found in the animation aspect with a total of 1146, and the negative sentiment is found in the plot aspect with a total of 362. The dataset used went through cleaning data, including case folding and removing HTML tags, punctuation, numbers, and special characters. This research uses the BERTBASE-UNCASE model with four experiments using hyperparameters max\_epoch 10, batch size 16, and learning rates of  $1e-4$ ,  $5e-5$ ,  $3e-5$ , and  $2e-5$ . The research results show that, from all experiments, the best accuracy value is achieved in the third experiment using a learning rate of  $3e-5$ , which is 82,32%. Meanwhile, the best precision, recall, and f1-score values for the "animation" aspect are 86%, 85%, and 85%.*

**Keyword:** Review film; Multi-aspect; Sentiment analysis; Bidirectional Encoder Representations from Transformer (BERT); BERT Base-Uncase.