

ANALISIS SENTIMEN DOSEN PEMBIMBING DALAM MEREVISI SKRIPSI MAHASISWA PRODI INFORMATIKA MENGGUNAKAN METODE BERT

Andini Banuri¹, Muhammad Fhadli², Assaf Arief³, Muhammad Ridha Albaar⁴
Program Studi Teknik Informatika, Fakultas Teknik, Universitas Khairun
Jl. Jati Metro, Kota Ternate Selatan

Email: andinibanuri22@gmail.com¹, muhammadfhadli20@gmail.com², assaf.rief @unkhair.ac.id³
mridha.albaar84@gmail.com⁴

Abstrak

Sistem penilaian pembelajaran di perguruan tinggi untuk mengetahui indikator kinerja dosen. Namun dalam kenyataannya, proses bimbingan skripsi mahasiswa seringkali mengalami hambatan dengan dosen pembimbing. Analisis sentimen dilakukan pada komentar mahasiswa terhadap dosen pembimbing untuk mengklasifikasikan ke dalam tiga kategori sentimen yaitu positif, negatif, dan netral. Metode analisis menggunakan BERT memanfaatkan arsitektur transformer dengan pendekatan bidirectional, yaitu melihat konteks secara dua arah sehingga mampu memahami konteks yang lebih luas. Pelabelan data menggunakan GPT-3 dan manual oleh pakar bahasa. Data sebanyak 700 kalimat telah dilabeli dengan 330 kalimat positif, 351 kalimat netral, dan 19 kalimat negatif. Setelah pemrosesan data melalui case folding, filtering, dan tokenization, model BERT diaplikasikan menggunakan IndoBERT-Base. Pelatihan model dengan fine-tuning menggunakan hyperparameter batch size 32, learning rate 2e-5, epoch 10. Hasil dalam menggunakan BERT didapatkan nilai tertinggi untuk akurasi sebesar 84%, precision netral sebesar 87%, precision positif sebesar 81% dan precision negatif adalah 0. Untuk recall netral sebesar 87%, recall positif sebesar 93%, dan recall negatif adalah 0. Untuk f1-score netral sebesar 87%, f1-score positif sebesar 87% dan f1-score negatif adalah 0. Hasil ini merekomendasikan penggunaan BERT untuk analisis sentimen dengan data yang seimbang.

Kata kunci: Analisis Sentimen, BERT, Bimbingan Skripsi, Komentar Mahasiswa

ANALYSIS OF THE SENTIMENT OF SUPERVISORS IN REVISING THE THESES OF INFORMATICS STUDY PROGRAM STUDENTS USING THE BERT METHOD

Abstract

The learning assessment system in higher education is used to determine lecturer performance indicators. However, in reality, the process of supervising students' theses often encounters obstacles with their supervisors. Sentiment analysis was conducted on student comments to their supervisors to classify them into three sentiment categories: positive, negative, and neutral. The analysis method using BERT utilizes a transformer architecture with a bidirectional approach, namely viewing the context in two directions to understand the broader context. Data labeling used GPT-3 and manual processing by language experts. The data of 700 sentences were labeled with 330 positive sentences, 351 neutral sentences, and 19 negative sentences. After data processing through case folding, filtering, and tokenization, the BERT model was applied using IndoBERT-Base. Model training with fine-tuning using hyperparameter batch size 32, learning rate 2e-5, epoch 10. The results in using BERT obtained the highest value for accuracy of 84%, neutral precision of 87%, positive precision of 81% and negative precision is 0. For neutral recall of 87%, positive recall of 93%, and negative recall is 0. For neutral f1-score of 87%, positive f1-score of 87% and negative f1-score of 0. These results recommend the use of BERT for sentiment analysis with balanced data.

Keywords: Satisfaction, Website, Misbahula Aulad Labuha Vocational School, Webqual 4.0.

1. PENDAHULUAN

Adanya sistem penilaian pembelajaran perguruan tinggi dan universitas dapat membantu perguruan tinggi dapat melihat indikator kinerja

dosen [1]. Kegiatan bimbingan skripsi merupakan sebuah tugas dan kewajiban seorang dosen dalam menjalankan tri dharma perguruan tinggi sebagaimana yang dijelaskan dalam undang-undang No 14 Tahun 2005 Pasal 60 tentang tugas dan

kewajiban dosen meliputi, pelaksanaan pendidikan, penelitian dan pengabdian masyarakat (UU No 14 2005). Dalam kasus tertentu, proses bimbingan skripsi mahasiswa seringkali mengalami hambatan dengan dosen pembimbingnya [2].

Analisis sentimen diperlukan karena dapat memberikan cara pandang tentang persepsi, dan emosi orang tentang berbagai hal [3]. Menganalisis sentimen dalam ulasan, dapat diketahui aspek mana yang memuaskan dan mana yang perlu ditingkatkan. Informasi ini dapat digunakan untuk perbaikan, inovasi atau pengembangan yang sesuai [4].

Analisis sentimen dosen pembimbing dilakukan karena dosen pembimbing berperan sangat penting dalam mendampingi dan membimbing mahasiswa selama proses penelitian atau proses penulisan skripsi. Dosen pembimbing lebih awal dalam membimbing mahasiswa melalui proses penelitian dan penulisan skripsi. Selama proses penulisan skripsi, mahasiswa terlibat dalam interaksi yang lebih dalam dan berkelanjutan dengan dosen pembimbing [5].

Analisis sentimen dosen pembimbing dapat digunakan untuk meningkatkan kualitas bimbingan akademik secara keseluruhan di program studi. Menggunakan persepsi mahasiswa terhadap dosen pembimbing, program studi dapat meningkatkan dan mengembangkan metode dalam bimbingan [6]. Analisis sentimen yang dilakukan pada program studi dapat menyesuaikan dan dapat mengambil tindakan yang diperlukan untuk lebih memenuhi kebutuhan dan harapan mahasiswa [7].

Berdasarkan permasalahan tersebut maka penelitian ini melakukan analisis sentimen pada data responden mahasiswa terhadap penilaian kinerja dosen pembimbing menggunakan metode Bidirectional Encoder Representations from Transformers (BERT). Pemilihan BERT sebagai metode analisis sentimen dalam penelitian ini didasarkan pada kemampuan BERT dalam pendekatan bidirectional (dua arah), BERT dapat memperhitungkan konteks yang lebih luas dan mengenali hubungan antar kata dalam teks [8]. Kemampuan metode BERT diharapkan mampu menjawab pertanyaan pada masalah penelitian tersebut.

2. TINJAUAN PUSTAKA

2.1 Analisis Sentimen

Analisis sentimen dalam arti luas mengacu pada bidang-bidang seperti ilmu komputasional, pemrosesan bahasa alami dan penambangan teks. Tujuannya adalah untuk menentukan sikap pembicara atau penulis terhadap topik tertentu [9].

Analisis sentimen adalah teknik yang digunakan pemrosesan bahasa alami untuk menentukan dan memperoleh informasi dari data. Dalam beberapa penelitian sebelumnya yaitu menentukan apakah

komentar tersebut menunjukkan sentimen positif atau negatif maupun netral [10].

2.2 Transformer

Transformer adalah model pembelajaran mendalam yang diperkenalkan pada tahun 2017 di jurnal Attention is all you need. Transformer digunakan di bidang pemrosesan NLP. Transformer dirancang untuk memproses data berurutan, seperti pemrosesan bahasa alami untuk tugas-tugas seperti penerjemahan dan peringkasan [11].

Transformer adalah arsitektur encoder-decoder. Encoder terdiri dari satu set lapisan pengkodean yang secara iteratif memproses input satu demi satu, dan decoder terdiri dari satu set lapisan decoding yang melakukan hal yang sama untuk output dari encoder. Fungsi dari setiap lapisan encoder adalah memproses inputnya untuk menghasilkan pengkodean, berisi informasi tentang bagian mana dari input yang terkait satu sama lain, dan kemudian meneruskan rangkaian pengkodean sebagai input ke lapisan pembuat encoder berikutnya. Setiap lapisan decoder mengambil semua pengkodean dan memprosesnya, menggunakan informasi kontekstual gabungan untuk menghasilkan urutan keluaran.

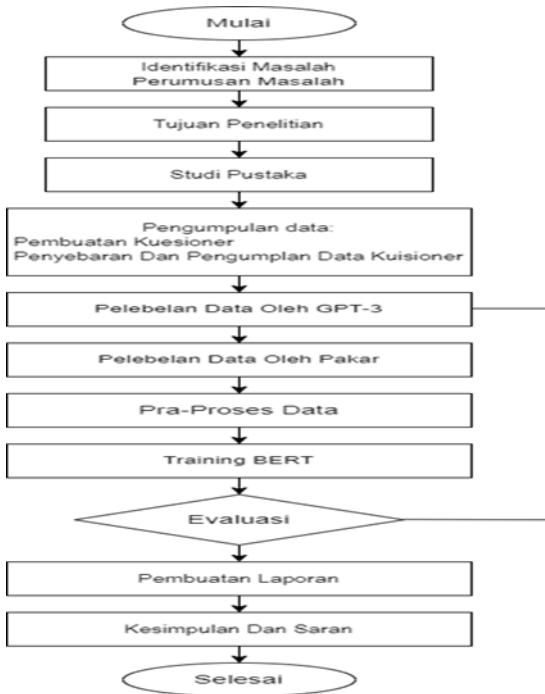
2.3 Bidirectional Encoder Representations from Transformers (BERT)

Bidirectional Encoder Representations from Transformers (BERT) adalah teknik pembelajaran mesin berbasis transformer untuk pemrosesan bahasa alami (NLP) pra-pelatihan, dikembangkan oleh Google. BERT dibuat dan dirilis pada tahun 2018 oleh Jacob Devlin dan rekan-rekannya di Google, dan pada tahun 2019, Google memanfaatkan BERT untuk lebih memahami penelusuran pengguna [12].

Salah satu model representasi bahasa terbaru adalah BERT, menggunakan model bahasa masked untuk secara acak menutupi berbagai token dari input dan menggabungkan konteks kiri dan kanan, sehingga menerapkan transformator dua arah [13].

3. METODE PENELITIAN

Penelitian ini menganalisis komentar dari mahasiswa selama proses bimbingan skripsi menggunakan algoritma (BERT) dengan tahapan penelitian yang dapat dilihat pada gambar 1.



3.1. Pengumpulan Data

Data yang akan digunakan pada penelitian ini yaitu data yang diambil menggunakan *google form* yang telah diisi oleh mahasiswa prodi informatika unkhair. Pada penelitian ini akan diambil sebanyak

100 responden dan menggunakan tiga kategori sentimen analisis yaitu positif, negatif dan netral.

3.2 Pra-Proses Data.

Tahap ini dilakukan untuk memproses data mentah agar digunakan sebagai data training penelitian. Tahapan preprocessing data yaitu sebagai berikut: (1) Case Folding digunakan untuk mengubah seluruh data opini menjadi huruf huruf kecil. (2) Data yang tidak diperlukan seperti tanda baca dan kata-kata non-alfabet dihilangkan. (3) Tokenization adalah proses pemisahan rangkaian string menjadi fragmen-fragmen string yang selanjutnya disebut token. Pada penelitian ini, setiap kalimat atau frase akan dipecah menjadi kata sesuai dengan kosakata BERT..

3.3 Pelabelan Data

Pelabelan data menggunakan GPT-3 dapat Pada melihat peluang paling besar hingga paling kecil sehingga dapat diurutkan hasil sentimennya. Data yang diambil akan digunakan sebesar 75% dalam pelabelan menggunakan GPT-3 yang diambil

berdasarkan probabilitas yang paling tinggi dan sisanya 25% pada probabilitas yang rendah maka akan dilakukan pelabelan ulang oleh si pakar bahasa sebagai perbandingan untuk menentukan hasil sentimennya.

3.4 Implementasi BERT

Pada bagian ini akan dilakukan pelatihan bert menggunakan IndoBERT yang telah dilakukan penelitian sebelumnya oleh koto tahun 2021. IndoBERT adalah model BERT untuk bahasa Indonesia, IndoBERT memiliki 3 model IndoBERT-liteBase, IndoBERTBase, IndoBERTLarge.

BERT memanfaatkan transformer, mekanisme attention yang mempelajari hubungan kontekstual antara kata (atau subkata) dalam teks. Secara formal, transformator terdiri dari dua mekanisme independen: pembuat encode yang membaca input teks dan decoder yang menghasilkan prediksi untuk tugas tersebut [14].

Ada 2 langkah dalam menggunakan metode IndoBERT, pre-training dan fine-tuning yaitu sebagai berikut:

1. Pre-training

a. Masked Language Modeling (MLM)

Menggantikan secara acak 15% kata di setiap urutan token dengan [MASK]. Pada tugas MLM, model BERT akan diberikan tugas untuk memprediksi kata bertopeng ([MASK]).

b. Next sentence prediction (NSP)

Selama pelatihan BERT, model memilih kalimat A dan kalimat B untuk setiap contoh pada prapelatihan sebelumnya, 50% dari pra-pelatihan yang benar kalimat B adalah kalimat berikutnya dari kalimat A (ditandai sebagai IsNext), dan 50% lainnya yaitu adalah acak (berlabel NotNext) dari korpus.

2. Fine-tuning

Pada tahap ini, karena model yang diinginkan adalah bentuk analisis sentimen, maka memiliki kemiripan dengan klasifikasi kalimat berikutnya dengan menambahkan layer di atas keluaran transformers dari token [CLS]. Sambungkan input dan output spesifik tugas ke BERT dan lengkapi semua parameter [15]. Untuk lebih jelasnya dapat dilihat pada tabel 1.

Tabel 1 Nilai parameter BERT

Parameter	Nilai
Dropout	0.1
Batch size	32

Learning Rate (AdamW)	2e-5
Epoch	x10

Bidirectional Encoder Representations from Transformers (BERT) memiliki beberapa parameter penting yang mempengaruhi kinerja model, yaitu

dropout digunakan untuk mencegah overfitting data dan membantu meningkatkan kemampuan generalisasi model, Batch size mempengaruhi kinerja dan kecepatan pelatihan, tingkat pembelajaran dan Pengoptimal adalah digunakan bersama seperti AdamW untuk meningkatkan bobot model dari waktu ke waktu dan epoch dibutuhkan untuk pelatihan berdasarkan ukuran kumpulan data, kompleksitas model, dan tingkat pengawasan yang diterapkan [16].

3.5. Evaluasi BERT

Setelah menyelesaikan tahap implementasi IndoBERT, uji model dengan menunjukkan hasil klasifikasi sentimen dari data test. Klasifikasi sentimen yang dihasilkan kemudian dievaluasi menggunakan confusion matrix. Hal ini menunjukkan akurasi dan performa IndoBERT saat melakukan analisis sentimen. Tahap evaluasi bertujuan untuk melihat hasil analisis sentimen kalimat ada dalam kumpulan data. Tahap evaluasi dilakukan untuk mengetahui keakuratan pemodelan yang diterapkan pada data pelatihan. Kemudian bandingkan hasil dari dua dataset berbeda dengan menerapkan confusion matrix untuk menghitung precision, recall, F-score, dan akurasi.

4 HASIL DAN PEMBAHASAN

4.1 Data Penilitian

Data	
saya sangat berterimakasih dan sangat mengapresiasi waktu yang diberikan oleh pak AS dan pak SF	saya sangat berterimakasih dan sangat mengapre...
Terkait kesediaan waktu dosen pembimbing untuk melakukan konsultasi cukup baik.	terkait kesediaan waktu dosen pembimbing untuk...
Sangat baik	sangat baik
dosen pembimbing saya pak RH dan pak FH memiliki waktu yang sangat fleksibel sesuai dengan jam	dosen pembimbing saya pak rh dan pak fh memiliki...
Menurut saya dosen pembimbing saya yg pertama (Bapak AF), kurang kesediaan waktu trkait	menurut saya dosen pembimbing saya yg pertama ...
Alhamdulillah lancar,, tergantung saya	alhamdulillah lancar,, tergantung saya
Kesediaan waktu terhadap pembimbing 1 pak RH sendiri menurut saya sudah lumayan bagus, namun	kesediaan waktu terhadap pembimbing 1 pak rh s...
pak AF waktunya selalu sibuk, setiap dosen memiliki jadwal bimbingannya tersendiri jadi para	lakukan konsultasi dengan dosen tergantung dar...
Lakukan konsultasi dengan dosen tergantung dari kesibukan dosen itu sendiri. Waktu yang	bagus
disediakan pak AS cukup untuk bimbingan, namun untuk pak AF melakukan bimbingan yang lama.	tergantung dosen itu sendiri, pak rh sering m...
Bagus	cukup baik
Tergantung dosen itu sendiri, pak RH sering meluangkan waktu dengan sangat cukup dan pak	cukup baik
Cukup baik	
untuk saya pribadi selama bimbingan, pak RH dan pak AF selalu bersedia dengan senang hati	
Waktu yang di berikan dosen sangat baik	
Menurut saya, dosen pembimbing saya pak AM dan pak FH sangat welcome dan juga baik. Mereka	

Gambar 2. Dataset

4.2 Pra-Processing

Berikut adalah hasil pra-processing data:

1. Case Folding

Tahapan pertama dalam preprocessing yaitu case folding yang digunakan untuk menghasilkan

bentuk data teks yang sama. Untuk lebih jelasnya dapat dilihat pada gambar 3 dan 4:

```
# Langkah 2: Case Folding
def case_fold_text(text):
    return text.lower()

# Melakukan case folding pada kolom 'DATA'
df['CASE_FOLDED_DATA'] = df['DATA'].apply(case_fold_text)

[ ] # Import modul yang diperlukan
import pandas as pd
from IPython.display import display
from google.colab.data_table import DataTable

# Simpan beberapa baris pertama dari kolom 'DATA' dan 'CASE_FOLDED_DATA' dalam dataframe
data = ['DATA':df['DATA'].head(15), 'CASE_FOLDED_DATA': df['CASE_FOLDED_DATA'].head(15)]
df_display = pd.DataFrame(data)

# # Tampilkan dataframe dalam bentuk tabel interaktif
# data_table.DataTable(df_display, include_index=False)

# Tampilkan tampilan interaktif dari file Excel
display(df_display)
```

Gambar 3. Kode Program untuk Case Folding

	Data	CASE_FOLDED_DATA
0	saya sangat berterimakasih dan sangat mengapre...	saya sangat berterimakasih dan sangat mengapre...
1	terkait kesediaan waktu dosen pembimbing untuk...	terkait kesediaan waktu dosen pembimbing untuk...
2	Sangat baik	sangat baik
3	dosen pembimbing saya pak RH dan pak FH memiliki...	dosen pembimbing saya pak rh dan pak fh memiliki...
4	Menurut saya dosen pembimbing saya yg pertama ...	menurut saya dosen pembimbing saya yg pertama ...
5	Alhamdulillah lancar,, tergantung saya	alhamdulillah lancar,, tergantung saya
6	Kesediaan waktu terhadap pembimbing 1 pak RH s...	kesediaan waktu terhadap pembimbing 1 pak rh s...
7	Lakukan konsultasi dengan dosen tergantung dar...	lakukan konsultasi dengan dosen tergantung dar...
8	Bagus	bagus
9	Tergantung dosen itu sendiri, pak RH sering m...	tergantung dosen itu sendiri, pak rh sering m...
10	Cukup baik	cukup baik

Gambar 4. Hasil proses Case Folding

2. Filtering

Tahapan selanjutnya dalam preprocessing yaitu filtering yang digunakan untuk menghapus karakter pada data yang ada. Untuk lebih jelasnya dapat dilihat pada gambar 5 dan 6:

```
# Simpan beberapa baris pertama dari kolom 'DATA' dan 'CASE_FOLDED_DATA' dalam dataframe
data = [df['CASE_FOLDED_DATA'].head(15), df['CASE_FOLDED_DATA'].head(15)]
df_display = pd.DataFrame(data)

# # Tampilkan dataframe dalam bentuk tabel interaktif
# data_table.DataTable(df_display, include_index=False)

# Tampilkan tampilan interaktif dari file Excel
display(df_display)
```

Gambar 6 Hasil proses Filtering

3. Tokenization

Tahapan selanjutnya dalam preprocessing yaitu tokenization proses pemecahan data yang berupa

kumpulan kalimat-kalimat akan dipecah menjadi kata-kata. Untuk lebih jelasnya dapat dilihat pada gambar 7 dan 8:

```
[ ] # Fungsi untuk tokenisasi
def tokenize_text(text):
    # Tokenisasi teks menjadi kata-kata
    tokens = word_tokenize(text)
    return tokens

# Tokenisasi kolom 'TOKENIZED_DATA' dan simpan dalam kolom 'TOKENIZED_DATA'
df['TOKENIZED_DATA'] = df['FILTERED_DATA'].apply(tokenize_text)
```

Index	Hasil	Persentase	Persentase Pakar
1	Positif	-7,60086E+16	✓
2	Positif	-7,40514E+16	✓
3	Netral	-7,20047E+16	✓
4	Positif	-7,37898E+16	✓
5	Netral	-7,14007E+16	✓
6	Positif	-8,00401E+16	✓
7	Netral	-8,73038E+16	✓
8	Positif	-8,58031E+16	✓
9	Netral	-8,54886E+16	✓

Gambar 10. Hasil pelabelan sentimen oleh pakar

Gambar 7. Kode Program *tokenization*

	FILTERED_DATA	TOKENIZED_DATA
0	saya sangat berterimakasih dan sangat mengapre... [saya, sangat, berterimakasih, dan, sangat, mengapresiasi, 'waktu', 'yang', 'diberikan', 'oleh', 'pak', 'as', 'dan', 'pak', 'sf']	Positif
1	terkait kesediaan waktu dosen pembimbing untuk... [terkait, kesediaan, waktu, 'dosen', pembimbing, 'untuk', 'melakukan', 'konsultasi', 'skripsi', 'cukup', 'baik']	Netral
2	sangat baik	Positif
3	dosen pembimbing saya pak rh dan pak fh memill... [dosen, pembimbing, saya, pak, rh, dan, pak, f...]	Netral
4	menurut saya dosen pembimbing saya yg pertama ... [menurut, saya, dosen, pembimbing, saya, yg, p...]	Netral
5	ahamdullah lancar tergantung saya	Positif
6	kesediaan waktu terhadap pembimbing 1 pak rh ... [kesediaan, waktu, terhadap, pembimbing, 1, pa...]	Netral
7	lakukan konsultasi dengan dosen tergantung da... [lakukan, konsultasi, dengan, dosen, tergantun...]	Netral
8	bagus	Positif
9	tergantung dosen itu sendiri pak rh sering me... [tergantung, dosen, itu, sendiri, pak, rh, ser...]	Netral
10	cukup baik	Positif

Gambar 4.7 Hasil proses *tokenization*

4.3 Label data

1. Pelabelan Menggunakan GPT-3

Model GPT-3 digunakan untuk menghasilkan prediksi sentimen awal. Bentuk pengolahan data terkait dengan penggolongan positif, negatif dan netral pada data yang dilabeli menggunakan GPT-3. Bentuk pengolahan data disajikan pada gambar berikut.

Data	Sentimen	Skor Sentimen
['saya', 'sangat', 'berterimakasih', 'dan', 'sangat', 'mengapresiasi', 'waktu', 'yang', 'diberikan', 'oleh', 'pak', 'as', 'dan', 'pak', 'sf']	Positif	-1,03523E+11
['terkait', 'kesediaan', 'waktu', 'dosen', 'pembimbing', 'untuk', 'melakukan', 'konsultasi', 'skripsi', 'cukup', 'baik']	Netral	-2,24127145
['sangat', 'baik']	Positif	-6,98277074
['dosen', 'pembimbing', 'saya', 'pak', 'rh', 'dan', 'pak', 'fh', 'memiliki', 'waktu', 'yang', 'sangat', 'flexibel', 'sesuai', 'dengan', 'jam', 'kampus', 'baiknya', 'ada', 'yg', 'bersedia', 'memberikan', 'bimbining', 'dilar', 'jam', 'kampus']	Positif	-9,24051E+15
['menurut', 'saya', 'dosen', 'pembimbing', 'saya', 'yg', 'pertama', 'baik', 'af', 'kurang', 'kesediaan', 'waktu', 'trikat', 'bimbining', 'skripsi', 'dian', 'satu', 'sudah', 'waktu', 'bimbining', 'sering', 'kali', 'terlambat', 'tidak', 'sesuai', 'waktu', 'yg', 'ditentukan']	Negatif	-1,39636E+11
['ahamdullah', 'lancar', 'tergantung', 'saya']	Netral	-11,03032336
['kesediaan', 'waktu', 'terhadap', 'pembimbing', '1', 'pak', 'rh', 'sendiri', 'menurut', 'saya', 'sudah', 'jumayat', 'bagus', 'namun', 'pak', 'af', 'waktunya', 'selalu', 'sibuk', 'setiap', 'dosen', 'memiliki', 'jadian', 'bimbiningannya', 'tersendiri', 'jadi', 'para', 'mahasiswa', 'dapat', 'menyeusukan', 'berdasarkan', 'jadian', 'bimbiningan', 'tersebut']	Positif	-5,2331E+12
['lakukan', 'konsultasi', 'dengan', 'dosen', 'tergantung', 'dari', 'kesiukuan', 'dosen', 'itu', 'sendiri', 'waktu', 'yang', 'disediakan', 'pak', 'as', 'cukup', 'untuk', 'bimbiningan', 'namun', 'untuk', 'pak', 'af', 'melakukan', 'bimbiningan', 'yang', 'lama', 'mahasiswa', 'perlu', 'menunggu', 'jadian', 'bimbiningan', 'yang', 'diberikan', 'dospem']	Netral	-7,05794E+16
['bagus']	Positif	-4,74914E+16
['tergantung', 'dosen', 'itu', 'sendiri', 'pak', 'rh', 'sering', 'meluangkan', 'waktu', 'dengan', 'sangat', 'cukup', 'dan', 'mubarak', 'yang', 'sibuk', 'dengan', 'pekerjaan', 'tambahan', 'maka', 'dari', 'itu', 'kurangnya', 'waktu', 'bimbiningan']	Netral	-7,94772E+15
['cukup', 'baik']	Netral	-5,11951E+16

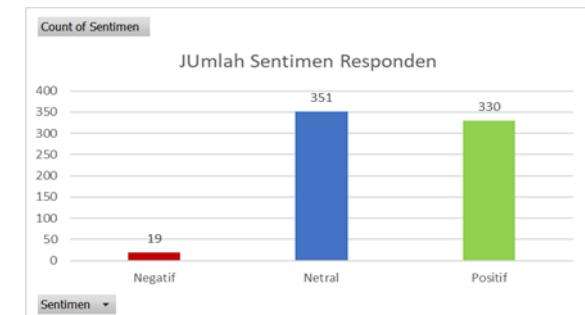
Gambar 9. Hasil pelabelan sentimen menggunakan GPT-3

2. Pelabelan Dengan Pakar

Setelah dilakukan pelabelan menggunakan GPT-3, untuk memastikan keakuratan dan kualitas label sentimen, maka 25% data pada penelitian ini dengan kemungkinan probabilitas paling rendah akan dikoreksi oleh pakar bahasa. Untuk lebih jelasnya dapat dilihat pada gambar 10.

Gambar 10. Hasil pelabelan sentimen oleh pakar

Seluruh data yang dikumpulkan berjumlah pada kuesioner berjumlah 700 data dengan rincian sentimen positif sebanyak 330, sentimen netral sebanyak 351 dan sentimen negatif sebanyak 19 data. Gambar 4.10 menunjukkan perbandingan data untuk setiap kategori sentimen. Perbedaan jumlah data pada tiap sentimen yang cukup signifikan membuat metode penyeimbangan data menggunakan BERT diterapkan pada proses selanjutnya.



Gambar 11. Perbandingan jumlah data tiap Sentimen

4.4 Implementasi BERT

Implementasi BERT pada penelitian ini menggunakan fine-tuning untuk lapisan output dalam tugas tertentu dan menggunakan library Transformers yang disediakan oleh Hugging Face, yang akan diimplementasikan menggunakan framework TensorFlow Google.

Kumpulan data harus disesuaikan dengan masukan yang dapat diterima BERT. Oleh karena itu, penelitian ini menggunakan Bert Tokenizer, sebuah tokenizer yang dirancang untuk memberi token pada kalimat dan menghasilkan masukan yang sesuai. Hal ini dilakukan karena BERT menggunakan kosa kata tertentu, tergantung model yang digunakan. Proses penyusunan kalimat menjadi representasi masukan pada BERT dilakukan oleh tokenize. Berikut pada gambar 12

untuk tokenizer BERT.

Gambar 12. Hasil Tokenizer BERT

```
Original: ['saya', 'merasa', 'sangat', 'puas', 'dengan', 'sarang', 'dan', 'nasukan', 'yang', 'telah', 'diberikan']
Tokenized: ['[', '**', 'saya', '***', 'merasa', '***', 'sangat', '***', 'puas', '***', 'dengan', '***', 'sarang', '***', 'dan', '***', 'nasukan', '***', 'yang', '***', 'telah', '***', 'diberikan', '**', ']']
Token IDs: [138, 112, 44334, 112, 117, 112, 86399, 112, 117, 112, 20514, 112, 117, 112, 25208, 10416, 112, 117, 112, 10722, 112, 117, 112, 96103, 112, 117, 112, 10214, 112, 117, 112, 33722, 10208, 112, 117, 112, 10290, 112, 117, 112, 12184, 112, 117, 112, 43496, 112, 140]
```

BERT akan menerima rangkaian kata atau kalimat sebagai inputan dan akan dilanjutkan melalui tumpukan encoder. Setiap encoder menerapkan self- attention dan memberikan output melalui jaringan feed-forward, yang kemudian dilanjutkan oleh encoder berikutnya.

Karena penelitian ini menggunakan model base IndoBERT, maka setiap token pada setiap posisi memberikan keluaran berupa vektor setelah melewati seluruh encoder. Untuk proses analisis sentimen, output yang dipertimbangkan adalah output dari posisi pertama yaitu token [CLS], karena token ini dianggap sebagai rata-rata token kata yang dikumpulkan untuk mendapatkan vektor kalimat dan diakhiri kalimat yaitu token [SEP]. Berikut pada gambar 13 untuk Hasil token [CLS] dan [SEP].

```
Original: ['saya', 'merasa', 'sangat', 'puas', 'dengan', 'sarang', 'dan', 'nasukan', 'yang', 'telah', 'diberikan']
Token IDs: [101, 138, 112, 44334, 112, 117, 112, 86399, 112, 117, 112, 20514, 112, 117, 112, 25208, 10416, 112, 117, 112, 10722, 112, 117, 112, 96103, 112, 117, 112, 10214, 112, 117, 112, 33722, 10208, 112, 117, 112, 10290, 112, 117, 112, 12184, 112, 117, 112, 43496, 112, 140, 102]
```

Gambar 13. Hasil token [CLS] dan [SEP]

Setiap kata dalam kalimat telah diubah menjadi token ID. Token ID adalah representasi numerik unik dari setiap kata dalam kamus tokenizer. Token[101] mewakili token khusus [CLS], yang merupakan token awal setiap kalimat. Token [102] adalah token [SEP] yang menandai akhir setiap kalimat. Dari [138, 112, 44334, 112, ...] kita dapat melihat rangkaian token yang mewakili kata-kata dalam kalimat tersebut. Tag 112 adalah tag khusus untuk sebuah kata dalam sebuah kalimat.

Proses pengembangan model BERT melibatkan penggunaan library yang disebut transformer dan kelas yang disebut BertFor Sequence Classification. Kelas ini khusus digunakan untuk fine-tuning. Langkah selanjutnya dalam pengembangan model BERT adalah memuat model bert_base_case yang telah dilatih sebelumnya di library transformers. Fungsi optimasi yang diterapkan pada model ini disebut AdamW.

1. Fine-Tuning

Penelitian ini menetapkan jumlah Epoch menjadi 10 dan menyimpan model terbaik untuk pengujian berdasarkan data validasi. Melakukan fine-tuning dengan menggunakan hyperparameters

yang digunakan dalam penelitian ini yaitu batch size 32, learning rate (Adam) 2e-5, sehingga hasil yang diperoleh menunjukkan penyempurnaan model BERT pada kumpulan data menggunakan label berbasis skor.

Dataset yang digunakan pada setiap tahap training, validasi, dan pengujian dilakukan tanpa menentukan random seed. Jadi setiap kali melakukan analisis sentimen, kumpulan data pelatihan, validasi, dan pengujian berbeda. Berikut dapat dilihat pada gambar 14 untuk training Epoch dan gambar 15 untuk hasil kurva training Epoch.

```
===== Epoch 1 / 10 =====
Training...
    Average training loss: 0.90
    Training epoch took: 0:00:09
Running Validation...
    Accuracy: 0.41
    Validation took: 0:00:00
===== Epoch 2 / 10 =====
Training...
    Average training loss: 0.77
    Training epoch took: 0:00:06
Running Validation...
    Accuracy: 0.78
    Validation took: 0:00:00
===== Epoch 3 / 10 =====
Training...
    Average training loss: 0.62
    Training epoch took: 0:00:05
Running Validation...
    Accuracy: 0.65
```

Gambar 14. Hasil training Epoch



Gambar 15. Kurva hasil training Epoch

Grafik garis diberi label kurva hasil performa *training epoch*, label ini menunjukkan bahwa diagram garis menunjukkan hasil pelatihan model. Sumbu X menunjukkan *Epoch*, yaitu jumlah iterasi pelatihan yang telah dilakukan. *Epoch* dimulai dari 0 dan meningkat menjadi 8. Sumbu Y menunjukkan loss, yang mengukur seberapa buruk model dalam memprediksi data pelatihan. *Loss* diukur pada skala logaritma, sehingga nilai kerugian yang lebih rendah menunjukkan kinerja yang lebih baik.

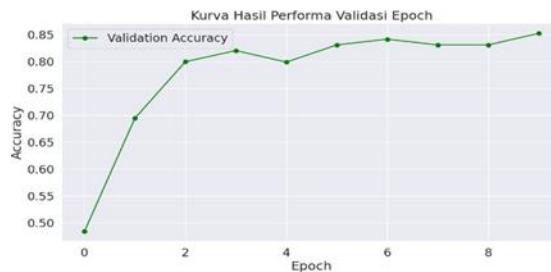
Grafik dimulai pada *epoch* 0, awal pelatihan dan *loss* pelatihan masih tinggi, sekitar 0,9. Pada *epoch* 0, model baru saja memulai pelatihan. Model juga tidak memiliki informasi tentang data pelatihan, sehingga *loss* pelatihan akan tinggi. Pada *epoch* 2, model telah mempelajari tentang data pelatihan, sehingga *loss* pelatihan akan lebih rendah dibandingkan pada *epoch*

0. *Epoch* 2, rata-rata latihan turun menjadi sekitar 0,8. Pada *epoch* 4, jumlah rata-rata pelatihan

turun menjadi sekitar 0,6. Selanjutnya pada *epoch* ke 6, model telah mempelajari seluruh data pelatihan, sehingga *loss* pelatihan akan mencapai titik terendah. *Epoch* 8 adalah titik di mana pelatihan model selesai. Pada titik ini, model tidak belajar lagi dari data pelatihan. *Loss* dipelatihan pada *Epoch* 8 mungkin tetap sama seperti *epoch* 6 atau bahkan sedikit meningkat karena model mungkin mulai melakukan *overfit* pada data pelatihan. *Loss* pelatihan selama periode ini akan mencapai titik terendah, yang menunjukkan bahwa model memprediksi data pelatihan dengan akurasi maksimum.

4.5 Evaluasi Model

Hasil kurva evaluasi penelitian ini pada gambar 16 menunjukkan hasil kurva validasi *epoch* dan gambar 17 menunjukkan perbandingan hasil *training* dan validasi *epoch* model IndoBERT. Berikut tampilan gambar hasil perbandingan antara akurasi yang ada diperoleh selama pelatihan dan validasi pada gambar dibawah.



Gambar 16. Kurva hasil performa validasi

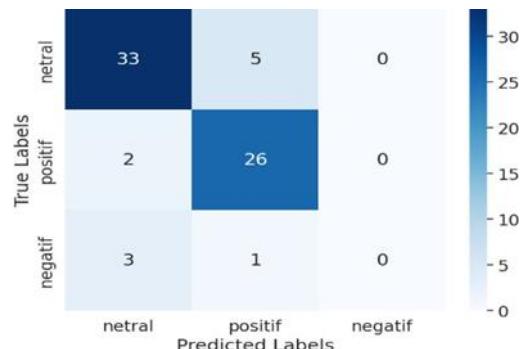
Gambar 16 pada epoch ke 8, akurasi model mencapai nilai maksimum 0,85 yaitu model dapat memprediksi label data validasi dengan akurasi 85%.



Gambar 17. Kurva perbandingan performa training dan validasi pada 10 perulangan (epoch). Selanjutnya penelitian ini menggunakan confusion matrix dan classification report guna melakukan pengecekan hasil pengujian. Dalam penelitian ini, *confusion*.

matrix memiliki tiga kategori: netral, positif dan negatif. Nilai dalam confusion matrix mewakili seberapa banyak data yang diprediksi dengan benar oleh model untuk setiap kategori. Untuk lebih

jelas- nya dapat dilihat pada gambar 18.



Gambar 18. Confusion matrix

Berdasarkan gambar confusion matrix akurasi model klasifikasi opini di atas, berikut dapat dilihat pada tabel 4.2 adalah nilai *True Positive* (TP), *False Positive* (FP), *False Negative* (FN), *True Neutral*

Tabel 2. Confusion matrix

		Predicted Class		
		Netral	Positif	Negatif
True Class	Netral	TNt = 33	FP = 5	FN = 0
	Positif	FNt = 2	TP = 26	FN = 0
	Negatif	FNt = 3	FP = 1	TN = 0

Tabel *confusion matrix* digunakan pada penelitian ini untuk menghitung matriks seperti Accuracy, Precision, Recall, dan fi-score untuk di setiap kategori emosi dan memberikan kemampuan model untuk dapat mengklasifikasikan data di setiap kategori, dengan melihat confusion matrix maka dapat dihitung nilai Accuracy, Precision, Recall, dan fi-score dari model klasifikasi.

Untuk mengetahui nilai akurasinya, perlu menghitung proporsi data yang diprediksi dengan benar terhadap keseluruhan data seperti pada persamaan (1).

$$Accuracy = \frac{33 + 26 + 0}{70} = 0.84 \dots \dots \dots (1)$$

Nilai Precision diperoleh dengan cara membandingkan data yang diklasifikasikan secara benar dengan prediksi jumlah data untuk kategori tersebut. Nilai Precision pada model klasifikasi ini dihitung pada persamaan (2), (3) dan (4).

$$P(\text{Netral}) \frac{33}{33+2+3} = 0.87 \dots \dots \dots (2)$$

$$P(\text{Positif}) \frac{26}{26+5+1} = 0.81 \dots \dots \dots (3)$$

$$P(\text{Negatif}) \frac{0}{0+0} = 0 \dots \dots \dots (4)$$

Recall menggambarkan keberhasilan model dalam mengambil informasi. Recall diperoleh

dengan menghitung data yang dapat diklasifikasikan dengan benar dibandingkan dengan seluruh data yang seharusnya masuk dalam kelas tersebut. Tingkat recall dihitung sebagai berikut.

F1-Score diperoleh dengan menggabungkan nilai Precision dan recall. Dihitung dengan membandingkan 2 kali recall dan presisi dengan jumlah recall dan Precision.

$$F1-Score_{Neutral} = \frac{2 * (0.87 * 0.87)}{0.87 + 0.87} = 0.87 \dots \dots \dots (8)$$

(TN), *False Neutral* (F_{Nt}), dan *False Negative* (FN) untuk masing-masing sentimen.

Setelah memahami model confusion matrix yang terlihat pada gambar, proses perhitungan dilakukan untuk menghitung nilai accuracy, precision, recall dan f1-score pada model BERT. Hasil dari perhitungan tersebut dapat dilihat pada gambar Tabel 3 dibawah menjelaskan akurasi model mendapat skor cukup tinggi. Akurasi model pada kumpulan data pelatihan adalah 84%, yang berarti model berhasil mengklasifikasikan 84% dari sentimen dalam kumpulan data training dengan benar. Untuk Precision netral 87%, Precision Positif 81%, Precision Negatif 0% dan pada Recall netral 87%, Recall Positif 93%, Recall Negatif 0% sedangkan pada F1-Score Netral 87%, F1-Score Positif 87%, dan F1-Score Negatif nya 0%.

Tabel 3 *Classification Report*

Accuracy: 0.84			
Classification Report			
	Precision	Recall	F-Score
Netral	0.87	0.87	0.87
Positif	0.81	0.93	0.87
Negatif	0.00	0.00	0.00

Hasil penelitian ini menunjukkan bahwa model BERT yang digunakan pada penelitian ini memberikan akurasi keseluruhan sebesar 84%. Proses fine-tuning dilakukan dengan jumlah epoch sebanyak 10, menggunakan batch size 32, dan learning rate (Adam) sebesar $2e-5$. Hasil fine-tuning model menunjukkan bahwa performa meningkat seiring bertambahnya jumlah epoch.

5 KESIMPULAN

Penelitian sebelumnya lebih condong pada analisis sentimen yang bersifat umum dan mengumpulkan data secara luas di sosial media, sedangkan penelitian ini fokus pada interaksi antara dosen dan mahasiswa di prodi informatika, walaupun fokus penelitiannya berbeda, namun semuanya menunjukkan efektivitas di model BERT dalam memahami dan melakukan mengklasifikasikan sentimen. Berdasarkan hasil penelitian yang telah dilakukan, analisis saya terhadap penelitian ini adalah sebagai berikut: Hasil pengujian menunjukkan BERT mampu melakukan klasifikasi sentimen dengan akurasi 84%. Nilai Precision dan recall tertinggi pada sentimen netral dan positif. Sedangkan pada sentimen negatif, model BERT mengalami kesulitan untuk mengklasifikasikan. Hasil penelitian sejalan dengan penelitian sebelumnya, yang menunjukkan kemampuan BERT dalam analisis sentimen. Namun, perbedaan metode dan di data dapat mempengaruhi hasil.

DAFTAR PUSTAKA

- [1] Haryati, R., 2012., Survey kinerja dosen pembimbing skripsi dan kualitas skripsi mahasiswa akuntansi stie malangkucewara. Jurnal Dinamika Akuntansi, 4(2), 121–128.
 - [2] Habibi, M., (2019)., Analisis Sentimen dan Klasifikasi Komentar Mahasiswa pada Sistem Evaluasi Pembelajaran Menggunakan Kombinasi KNN Berbasis Cosine Similarity dan Supervised Model. Departemen Ilmu. Komputer Dan Elektronika, Fakultas Matematika Dan Ilmu Pengetahuan Alam.
 - [3] Agus Sidiq Purnomo., 2020., Analisis Sentimen Untuk Respon Terhadap Universitas (Studi Kasus : Universitas Mercu Buana Yogyakarta). Journal Of Information System And Artificial Intelligence (JISAI), X(X), 53–60.
 - [4] Surya Agustian., 2022., Analisis Sentimen Masyarakat Terhadap Kebijakan Vaksinasi Covid-19 pada Media Sosial Twitter menggunakan Metode Logistic Regression. Jurnal CoSciTech (Computer Science and Information Technology), 3(2), 99–106. <https://doi.org/10.37859/coscitech.v3i2.3836>.
 - [5] Kinasih, H. W., 2021., Peran Dosen Pembimbing Dalam Lulus Tepat Waktu Mahasiswa : Study Pada Mahasiswa Akuntansi Universitas X. Proceeding SENDU, 208–214.
 - [6] Santoso., 2022., Analisis Sentimen Mahasiswa Terkait Pembelajaran Tatap Muka Menggunakan Metode Naive Bayes Classifier. Techno.Com, 21(3), 644–654. <https://doi.org/10.33633/tc.v21i3.6262>
 - [7] Wahyudi, M. D. (2019). Analisis Sentimen pada Indeks Kinerja Dosen Fakultas SANTEK UIN Menggunakan Naive Bayes Classifier. Jurnal Buana Informatika,. <https://doi.org/10.24002/jbi.v10i2.2250>.

- [8] Patwardhan, N., Marrone, S., & Sansone, C., 2023., Transformers in the Real World: A Survey on NLP Applications. *Information*. <https://doi.org/10.3390/info14040242>.
- [9] Saputra, R., 2022., Implementasi Bidirectional Encoder Representations From Transformers (BERT) Untuk Mendeteksi Hatespeech. *UIN SUSKA RIAU*.
- [10] and rating of tourist reviews using machine learning. *Journal of Hospitality and Tourism Insights*. <https://doi.org/10.1108/JHTI-02-2022-0078>
- [11] Vaswani, (2019). Attention is all you need. *Advances in Neural Information Processing Systems*, 2017-Decem(Nips), 5999–6009.
- [12] Devlin, J., (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. - *Proceedings of the Conference*, 1, 4171–4186.
- [13] Viñán-Ludeña., (2022). Discovering a tourism destination with social media data: BERT-based sentiment analysis. *Journal of Hospitality and Tourism Technology*.
- [14] <https://doi.org/10.1108/JHTT-09-2021-0259>. Latsch, S., & Thygesen, E., 2020., Fact Extraction and Verification in Danish. 1–55.
- [15] Hadiyan., (2021)., Deteksi Penggunaan Kalimat Abusive Pada Teks Bahasa Indonesia Menggunakan Metode IndoBERT. *E-Proceeding of Engineering*, Vol.8, No.(2), 3028–3038.
- [16] Raden., (2021)., Analisis Sentimen Customer Review Aplikasi Ruang Guru dengan Metode BERT (Bidirectional Encoder Representations from Transformers). *Jeisbi*, 02(03) Utami, L. A., Gani, A., & Suparni, S. 2020. Penerapan Metode Webqual 4.0 dan IPA Dalam Mengukur Kualitas Website VISLOG PT. Citra Surya Indonesia. *Komputika : Jurnal Sistem Komputer*, 9(1), 25–34. <https://doi.org/10.34010/komputika.v9i1.2849>.