

---

## Analisis Sentimen Perbandingan Lange Language Model Study kasus: Gpt, Gemini dan Llama Menggunakan BERT

Linda Safitri<sup>\*1</sup>, Adhitia Erfina<sup>2</sup>

<sup>1,2</sup>) Sistem Informasi, Fakultas Teknik Komputer dan Desain, Universitas Nusa Putra

Email: <sup>\*1</sup>[linda.safitri\\_si21@nusaputra.ac.id](mailto:linda.safitri_si21@nusaputra.ac.id), <sup>2</sup>[adhitia.erfina@nusaputra.ac.id](mailto:adhitia.erfina@nusaputra.ac.id)

(Naskah masuk: 22 Januari 2025, diterima untuk diterbitkan: 20 Juli 2025)

**Abstrak:** Perkembangan Large Language Models (LLMs) seperti GPT, Gemini, dan LLaMA telah mendorong penelitian lebih lanjut mengenai efektivitas dan sentimen yang dihasilkan oleh model-model ini dalam berbagai aplikasi. Studi ini bertujuan untuk melakukan analisis sentimen perbandingan terhadap ketiga model tersebut dengan pendekatan berbasis BERT (Bidirectional Encoder Representations from Transformers). Metode yang digunakan mencakup pengumpulan data dari keluaran masing-masing model dalam berbagai konteks, analisis sentimen menggunakan BERT, serta evaluasi perbedaan dalam distribusi sentimen positif, negatif, dan netral. Hasil penelitian ini menunjukkan model GPT menunjukkan sentimen positif tertinggi (49,54%) dibandingkan dengan Gemini (38,76%) dan Llama (35,98%). Sementara itu, model Llama memiliki sentimen negatif tertinggi (43,47%), diikuti oleh Gemini (41,68%) dan GPT (35,21%). Untuk sentimen netral, Llama memiliki persentase tertinggi (20,55%), sedangkan Gemini dan GPT memiliki persentase yang lebih rendah, masing-masing 13,56% dan 15,25%. Secara keseluruhan, GPT cenderung menghasilkan lebih banyak sentimen positif dibandingkan dengan model lainnya, sedangkan Llama dan Gemini lebih banyak menghasilkan sentimen negatif. Secara umum, dapat disimpulkan bahwa GPT menerima persepsi publik yang lebih positif dibandingkan Gemini dan LLaMA dalam konteks yang dianalisis dalam studi ini. Hal ini mengindikasikan bahwa pengguna atau pihak yang memberikan sentimen cenderung memiliki pengalaman atau pandangan yang lebih baik terhadap GPT. Hasil penelitian ini diharapkan dapat memberikan wawasan mengenai keunggulan dan kelemahan masing-masing LLM dalam memahami dan menghasilkan teks yang mencerminkan sentimen secara akurat. Temuan ini juga dapat membantu dalam pemilihan model yang lebih sesuai untuk aplikasi spesifik yang memerlukan analisis sentimen tingkat lanjut.

**Kata Kunci** – Analisis Sentimen; Large Language Models; GPT; Gemini; LLaMA; BERT

---

## Sentiment Analysis Comparative Lange Language Model Case Study: Gpt, Gemini and Llama Using BERT

**Abstract:** The development of Large Language Models (LLMs) such as GPT, Gemini, and LLaMA has encouraged further research into the effectiveness and sentiment generated by these models in various applications. This study aims to conduct comparative sentiment analysis of the three models using a BERT (Bidirectional Encoder Representations from Transformers) based approach. The methods used include collecting data from the output of each model in various contexts, sentiment analysis using BERT, as well as evaluating differences in the distribution of positive, negative and neutral sentiment. The results of this research show that the GPT model shows the highest positive sentiment (49.54%) compared to Gemini (38.76%) and Llama (35.98%). Meanwhile, the Llama model had the highest negative sentiment (43.47%), followed by Gemini (41.68%) and GPT (35.21%). For neutral sentiment, Llama has the highest percentage (20.55%), while Gemini and GPT have lower percentages, 13.56% and 15.25% respectively. Overall, GPT tends to generate more positive sentiment compared to other models, while Llama and Gemini generate more negative sentiment. In general, it can be concluded that GPT receives more positive public perception than Gemini and LLaMA in the context analyzed in this study. This indicates that the user or party providing the sentiment tends to have a better experience or view of GPT. It is hoped that the results of this research will provide insight into the strengths and weaknesses of each LLM in understanding and producing text that accurately reflects sentiment. These findings may also help in the selection of more appropriate models for specific applications that require advanced sentiment analysis.

**Keywords** – Sentiment Analysis; Large Language Models; GPT; Gemini; LLaMA; BERT

---

## 1. PENDAHULUAN

Perkembangan teknologi kecerdasan buatan (AI) telah membawa dampak besar dalam berbagai bidang, termasuk pemrosesan bahasa alami (NLP). Model bahasa besar (Large Language Models atau LLMs), seperti GPT dari OpenAI, LLaMA dari Meta, dan Gemini dari Google DeepMind, merupakan terobosan yang memungkinkan mesin memahami, menghasilkan, dan merespons teks dengan tingkat akurasi yang luar biasa. LLMs menggunakan arsitektur berbasis transformer untuk memproses informasi dalam jumlah besar, memungkinkan model ini mempelajari pola bahasa dari berbagai domain. Keunggulan ini menjadikan LLMs sebagai fondasi untuk berbagai aplikasi, mulai dari chatbot, sistem rekomendasi, hingga alat analisis teks canggih yang digunakan di industri dan akademisi.

Salah satu aplikasi utama NLP adalah analisis sentimen, yaitu teknik untuk mengidentifikasi opini atau emosi dalam teks. Analisis ini banyak diterapkan dalam berbagai bidang, seperti pemasaran, layanan pelanggan, dan politik. Dengan memanfaatkan model seperti BERT (Bidirectional Encoder Representations from Transformers), analisis sentimen dapat dilakukan dengan lebih akurat karena kemampuan BERT untuk memahami konteks dua arah dalam sebuah teks. Keunggulan ini menjadikan BERT sebagai salah satu model unggulan dalam klasifikasi teks dan tugas NLP lainnya, termasuk ekstraksi informasi dan pemrosesan bahasa lintas domain.

Berberapa penelitian terkait yang dilakukan oleh Alessio Buscemi, Daniele Proverbio 2024 [1] yang berjudul "GPT vs Gemini vs LLaMA on Multilingual Sentiment Analysis" Penelitian ini mengevaluasi kinerja model ChatGPT, Gemini, dan LLaMA dalam analisis sentimen multibahasa, khususnya pada teks ambigu atau ironis. Hasilnya menunjukkan bahwa ChatGPT dan Gemini mampu menangani skenario ambigu dengan baik, namun terdapat bias dan inkonsistensi performa di antara model dan bahasa yang dievaluasi. Lalu penelitian tentang "Perbandingan Performa Open Source LLM IndoBERT, OpenAI GPT, dan Gemini terhadap Sentiment Analysis dalam Bahasa Indonesia: Studi Kasus pada TokoTeknoPro PT VinTechnology" dilakukan oleh Arida Amalia Rosa 2024 [2] Studi ini membandingkan performa model IndoBERT, GPT, dan Gemini dalam analisis sentimen ulasan produk berbahasa Indonesia. Hasilnya menunjukkan bahwa IndoBERT lebih unggul dengan nilai F1 sebesar 0,9, akurasi 0,94, recall 0,89, dan presisi 0,9.. Studi-studi tersebut mencakup pengembangan metode fine-tuning model, peningkatan efisiensi algoritma, serta integrasi dengan teknologi lain seperti pemrosesan citra dan analitik data. Contohnya, penelitian terbaru menunjukkan bagaimana GPT dan BERT dapat digabungkan untuk menghasilkan sistem rekomendasi berbasis teks yang lebih canggih. Selain itu, terdapat penelitian tentang penerapan LLMs dalam bidang hukum dan medis, yang menunjukkan potensi besar model ini untuk mendukung keputusan berbasis data. Penelitian semacam ini tidak hanya memperkuat fondasi teoritis dalam NLP, tetapi juga membuka jalan bagi inovasi di berbagai sektor industri.

penelitian sebelumnya telah dilakukan yang berkaitan dengan judul "SentimentGPT: Exploiting GPT for Advanced Sentiment Analysis and its Departure from Current Machine Learning" oleh Kiana Kheiri dan Hamid Karimi 2023. Penelitian ini membahas berbagai metode GPT dalam analisis sentimen, termasuk teknik prompt engineering, fine-tuning, dan pendekatan embedding classification. Hasil penelitian menunjukkan keunggulan signifikan dari pendekatan berbasis GPT dalam hal kinerja prediktif dibandingkan model lain yang ada [3].

"Analisis Sentimen Pengguna Aplikasi Gen Ai dari Appstore dan Googleplay Menggunakan Algoritma C45" oleh Widia Ningsih USTI et al. 2024. Penelitian ini menganalisis sentimen pengguna terhadap aplikasi Generative AI seperti ChatGPT, Bing AI, Microsoft Co-Pilot, Gemini AI, dan Da Vinci AI yang tersedia di AppStore dan Google Play. Data ulasan pengguna diklasifikasikan menjadi sentimen positif, negatif, atau netral menggunakan algoritma C4.5. Hasil penelitian menunjukkan bahwa algoritma C4.5 mampu mengklasifikasikan sentimen dengan akurasi, presisi, dan recall yang tinggi [4].

"Analisis Sentimen Chat GPT sebagai Masa Depan Pekerja pada Media Sosial Youtube menggunakan Algoritma Klasifikasi Naive Bayes" oleh Muhammad Yusril Pratama et al. 2023. Studi ini fokus pada implementasi analisis sentimen dalam percakapan di YouTube dengan bantuan Chat

GPT dan teknik klasifikasi Naive Bayes. Hasilnya menunjukkan akurasi yang tinggi dalam klasifikasi sentimen, meskipun tantangan seperti variasi bahasa dan konteks percakapan dapat mempengaruhi hasil analisis [5].

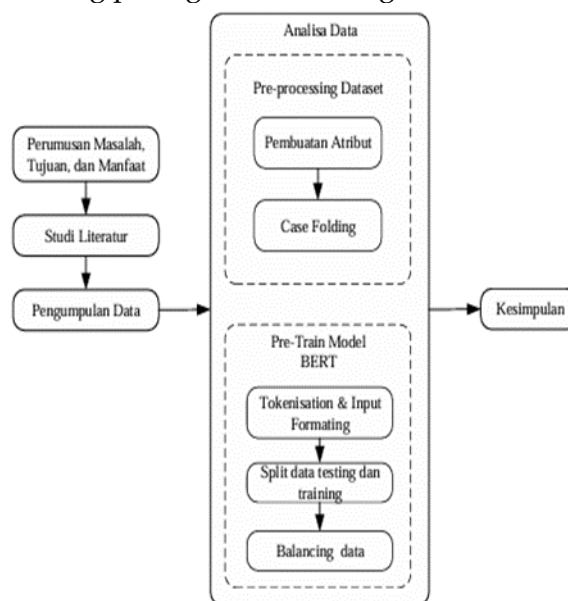
"Analisis Sentimen terhadap Pelayanan Kesehatan berdasarkan Ulasan Google Maps menggunakan BERT" oleh Ardiansyah et al. 2023. Penelitian ini menganalisis sentimen ulasan pengguna terhadap pelayanan kesehatan menggunakan model BERT. Hasil penelitian menunjukkan bahwa BERT efektif dalam mengklasifikasikan sentimen dengan akurasi yang tinggi [6].

"Bidirectional Encoder Representations from Transformers (BERT): A Sentiment Analysis Odyssey" oleh Shivaji Alaparthi dan Manit Mishra 2020. Studi ini membandingkan efektivitas empat teknik analisis sentimen, termasuk model BERT, dan menemukan bahwa BERT menunjukkan keunggulan signifikan dalam analisis sentimen teks[7].

"Analisis Sentimen Sederhana Menggunakan Algoritma LSTM dan BERT untuk Klasifikasi Data Spam dan Non-Spam" oleh Prismahardi Aji Riyantoko et al. (2022). Penelitian ini menganalisis sentimen menggunakan model LSTM dan BERT untuk mengklasifikasikan data spam dan non-spam, dengan hasil menunjukkan bahwa BERT mencapai tingkat akurasi yang lebih tinggi[8].

## 2. METODE PENELITIAN

Langkah awal penelitian adalah menentukan permasalahan yang akan diselesaikan, tujuan, manfaat dan mencari literatur sebagai pendukung penyelesaian permasalahan yang sudah ditentukan. Hasil literatur disusun untuk merumuskan hipotesis terhadap sumber masalah dan data yang akan diolah. Dari hasil hipotesis, lalu dilakukan analisis dengan Bidirectional Encoder Representations from Transformers (BERT), dan dilanjutkan membuat sentimen analisis dan performansi dari BERT. Hasil sentimen analisis tersebut, peneliti membuat kesimpulan. Tahapan metode penelitian diatas tertuang pada gambar 1, sebagai berikut :



Gambar 1. Tahapan BERT

### 2.1. Pengumpulan Data

Pengumpulan data dilakukan proses mendapatkan informasi dari berbagai sumber untuk analisis lebih lanjut. Dalam konteks komentar YouTube, pengumpulan data melibatkan ekstraksi komentar dari video YouTube untuk dianalisis. Menggunakan YouTube Data API untuk memungkinkan pengembang mengakses data dari platform YouTube. API ini memungkinkan pengguna untuk mengambil data seperti video, komentar, playlist, dan informasi kanal. Data yang di ambil dari 3 video youtube yang membahas GPT, Gemini dan Llama.

## 2.2. Pre-Processing Text

Tahapan pre-processing merupakan langkah penting untuk penyempurnaan Data mentah menjadi bisa diolah. Data yang masih terdapat simbol, hastag, dan kata yang tidak mengandung makna akan diperbaiki melalui tahapan ini. pre-processing pada penelitian ini terdiri dari case folding dan pembuatan atribut.

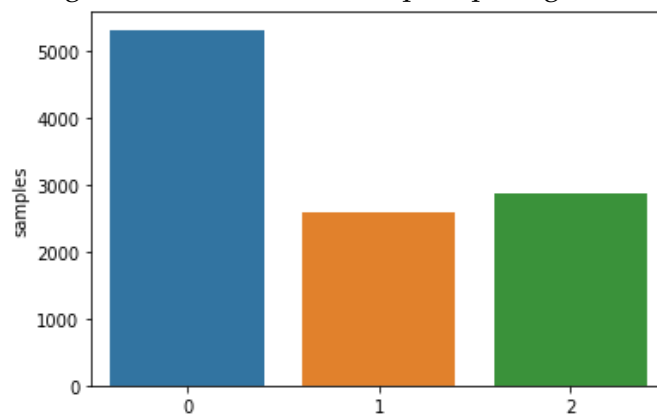
## 2.3. Pelabelan Data

Hasil dari pre processing komentar bahasa Inggris bersih, kemudian dilakukan pelabelan. Untuk optimalisasi, dipakai. Berikut contoh hasil pelabelan komentar. Hasil pelabelan terlihat pada tabel

Tabel 1. Pelabelan komentar

Komentar	Sentimen	Label
gpt is the end of the world now	Negative	0
The potentials of ChatGPT is very huge. It helps me a lot in	Positive	1
Moderation? But of course.	Netral	2

Sentimen negative adalah komentar yang berisi kata kata hasutan, hinaan terhadap video yang diunggah. Komentar tersebut akan diberikan label 0. Sentimen positif adalah komentar yang memuji, memberikan dukungan, atau memberikan umpan balik positif video yang diunggah. Komentar tersebut akan diberi label 1. Komentar dengan sentimen netral dimana tidak terlihat pendapat positif ataupun negatif akan diberi label 2 seperti pada gambar 2 berikut.



Gambar 2. Pelabelan Data

## 2.4. Split Data

Sebelum dilakukan klasifikasi, data yang telah dilabeli dibagi menjadi 3 yaitu data train, data test, dan data validation dengan menggunakan metode sklearn train test split. Data train digunakan pengembangan model, data test digunakan untuk menguji dan melihat keakuratan model, sedangkan data validation digunakan untuk memvalidasi kinerja model dan meminimalisir overfitting yang sering terjadi. Scikit learn merupakan library Python yang sering digunakan untuk split data. Pada penelitian ini sebanyak 80% data berlabel digunakan untuk membangun model, sedangkan data lainnya akan digunakan untuk menerapkan model melakukan klasifikasi sentiment. Pembagian data ke dalam data Training dan data Testing adalah sebanyak 80% data Training,

10% data validation, dan 10% data Testing. Hasil dari pembagian label terlihat pada gambar 3 di bawah ini.

```

== Train ==
Input: (405, 275)
Label: (405,)
Mask: (405, 275)

== Validation ==
Input: (45, 275)
Label: (45,)
Mask: (45, 275)

== Test ==
Input: (50, 275)
Label: (50,)
Mask: (50, 275)
    
```

Gambar 3. Pebagian Data yang telah dilabeli.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Hasil

Data yang telah di bagi menjadi 3 bagian, akan dilakukan proses klasifikasi dengan tahapan sebagai berikut.

##### 3.1.1. Pre-Train Model BERT

Dataset yang digunakan harus disesuaikan dengan representasi input yang diterima oleh BERT. Setelah didapatkan dataset yang bersih hasil preprocessing selanjutnya dilakukan penambahan token khusus yaitu [CLS] pada awal kalimat, dan token [SEP] pada tiap akhir kalimat sebagai token pemisah antar kalimat. Kemudian dilakukan pengkodean (encoding) menggunakan tokenizer sesuai indeks vocabulary dari pre-trained BERT. BERT telah memiliki vocabulary yang nantinya akan digunakan sebagai proses transfer learning.

Pada penelitian ini, jumlah kata paling banyak dari semua komentar adalah 272 kata. Agar kalimat tidak terjadi pemisahan kata, maka ditetapkan nilai maksimal untuk kata yang digunakan dalam 1 data yaitu 275 kata. Jika kalimat lebih pendek akan dilakukan padding dengan menambahkan token [PAD]. Untuk kata yang tidak ada pada vocabulary (out-of vocabulary) dilakukan pemisahan menjadi sub kata (subword) dengan symbol ##. Sebagai contoh proses penyesuaian input yang dapat diterima BERT disajikan pada tabel berikut.

Tabel 4. Proses Pre-Train Model

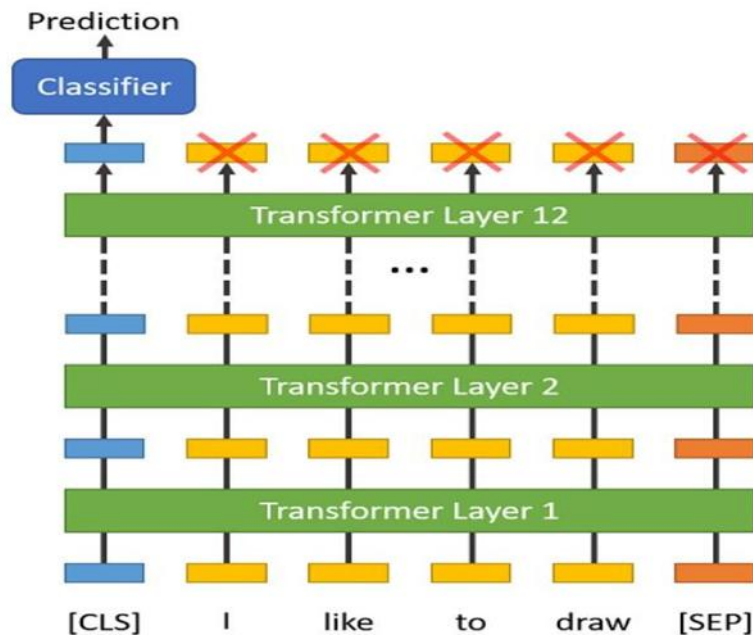
<b>Kalimat</b>	chat gpt is the end of the world now!
Hasil tokenisasi	'ch', '##at', 'gp', '##t', 'i', '##s', 'end', 'of', 'the', worl '##d',
Penambahan token khusus	'[CLS]', 'bag', '##us', 'tu', '##h', 'hap', '##e', 'nya', 'bang', 'keren',
Encoding	101, 35126, 10258, 10689, 10243, 55558, 10111, 19483, 12221,
Attention mask	1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1

Hasil dari proses tokenisasi kemudian ditambahkan token khusus yaitu token [CLS] di awal kalimat dan token [SEP] di akhir kalimat. Selanjutnya pada tiap-tiap token dikodekan sesuai dengan indeks vocabulary. Token [CLS] memiliki id 101, token 'bag' memiliki id 35126, dan seterusnya

Attention mask digunakan untuk membedakan nilai token kata dengan nilai padding. Toked padding [PAD] memiliki nilai 0 sedangkan token kata diberi nilai 1. Untuk memudahkan proses tokenisasi dibuat suatu data loader untuk masing-masing dataset yaitu data loader untuk Training, data loader untuk Testing, dan data loader untuk validasi.

### 3.1.2. Fine-tuning BERT

Selanjutnya input yang telah disesuaikan kemudian diteruskan ke dalam jaringan BERT yaitu tumpukan 12 layer Transformers Encoder. Tiap layer encoder memiliki dua sub layer, yang pertama adalah multi-head self-attention mechanism, dan yang kedua fully connected feed forward network. Setelah melewati semua encoder, didapatkan vector output dari tiap token. Namun hanya vector output dari token [CLS] yang akan digunakan sebagai vector input untuk classifier.



Gambar 4. Ilustrasi Fine-tuning BERT untuk Klasifikasi

Proses fine-tuning BERT untuk klasifikasi dilakukan dengan menambahkan layer klasifikasi. Library Transformers memiliki kelas BertForSequenceClassification yang didesain untuk tugas klasifikasi. Kelas BertForSequenceClassification bekerja dengan cara memasukkan output dari pooler untuk menghitung logits. Nilai logika yang dihasilkan kemudian digunakan untuk mendapatkan nilai prediksi menggunakan softmax. Pada penelitian ini, penulis menggunakan learning rate sebesar

2e-5 dengan optimizer adam dengan epsilon 1e-8. Berikut hasil Training dengan epoch 10.

```

===== Epoch 1 / 10 =====
Training...
Average training loss: 0.95
Training epoch took: 0:00:19
Running Validation...
Accuracy: 0.64
Validation took: 0:00:01
===== Epoch 2 / 10 =====
Training...
Average training loss: 0.81
Training epoch took: 0:00:16
Running Validation...
Accuracy: 0.64
Validation took: 0:00:01
===== Epoch 3 / 10 =====
Training...
Average training loss: 0.72
Training epoch took: 0:00:16
Running Validation...
Accuracy: 0.67
Validation took: 0:00:01
===== Epoch 4 / 10 =====
Training...
Average training loss: 0.65
Training epoch took: 0:00:17
Running Validation...
Accuracy: 0.62
Validation took: 0:00:01
===== Epoch 5 / 10 =====
Training...
Average training loss: 0.52
Training epoch took: 0:00:17
Running Validation...
Accuracy: 0.60
Validation took: 0:00:01
===== Epoch 6 / 10 =====
Training...
Average training loss: 0.41
Training epoch took: 0:00:17
Running Validation...
Accuracy: 0.69
Validation took: 0:00:01
===== Epoch 7 / 10 =====
Training...
Average training loss: 0.34
Training epoch took: 0:00:17
Running Validation...
Accuracy: 0.69
Validation took: 0:00:01
===== Epoch 8 / 10 =====
Training...
Average training loss: 0.28
Training epoch took: 0:00:17
Running Validation...
Accuracy: 0.69
Validation took: 0:00:01
    
```

Gambar 5. Hasil Training dengan 10 Epoc

Pada uji coba BERT, peneliti melakukan Training dan evaluasi pada 10 epoch, dimana epoch ini berfungsi untuk menentukan beberapa kali model dapat melihat dataset secara keseluruhan. Rata-rata Waktu eksekusi pada tiap epoch yang dibutuhkan adalah 20 detik dan menghasilkan rata-rata akurasi data 0.60 pada tiap epoch. Ratarata Training loss pada epoch pertama cukup tinggi

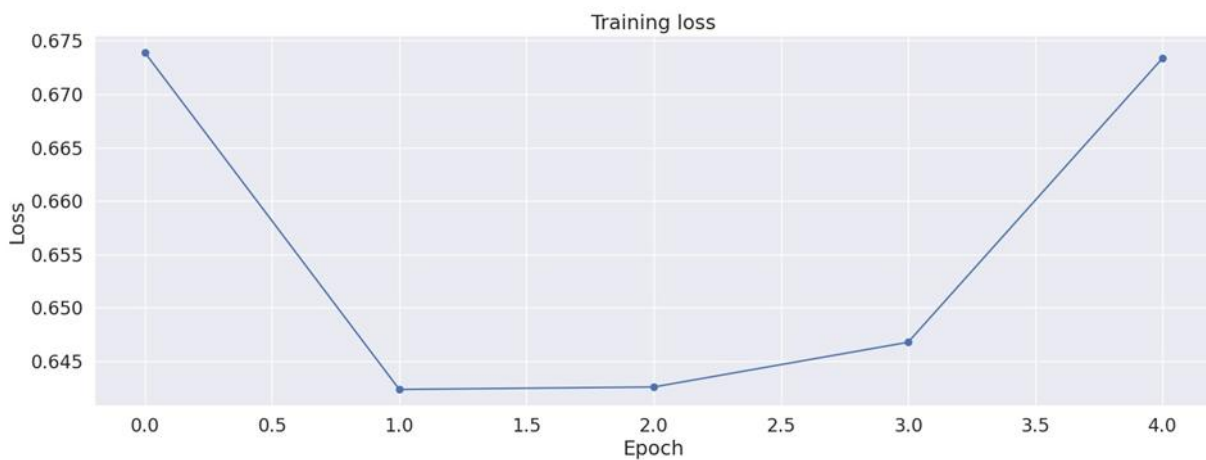
yaitu 0.95. tetapi pada epoch kelima nilai rata-rata Training loss turun menjadi 0.53, dan di epoch terakhir mendapat nilai Training loss sebesar 0.22 seperti pada gambar 6



Gambar 6. Grafik Training Loss dengan 10 Epoch

Pada pengujian lain, dengan model yang telah dibuat namun dengan perbedaan nilai epoch. Pada percobaan menggunakan nilai epoch 5 adalah Rata-rata Waktu eksekusi pada tiap epoch yang dibutuhkan adalah 20 detik dan menghasilkan rata-rata akurasi data 0.64 pada tiap epoch. Ratarata Training loss pada epoch pertama cukup tinggi yaitu 0.67. tetapi pada epoch ketiga nilai rata-rata Training loss turun menjadi

0.64, dan di epoch terakhir mendapat nilai Training loss naik sebesar 0.67 seperti pada gambar berikut.



Gambar 7. Grafik Training Loss dengan 5 Epoch

### 3.2. Evaluasi

Setelah dilakukan proses Training, maka selanjutnya dicoba model untuk mengetahui performa terhadap data baru. Hasil percobaan implementasi model terhadap data Testing dengan menggunakan Confusion Matrix dengan jumlah epoch yang berbeda bisa dilihat dari tabel berikut

Tabel 5. Perbandingan Nilai Akurasi Menurut Jumlah Epoch

Jumlah Epoch	MCC	Akurasi
5	0.292	0.700
10	0.572	0.800

Berdasarkan hasil implementasi model terhadap data Testing dengan menggunakan Confusion Matrix, model dengan epoch 10 menjadi model yang paling baik nilai akurasinya yaitu sebesar 0.80. Semakin tinggi nilai akurasi, semakin baik pula model yang dihasilkan. Selain nilai, didapat juga nilai Matthew Correlation Coefficient (mcc) sebesar 0.572 . Semakin nilai MCC mendekati +1 maka semakin baik kinerja algoritma klasifikasinya. Sebaliknya, jika nilai nya mendekati -1, maka semakin buruk kinerja algoritma klasifikasi.

```
[ ] from sklearn.metrics import matthews_corrcoef

flat_prediction = [item for sublist in prediction for item in sublist]
flat_prediction = np.argmax(flat_prediction, axis=1).flatten()

flat_true_labels = [item for sublist in true_labels for item in sublist]

mcc = matthews_corrcoef(flat_true_labels, flat_prediction)

print("MCC: %.3f" %mcc)

MCC: 0.572

▶ from sklearn.metrics import accuracy_score

acc = accuracy_score(flat_true_labels, flat_prediction)

print("ACC: %.3f" %acc)

ACC: 0.800
```

Gambar 9. Nilai akurasi dan MCC dengan epoch 10

Tabel 6. Visualisasi Sentimen

Sentiment	Positif	netral	negatif
GPT	49,54%	15,25%	35,21%
Gemini	38,76%	13,56%	41,68%
Llama	35.98 %	20,55%	43,47

#### 4. KESIMPULAN

Performansi model BERT dengan pembagian data 80% data Training, 10% data validation, dan 10% data Testing mendapatkan nilai akurasi sebesar 0.80 dengan jumlah epoch 10 mengungguli percobaan yang lain dengan jumlah epoch 5 dengan nilai akurasi 0.70, Hasil evaluasi menggunakan confusion matrix menunjukkan bahwa model BERT Cukup baik untuk melakukan klasifikasi sentiment Berdasarkan penelitian yang dilakukan oleh peneliti, dari data sentimen yang diberikan, model GPT menunjukkan sentimen positif tertinggi (49,54%) dibandingkan dengan Gemini (38,76%) dan Llama (35,98%). Sementara itu, model Llama memiliki sentimen negatif tertinggi (43,47%), diikuti oleh Gemini (41,68%) dan GPT (35,21%). Untuk sentimen netral, Llama memiliki persentase tertinggi (20,55%), sedangkan Gemini dan GPT memiliki persentase yang lebih rendah, masing-masing 13,56% dan 15,25%. Secara keseluruhan, GPT cenderung menghasilkan lebih banyak sentimen positif dibandingkan dengan model lainnya, sedangkan Llama dan Gemini lebih banyak menghasilkan sentimen negatif. Secara umum, dapat disimpulkan bahwa GPT menerima persepsi publik yang lebih positif dibandingkan Gemini dan LLaMA dalam konteks yang dianalisis dalam studi ini. Hal ini mengindikasikan bahwa pengguna atau pihak yang memberikan sentimen cenderung memiliki pengalaman atau pandangan yang lebih baik terhadap GPT .

### UCAPAN TERIMA KASIH

Penulis mengucapkan banyak terima kasih kepada semua pihak yang telah memberikan dukungan dan kontribusinya dalam penyelesaian penelitian ini, terutama untuk Dosen Pembimbing Penelitian Saya Bapak Adhita Erfina, S.T.,M.Kom yang telah memberikan saran dan masukan perihal penelitian ini, tidak lupa saya ucapkan terimakasih juga untuk teman teman saya yang selalu suport saya dalam segala hal apapun demi menyelesaikan Penelitian.

### DAFTAR PUSTAKA

- [1] A. Buscemi and D. Proverbio, "GPT vs Gemini vs LLaMA on Multilingual Sentiment Analysis," *J. Artif. Intell. Res.*, , 2024.
- [2] A. A. Rosa, "Perbandingan Performa Open Source LLM IndoBERT, OpenAI GPT, dan Gemini terhadap Sentiment Analysis dalam Bahasa Indonesia: Studi Kasus pada TokoTeknoPro PT VinTechnology," *J. Comput. Linguist.*, vol. 5, no. 2, pp. 123-140, 2024.
- [3] K. Kheiri and H. Karimi, "SentimentGPT: Exploiting GPT for Advanced Sentiment Analysis and its Departure from Current Machine Learning," *Int. J. Mach. Learn. AI*, 2023.
- [4] W. Ningsih USTI et al., "Analisis Sentimen Pengguna Aplikasi Gen Ai dari Appstore dan Googleplay Menggunakan Algoritma C45," *J. Data Sci. AI*, vol. 13, no. 5, 2024.
- [5] M. Y. Pratama et al., "Analisis Sentimen Chat GPT sebagai Masa Depan Pekerja pada Media Sosial YouTube menggunakan Algoritma Klasifikasi Naive Bayes," *J. Soc. Media AI Res.*, , 2023.
- [6] Ardiansyah et al., "Analisis Sentimen terhadap Pelayanan Kesehatan berdasarkan Ulasan Google Maps menggunakan BERT," *J. Health Inform. AI*, , 2023.
- [7] S. Alaparathi and M. Mishra, "Bidirectional Encoder Representations from Transformers (BERT): A Sentiment Analysis Odyssey," *J. Nat. Lang. Process.*, 2020.
- [8] P. A. Riyantoko et al., "Analisis Sentimen Sederhana Menggunakan Algoritma LSTM dan BERT untuk Klasifikasi Data Spam dan Non-Spam," *J. Data Min. AI*
- [9] N. A. R. Putri and Ardiansyah, "Analisis Sentimen Terhadap Kemajuan Kecerdasan Buatan di Indonesia Menggunakan BERT dan RoBERTa," *J. Sains dan Inform.*, vol. 9, no. 2, pp. 136-145, 2023, doi: 10.34128/jsi.v9i2.649.
- [10] F. T. Saputra, Y. Nurhadryani, S. H. Wijaya, and D. Defina, "Analisis Sentimen Bahasa Indonesia pada Twitter Menggunakan Struktur Tree Berbasis Leksikon," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 135, 2021, doi: 10.25126/jtiik.0814133.
- [11] A. Jaya, "Analisis Sentimen Pandangan Public Profesi PNS (Pegawai Negeri Sipil) dari Twiter menerapkan indonesian Roberta Base Sentiment Classifier," *Indones. J. Data Sci.*, vol. 4, no. 1, pp. 38-44, 2023, doi: 10.56705/ijodas.v4i1.66.
- [12] M. R. Fahlevvi, "Analisis Sentimen Terhadap Ulasan Aplikasi Pejabat Pengelola Informasi Dan Dokumentasi Kementerian Dalam Negeri Republik Indonesia Di Google Playstore Menggunakan Metode Support Vector Machine," *J. Teknol. dan Komun. Pemerintah.*, vol. 4, no.1, pp. 1-13, 2022, doi: 10.33701/jtkp.v4i1.2701.
- [13] M. H. Al-Areef and K. Saputra S, "Analisis Sentimen Pengguna Twitter Mengenai Calon Presiden Indonesia Tahun 2024 Menggunakan Algoritma LSTM," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 22, no. 2, p. 270, 2023, doi: 10.53513/jis.v22i2.8680.
- [14] I. Akbar et al., "Penerapan Long Short-Term Memory untuk Klasifikasi Multi-Label," vol. 7, no. 1, 2026.
- [15] A. Kaharudin, A. A. Supriyadi, and ..., "Analisis Sentimen pada Media Sosial dengan Teknik Kecerdasan Buatan Naïve Bayes: Kajian Literatur Review," *OKTAL J. Ilmu ...*, vol. 2, no. 6, pp. 1642-1649, 2023, [Online]. Available:

<https://journal.mediapublikasi.id/index.php/oktal/article/view/2944%0Ahttps://journal.mediapublikasi.id/index.php/oktal/article/download/2944/1371>