BAB I
PENDAHULUAN

1.1 Latar Belakang

Pesatnya pertumbuhan *e-commerce* di Indonesia dengan proyeksi 79,85 juta pengguna pada tahun 2026 (Statista, 2025) telah mendorong bertambahnya volume ulasan produk dalam skala besar, yang berperan krusial bagi penjual dalam membantu mengidentifikasi kelemahan produk serta memperbaiki kualitas produk agar lebih memuaskan pelanggan (Riski dan Darmawan, 2024), maupun bagi pembeli sebagai bahan pertimbangan sebelum melakukan transaksi (Chen et al., 2022). Namun, pemanfaatan informasi emosional ini tidaklah sederhana. Kompleksitas bahasa pengguna termasuk variasi informal, singkatan, dan gaya bertutur serta tingginya volume ulasan membuat analisis manual menjadi tidak efisien dan rentan inkonsistensi. Untuk mengatasi keterbatasan tersebut, berbagai pendekatan otomatis mulai dikembangkan dalam bidang pemrosesan bahasa alami.

Berbagai penelitian sebelumnya telah menunjukkan bahwa pendekatan berbasis *Bidirectional Encoder Representations from Transformers* (BERT) merupakan salah satu metode yang banyak digunakan dalam tugas klasifikasi emosi berbahasa Indonesia. Salah satu studi oleh Christian et al. (2025) menunjukkan bahwa model seperti IndoBERT mampu mencapai akurasi 80% pada domain ulasan *e-commerce* melalui pendekatan *fine-tuning*. Meskipun cukup efektif, temuan tersebut juga memperlihatkan bahwa pendekatan berbasis *fine-tuning* membutuhkan dataset berlabel besar serta sumber daya komputasi yang besar pada tahap pelatihan, sehingga dapat menjadi tantangan bagi organisasi dengan infrastruktur terbatas.

Sebagai alternatif, tren terbaru menunjukkan meningkatnya pemanfaatan *Generative Artificial Intelligence* (Generative *AI*) dalam berbagai tugas pemrosesan bahasa alami, khususnya karena kemampuannya melakukan *zero-shot* dan *few-shot learning* tanpa memerlukan proses pelatihan ulang yang kompleks. Merujuk pada penelitian oleh (Brown et al., 2020), pendekatan *few-shot* menunjukkan peningkatan performa yang nyata dibandingkan *zero-shot*, dengan lonjakan akurasi dari kisaran 20–25% menjadi 60–65%. Pendekatan tersebut mendorong eksplorasi lebih lanjut terhadap pemanfaatan Generative AI dalam berbagai skenario pemrosesan bahasa alami.

Selain itu, penelitian yang dilakukan oleh Nasution et al. (2025) menunjukkan bahwa model GPT-4 yang hanya melalui pendekatan *zero-shot learning* mencapai *F1 score macro average* sebesar 0,840, yang melampaui capaian model *state of the art* sebelumnya, yakni model IndoBERT untuk tugas klasifikasi sentimen pada domain media sosial. Penelitian tersebut juga menyoroti adanya perbedaan waktu inferensi yang cukup nyata antar model, sehingga efisiensi komputasi menjadi faktor penting dalam implementasi praktis. Meskipun menunjukkan potensi yang menjanjikan, capaian tersebut sangat dipengaruhi oleh domain, bahasa, serta skenario evaluasi yang digunakan, sehingga performa Generative AI belum tentu dapat digeneralisasi secara langsung ke khususnya domain ulasan produk *e-commerce* berbahasa Indonesia.

Pemanfaatan model Generative AI dalam memproses teks berbahasa Indonesia khususnya dalam domain ulasan *e-commerce* yang memiliki struktur bahasa khas belum

dievaluasi secara mendalam. Selain dilihat dari efektivitas model, kebutuhan industri juga menuntut model yang efisien dari sisi komputasi, termasuk kecepatan prediksi dan kemampuan menangani permintaan dalam jumlah besar. Pertimbangan tersebut penting untuk memastikan bahwa solusi yang dikembangkan tidak hanya efektif, tetapi juga layak diterapkan dalam sistem operasional berskala luas. Melihat adanya celah penelitian tersebut, diperlukan evaluasi komprehensif yang dapat membandingkan pendekatan generatif dan *fine-tuning* model berbasis BERT secara langsung dalam konteks ulasan *e-commerce* Indonesia.

Berdasarkan uraian di atas, penelitian ini bertujuan untuk menganalisis sejauh mana performa pendekatan *zero-shot* dan *few-shot learning* menggunakan model Generative AI, yaitu LLaMA 3.1:8B dan Gemma 3:12B melalui Ollama, serta GPT-5.1 dan Gemini 2.5 Flash melalui API, dibandingkan dengan pendekatan *fine-tuning* model berbasis BERT menggunakan model IndoBERT, DistilBERT, RoBERTa, dan mBERT dalam klasifikasi emosi ulasan produk *e-commerce* berbahasa Indonesia. Penelitian ini juga menilai pengaruh penambahan ukuran set contoh pada pendekatan *few-shot learning* terhadap performa model Generative AI, serta mengimplementasikan model terbaik ke dalam antarmuka berbasis web sebagai media penerapan sistem klasifikasi emosi. Hasil penelitian diharapkan dapat menjadi rekomendasi metode yang paling optimal untuk mendukung pengembangan solusi kecerdasan buatan di sektor *e-commerce*.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah dalam penelitian ini dirumuskan menjadi tiga pertanyaan utama sebagai berikut:

1. Bagaimana performa pendekatan *zero-shot learning* dan *few-shot learning* menggunakan Generative AI serta *fine-tuning* model berbasis BERT dalam klasifikasi emosi teks ulasan produk *e-commerce* berbahasa Indonesia, baik dari sisi efektivitas maupun efisiensi?
2. Bagaimana pengaruh penambahan ukuran set contoh pada pendekatan *few-shot learning* terhadap performa model Generative AI dalam hal efektivitas dan efisiensi?
3. Bagaimana cara mengimplementasikan model terbaik ke dalam sebuah *website* sebagai media penerapan sistem klasifikasi emosi teks ulasan produk *e-commerce* berbahasa Indonesia?

1.3 Batasan Masalah

Agar penelitian lebih terarah dan tidak melebar dari fokus yang ditetapkan, maka batasan masalah dalam penelitian ini ditentukan sebagai berikut:

1. Sumber data yang digunakan dalam penelitian ini terbatas pada PRDECT-ID Dataset.
2. Kategori emosi yang diklasifikasikan dibatasi pada lima label, yaitu *anger*, *fear*, *happy*, *love*, dan *sad*, sesuai dengan anotasi yang tersedia dalam dataset.
3. Model *Generative AI* yang dievaluasi dibatasi pada lima model, yaitu LLaMA 3.1:8B, Gemma 3:12B melalui Ollama, serta GPT-5.1, dan Gemini 2.5 Flash melalui API.

4. Model berbasis BERT yang digunakan untuk baseline *fine-tuning* dibatasi pada IndoBERT, DistilBERT, RoBERTa, dan mBERT.
5. Evaluasi performa penelitian ini mencakup dua aspek, yaitu efektivitas yang diukur menggunakan *precision*, *recall*, *F1 score*, dan *accuracy*, serta efisiensi yang dinilai melalui *latency* per prediksi dan *throughput* model per menit.

1.4 Tujuan Penelitian

Adapun tujuan penelitian ini yaitu:
1. Menganalisis sejauh mana performa teknik *zero-shot* dan *few-shot learning* pada model Generative AI dibandingkan dengan model berbasis BERT hasil *fine-tuning*, baik dari sisi efektivitas maupun efisiensi dalam klasifikasi emosi ulasan produk *e-commerce* berbahasa Indonesia.
2. Mengevaluasi pengaruh penambahan ukuran set contoh pada pendekatan *few-shot learning* terhadap performa model Generative AI dalam hal efektivitas dan efisiensi.
3. Mengimplementasikan model terbaik ke dalam sebuah *website* sebagai media penerapan sistem klasifikasi emosi teks ulasan produk *e-commerce* berbahasa Indonesia.

1.5 Manfaat Penelitian

Manfaat yang diharapkan pada penelitian ini yaitu:
1. Menyediakan pemahaman yang lebih jelas mengenai performa model *Generative AI* pada skenario *zero-shot* dan *few-shot learning* dibandingkan dengan pendekatan *fine-tuning* BERT.
2. Menjadi acuan bagi pengembang sistem AI dalam memilih metode klasifikasi emosi ulasan *e-commerce* Indonesia yang paling efisien dan efektif sesuai kebutuhan operasional.
3. Membantu pelaku industri *e-commerce* dan konsumen dalam memahami emosi pelanggan secara otomatis, cepat, dan dalam skala besar.

1.6 Landasan Teori

1.6.1 Klasifikasi emosi pada ulasan

Klasifikasi emosi pada ulasan merupakan bagian dari bidang *Natural Language Processing* (NLP) yang bertujuan untuk mengidentifikasi dan mengelompokkan emosi yang terkandung dalam teks secara otomatis. Berbeda dengan analisis sentimen yang umumnya hanya mengklasifikasikan opini ke dalam kategori positif, negatif, atau netral, klasifikasi emosi berfokus pada pengenalan emosi yang lebih spesifik seperti senang, marah, sedih, takut, dan emosi lainnya yang merepresentasikan kondisi psikologis penulis teks (Acheampong et al., 2020).

Dalam konteks ulasan produk *e-commerce*, emosi yang diekspresikan pengguna memiliki peran penting dalam menggambarkan pengalaman, kepuasan, maupun kekecewaan terhadap suatu produk atau layanan. Ulasan berbasis emosi dinilai mampu

memberikan informasi yang lebih mendalam dibandingkan sekadar polaritas sentimen, sehingga bermanfaat bagi pelaku bisnis dalam pengambilan keputusan dan peningkatan kualitas layanan (Zhang et al., 2018).

1.6.2 Generative Artificial Intelligence

Generative Artificial Intelligence (Generative AI) merupakan cabang kecerdasan buatan yang berfokus pada pengembangan model yang mampu menghasilkan konten baru, seperti teks, gambar, atau data lainnya, berdasarkan pola yang dipelajari dari data sebelumnya. Dalam konteks NLP, Generative AI umumnya direpresentasikan oleh model bahasa yang dirancang untuk memahami konteks dan menghasilkan teks yang koheren serta bermakna (Brown et al., 2020).

Perkembangan Generative AI mengalami kemajuan signifikan dengan diperkenalkannya arsitektur Transformer yang memungkinkan model mempelajari dependensi jangka panjang dalam teks secara lebih efektif. Generative AI dilatih menggunakan data dalam jumlah besar dan mampu melakukan berbagai tugas NLP, tanpa memerlukan pelatihan ulang secara spesifik untuk setiap tugas (Radford et al., 2019). Dalam implementasinya, penelitian ini mengevaluasi beberapa model Generative AI yang merepresentasikan pendekatan *open-source* dan berbasis API. Penjelasan lebih lanjut mengenai model yang digunakan disajikan pada bab metode penelitian.

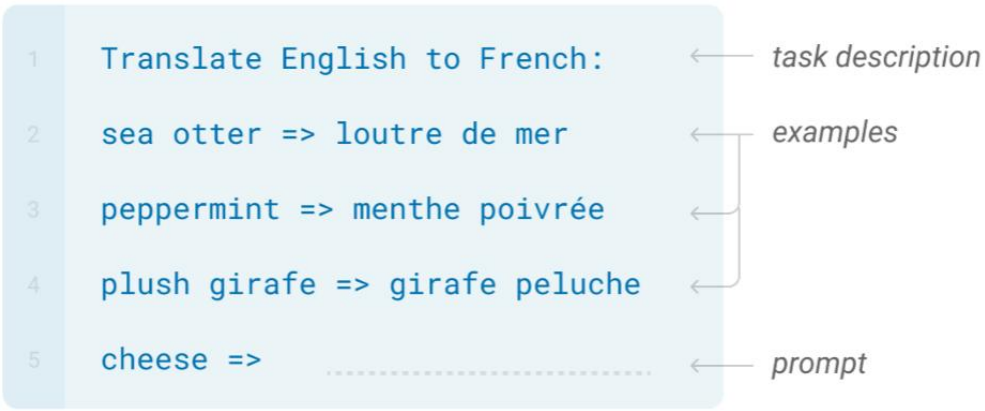
1.6.3 Zero-shot & few-shot learning

Zero-shot learning dan *few-shot learning* merupakan pendekatan pembelajaran dalam kecerdasan buatan yang bertujuan untuk mengatasi keterbatasan data berlabel dalam proses pelatihan model. Kedua pendekatan ini menjadi semakin relevan seiring berkembangnya *Large Language Models* (LLMs), khususnya dalam konteks NLP.

Zero-shot learning merujuk pada kemampuan model untuk menyelesaikan suatu tugas tanpa menggunakan contoh data berlabel secara eksplisit pada saat inferensi. Model hanya diberikan instruksi atau deskripsi tugas dalam bentuk *prompt*, tanpa disertai contoh input dan output. Pendekatan ini dimungkinkan karena model telah dilatih sebelumnya menggunakan data berskala besar dan beragam, sehingga mampu melakukan generalisasi terhadap tugas baru berdasarkan pemahaman semantik yang telah dipelajari (Wei et al., 2022). Seperti yang diilustrasikan pada gambar 1, dalam mekanisme ini model hanya diberikan deskripsi tugas dan input prompt tanpa adanya set contoh.

Gambar 1 Contoh *prompt zero-shot learning* (Brown et al., 2020)

Sebaliknya, *few-shot learning* memungkinkan model untuk meningkatkan performanya dengan memberikan sejumlah kecil contoh (*shot*) sebagai panduan sebelum melakukan prediksi. Contoh-contoh tersebut berfungsi sebagai konteks tambahan yang membantu model memahami pola tugas yang diinginkan. Penelitian menunjukkan bahwa penyertaan contoh dalam *prompt* dapat meningkatkan *accuracy* dan stabilitas prediksi model, khususnya pada tugas klasifikasi teks dan analisis sentimen atau emosi (Brown et al., 2020). Hal ini diperjelas pada gambar 2, yang menunjukkan bahwa model diberikan deskripsi tugas serta beberapa contoh (*shots*) sebagai konteks sebelum input *prompt* akhirnya diberikan.



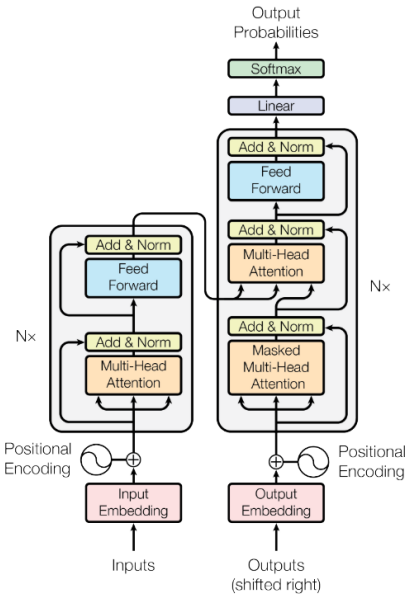
Gambar 2 Contoh *prompt few-shot learning* (Brown et al., 2020)

Dalam konteks model Generative AI berbasis Transformer, *zero-shot* dan *few-shot learning* tidak melibatkan proses pembaruan parameter model (*fine-tuning*), melainkan mengandalkan teknik *in-context learning*. Pada pendekatan ini, model belajar dari konteks *prompt* yang diberikan secara sementara selama proses inferensi, tanpa menyimpan pengetahuan baru secara permanen (Dong et al., 2024). Hal ini menjadikan *zero-shot* dan *few-shot learning* lebih efisien dari segi waktu dan sumber daya komputasi dibandingkan metode *fine-tuning* tradisional.

1.6.4 Bidirectional Encoder Representations from Transformers

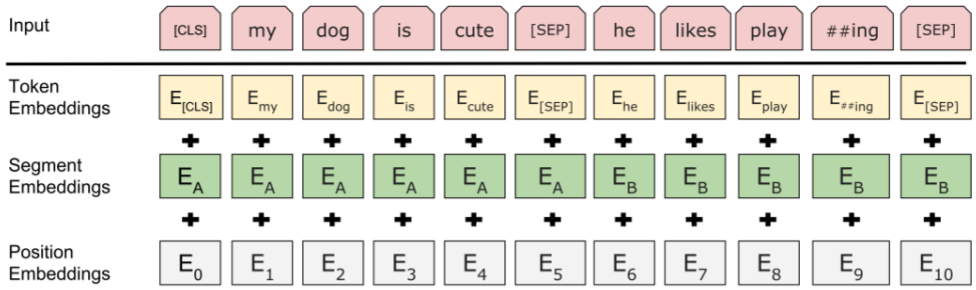
Bidirectional Encoder Representations from Transformers (BERT) merupakan model *pre-trained language* representation berbasis arsitektur Transformer *encoder* yang dikembangkan untuk memahami konteks bahasa secara *bidirectional*, yaitu dengan mempertimbangkan informasi dari sisi kiri dan kanan token secara simultan. Berbeda dengan model bahasa sebelumnya yang bersifat *unidirectional* atau hanya memanfaatkan konteks terbatas, BERT mampu menghasilkan representasi semantik yang lebih kaya dan kontekstual melalui mekanisme *self-attention* pada arsitektur Transformer. Pendekatan ini memungkinkan BERT untuk menangkap hubungan sintaksis dan semantik antar kata dalam satu kalimat maupun antar kalimat secara lebih efektif, sehingga sangat sesuai untuk berbagai tugas NLP (Devlin et al., 2019).

Arsitektur Transformer yang menjadi dasar BERT terdiri dari beberapa lapisan *encoder* yang masing-masing mengandung komponen *multi-head self-attention* dan *fully connected feed-forward network*. Mekanisme *self-attention* memungkinkan model untuk menentukan tingkat relevansi setiap token terhadap token lainnya dalam satu urutan teks tanpa bergantung pada pemrosesan sekuensial seperti pada RNN. Struktur ini juga didukung oleh *residual connection* dan *layer normalization* yang berfungsi untuk menjaga stabilitas pelatihan dan meningkatkan konvergensi model. Arsitektur Transformer dapat dilihat pada gambar 3, yang menggambarkan alur pemrosesan input melalui tumpukan *encoder* Transformer hingga menghasilkan representasi kontekstual yang kaya.



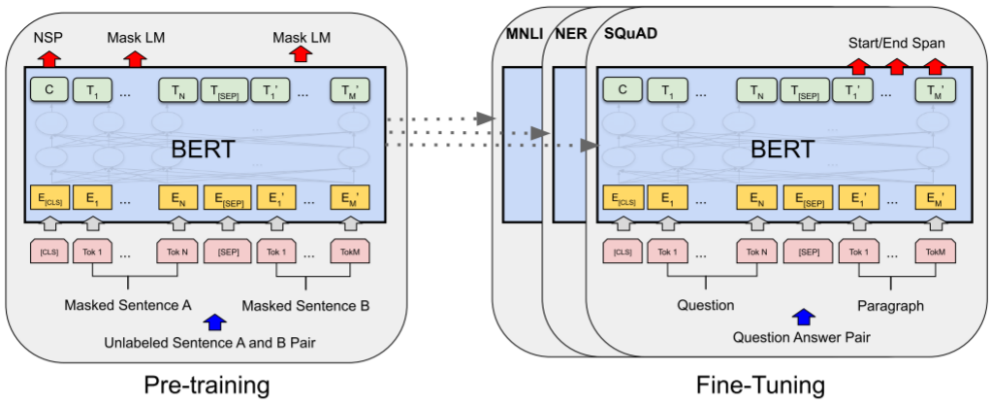
Gambar 3 Arsitektur Transformer (Vaswani et al., 2017)

Dalam proses pemrosesan input, BERT menggunakan representasi *embedding* yang merupakan hasil penjumlahan dari *token embedding*, *segment embedding*, dan *position embedding*. *Token embedding* merepresentasikan token hasil tokenisasi *WordPiece embeddings* yang menggunakan sebanyak 30.000 kosakata token (Wu et al., 2016), *Segment embedding* digunakan untuk membedakan segmen kalimat dalam satu input, sedangkan *position embedding* menyandikan informasi urutan token dalam kalimat. Selain itu, BERT menggunakan token khusus seperti ([CLS]) yang diletakkan di awal input sebagai representasi global untuk tugas klasifikasi, serta ([SEP]) sebagai pemisah antar segmen teks. Representasi input BERT dapat dilihat pada gambar 4, yang menunjukkan bagaimana ketiga jenis *embedding* tersebut digabungkan sebelum diproses oleh *encoder* Transformer.



Gambar 4 Representasi input BERT (Devlin et al., 2019)

BERT dilatih melalui dua tahap utama, yaitu *pre-training* dan *fine*-tuning. Ilustrasi perbedaan proses *pre-training* dan *fine-tuning* pada BERT merujuk pada skema yang diperkenalkan oleh (Devlin et al., 2019), yang menunjukkan bahwa pada tahap *pre-training* BERT dilatih menggunakan pasangan kalimat tanpa label yang diawali dengan token ([CLS]) dan dipisahkan oleh ([SEP]), dengan sebagian token disembunyikan secara acak untuk tugas *Masked Language Modeling* (MLM) serta pembelajaran hubungan antar kalimat melalui *Next Sentence Prediction* (NSP), sedangkan pada tahap *fine-tuning* struktur input tetap dipertahankan namun data yang digunakan telah berlabel dan disesuaikan dengan tugas spesifik, seperti *question answering* atau klasifikasi teks, yang umumnya direpresentasikan dalam bentuk *question–answer pair*, sehingga model *pre-trained* dapat diadaptasi secara efektif untuk menyelesaikan tugas *downstream* sesuai tujuan penelitian. Ilustrasi *pre-training* dan *fine-tuning* BERT dapat dilihat pada gambar 5, dengan penjelasan lebih lanjut mengenai *fine-tuning* model berbasis BERT yang digunakan disajikan pada bab metode penelitian.



Gambar 5 Tahapan *pre-training* dan *fine-tuning* BERT (Devlin et al., 2019)

1.6.5 Fine-tuning

Fine-tuning merupakan tahap lanjutan dalam pendekatan *transfer learning* pada model bahasa berbasis *deep learning*, yang menunjukkan bahwa model yang telah melalui proses *pre-training* pada *corpus* teks berskala besar disesuaikan kembali untuk menyelesaikan tugas spesifik menggunakan dataset berlabel. Pada tahap ini, parameter

model *pre-trained* tidak dibekukan, melainkan diperbarui secara menyeluruh atau sebagian selama proses pelatihan agar model dapat menyesuaikan representasi yang telah dipelajari dengan karakteristik data dan tujuan tugas *downstream*. Pendekatan *fine-tuning* memungkinkan model memanfaatkan pengetahuan linguistik umum yang diperoleh selama *pre-training*, seperti struktur sintaksis dan semantik bahasa, sehingga kebutuhan data berlabel dan waktu pelatihan dapat ditekan dibandingkan pelatihan model dari awal (Sun et al., 2019).

1.6.6 Pearson's correlation coefficient

Pearson's correlation coefficient merupakan metode statistik yang digunakan untuk mengukur kekuatan dan arah hubungan linier antara dua variabel numerik, dengan nilai berkisar antara -1 hingga +1, yang menunjukkan bahwa nilai mendekati +1 menunjukkan bahwa kedua variabel bergerak searah, 0 menunjukkan tidak ada hubungan linier yang nyata, dan nilai mendekati -1 menunjukkan bahwa kedua variabel bergerak berlawanan arah. *Pearson's correlation coefficient* sering digunakan dalam analisis data untuk mengukur hubungan linier antara fitur yang berbeda dalam suatu eksperimen atau dataset karena menggabungkan informasi kovarians dan varians dari kedua variabel. Secara matematis, *Pearson's correlation coefficient* merupakan ukuran linieritas yang normalisasi dari kovarians dua variabel dan memungkinkan perbandingan dalam konteks yang berbeda (Egghe & Leydesdorff, 2009). *Pearson's correlation coefficient*, dilambangkan dengan r, dapat dihitung dengan persamaan berikut:

$$r = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \quad (1)$$

Keterangan:
r: nilai *Pearson's correlation coefficient*
X_i: nilai variabel pertama ke-i
Y_i: nilai variabel kedua ke-i
$\bar{X}$: rata-rata variabel X
$\bar{Y}$: rata-rata variabel Y
n: jumlah pasangan data

Untuk menguji signifikansi statistik dari *Pearson's correlation coefficient*, *nilai t-statistic* dapat dihitung untuk menilai apakah korelasi yang diperoleh antara dua variabel benar-benar signifikan atau terjadi secara kebetulan. Persamaan perhitungan *t-statistic* adalah sebagai berikut:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad (2)$$

Keterangan:
t: nilai *t-statistic*
r: nilai *Pearson's correlation coefficient*
n: jumlah pasangan data

Nilai *t-statistic* ini kemudian dibandingkan dengan *t-table* pada derajat kebebasan $df = n - 2$, yang ditampilkan pada gambar 6, untuk menentukan signifikansi korelasi. Jika nilai *p-value* yang dihitung lebih kecil dari tingkat signifikansi yang ditentukan, maka korelasi dianggap signifikan secara statistik, menunjukkan bahwa hubungan linier antara kedua variabel bukan terjadi secara kebetulan. Dengan demikian, perhitungan *t-statistic* memungkinkan peneliti tidak hanya mengetahui kekuatan dan arah hubungan, tetapi juga memastikan validitas hubungan linier secara statistik.

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Gambar 6 Tabel distribusi *t*
Sumber: https://www.tdistributiontable.com/

BAB II
METODE PENELITIAN

2.1 Metode Penelitian Kuantitatif

Penelitian ini menggunakan pendekatan kuantitatif untuk membandingkan performa model *Generative AI* menggunakan *zero-shot* dan *few-shot learning* dengan *fine-tuning* model berbasis BERT dalam tugas klasifikasi emosi teks ulasan produk *e-commerce* Indonesia. Perbandingan performa ini diukur berdasarkan dua aspek utama, yaitu efektivitas dan efisiensi. Selain itu, penelitian ini juga menerapkan analisis korelasi untuk mengkaji hubungan antara peningkatan jumlah contoh pada pendekatan *few-shot learning* dengan perubahan performa model Generative AI.

2.1.1 Metrik efektivitas

Efektivitas model klasifikasi dievaluasi menggunakan sejumlah metrik yang diturunkan dari *confusion matrix*. *Confusion matrix* merupakan tabel dua dimensi yang membandingkan hasil prediksi model dengan label sebenarnya, sehingga memberikan informasi mengenai jumlah prediksi yang benar dan salah pada setiap kelas. Struktur ini membantu peneliti memahami pola kesalahan model dan menilai kualitas prediksi secara lebih komprehensif, baik secara keseluruhan maupun per kelas (Tharwat, 2018).

Untuk memberikan gambaran mengenai cara kerja *confusion matrix* pada kasus klasifikasi multi-kelas, berikut disajikan contoh ilustrasi *confusion matrix* ketika model melakukan prediksi terhadap lima kelas (A–E), khususnya saat dievaluasi pada performa model dalam mengenali kelas B.

Tabel 1 Contoh ilustrasi *multi-class consufion matrix* dalam memprediksi kelas B

Confusion Matrix		**Predicted Values**				
Actual Values	**Classes**	**A**	**B**	**C**	**D**	**E**
	A	TN	FP	TN	TN	TN
	B	FN	TP	FN	FN	FN
	C	TN	FP	TN	TN	TN
	D	TN	FP	TN	TN	TN
	E	TN	FP	TN	TN	TN

Keterangan:
True Positive (TP): Kasus yang diprediksi positif dan benar-benar positif.
True Negative (TN): Kasus yang diprediksi negatif dan benar-benar negatif.
False Positive (FP): Kasus yang diprediksi positif tetapi sebenarnya negatif.
False Negative (FN): Kasus yang diprediksi negatif tetapi sebenarnya positif.

Warna hijau digunakan untuk menandai sel-sel yang menunjukkan prediksi benar oleh model, yaitu pada bagian *True Positive* (TP) dan *True Negative* (TN) yang dapat dilihat pada tabel 1. Sebaliknya, warna merah digunakan untuk menandai prediksi yang salah, yaitu pada bagian *False Positive* (FP) dan *False Negative* (FN). Misalnya, sel *True*

Positive (TP) yang berwarna hijau menunjukkan bahwa data dengan kelas B berhasil diprediksi dengan tepat sebagai kelas B oleh model. Sementara itu, sel berwarna merah pada *False Negative* (FN) menunjukkan bahwa data yang sebenarnya berasal dari kelas B justru diprediksi sebagai kelas lain, sehingga model gagal mengenali label yang benar.

Pengukuran yang diperoleh dari confusion matrix mencakup empat metrik utama, yaitu *precision*, *recall*, *F1 score*, dan *accuracy*. Masing-masing metrik memiliki fungsi evaluasi yang berbeda serta rumus perhitungan yang spesifik, sebagai berikut:

1. *Precision* merupakan metrik yang mengukur tingkat ketepatan model dalam memberikan prediksi positif. *Precision* menunjukkan seberapa besar proporsi prediksi positif yang benar-benar sesuai dengan label sebenarnya. Precision yang tinggi mengindikasikan bahwa model jarang menghasilkan prediksi positif yang keliru. Rumus perhitungan *precision* dapat ditunjukkan pada persamaan berikut.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

2. *Recall* merupakan metrik yang menilai kemampuan model dalam menemukan seluruh sampel yang sebenarnya termasuk pada kelas positif. Nilai *recall* yang tinggi menunjukkan bahwa model mampu mengenali sebagian besar data positif. Rumus perhitungan *recall* dapat ditunjukkan pada persamaan berikut.

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

3. *F1 score* merupakan nilai harmoni antara *precision* dan *recall*, sehingga memberikan penilaian yang lebih seimbang apabila terdapat ketidakseimbangan jumlah data antar kelas. Rumus perhitungan *F1 score* dapat ditunjukkan pada persamaan berikut.

$$F1\ score = 2 \times \frac{(Precision \times Recall)}{(Precision+Recall)} \quad (5)$$

4. *Accuracy* merupakan metrik yang mengukur proporsi prediksi model yang benar dibandingkan dengan seluruh data yang dievaluasi. Rumus perhitungan *accuracy* dapat ditunjukkan pada persamaan berikut.

$$Accuracy = \frac{TP}{TP+FN} \quad (6)$$

Dalam penelitian ini digunakan *Macro F1 score* sebagai metrik utama dalam proses evaluasi model. *Macro F1 score* menghitung *F1 score* untuk setiap kelas secara terpisah, kemudian merata-ratakannya tanpa bobot sehingga setiap kelas memiliki kontribusi yang sama dalam penilaian (Sokolova dan Lapalme, 2009). Pendekatan ini memberikan gambaran performa model yang lebih merata pada seluruh kelas dan mampu merepresentasikan kualitas klasifikasi secara lebih menyeluruh pada skenario multi-kelas.

2.1.2 Metrik efisiensi

Selain efektivitas, penelitian ini juga mengevaluasi efisiensi model untuk menilai kelayakan penerapan masing-masing pendekatan pada skenario operasional berskala besar. Dua metrik yang digunakan adalah *latency* dan *throughput*, yang secara umum digunakan untuk mengukur kinerja sistem inferensi model bahasa (Fernandez et al., 2025), sebagai berikut:

1. *Latency* merupakan waktu yang dibutuhkan model untuk menghasilkan satu prediksi sejak menerima input hingga output selesai diproses. *Latency* digunakan untuk mengukur kecepatan respons model. *Latency* yang lebih kecil menunjukkan bahwa model dapat memberikan respons lebih cepat. Rumus perhitungan *latency* dapat ditunjukkan pada persamaan berikut.

$$Latency\ (s/sample) = \frac{\sum_{i=1}^{N} t_i}{N} \quad (7)$$

   Dengan:
   N = jumlah sampel yang diuji
   t_i = waktu inferensi untuk sampel ke-i (dalam detik)

2. *Throughput* merupakan jumlah prediksi yang dapat dihasilkan model dalam satu satuan waktu tertentu, dinyatakan dalam satuan prediksi per menit. Metrik ini menggambarkan kapasitas pemrosesan model ketika memproses sejumlah data secara berurutan dalam satu rangkaian eksekusi. *Throughput* yang lebih tinggi menunjukkan kemampuan model untuk menyelesaikan lebih banyak prediksi dalam waktu yang sama. Rumus perhitungan *throughput* dapat ditunjukkan pada persamaan berikut.

$$Throughput_{per\ minute} = \frac{N}{T_{seconds}} \times 60 \quad (8)$$

   Dengan:
   N = jumlah total sampel yang diproses
   $T_{seconds}$ = total waktu pemrosesan seluruh sampel (dalam detik)

2.1.3 Tempat penelitian

Penelitian ini dilaksanakan kurang lebih di Laboratorium Rekayasa Perangkat Lunak (RPL), Program Studi Sistem Informasi, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin.

2.1.4 Analisis korelasi

Dalam penelitian ini, analisis *Pearson's correlation coefficient* digunakan untuk mengukur hubungan antara jumlah penambahan ukuran set contoh pada pendekatan *few-shot learning* dengan metrik performa model Generative AI, yaitu *F1 score*, *accuracy*, *latency*, dan *throughput*. Tingkat signifikansi yang digunakan adalah $\alpha = 0{,}05$ dan nilai *p-value* dimanfaatkan untuk menentukan apakah hubungan antar variabel signifikan secara statistik. Analisis korelasi dilakukan menggunakan *library* SciPy di Python, sehingga

memungkinkan perhitungan koefisien korelasi (*r*) serta *p-value* secara otomatis. Dengan pendekatan ini, peneliti dapat mengevaluasi sejauh mana penambahan ukuran set contoh *few-shot learning* memengaruhi performa model.

2.2 Sumber Data

Penelitian ini menggunakan *Product Reviews Dataset for Emotions Classification Tasks - Indonesian* (PRDECT-ID), yaitu kumpulan data ulasan produk berbahasa Indonesia yang telah dilabeli dengan kategori emosi dan sentimen (Sutoyo et al., 2022). Dataset ini berisi 5.400 ulasan produk yang diambil dari dari salah satu platform *e-commerce* terbesar di Indonesia, yaitu Tokopedia. Ulasan tersebut berasal dari 29 kategori produk yang berbeda dan masing-masing ulasan telah diberikan satu label emosi, yaitu *love*, *happy*, *anger*, *fear*, dan *sad*.

Proses pelabelan emosi pada dataset PRDECT-ID mengacu pada *Shaver's emotions model*, yang mengelompokkan emosi manusia ke dalam kategori dasar yang bersifat universal dan mudah dibedakan secara linguistik (Shaver et al., 2001). Anotasi dilakukan secara manual oleh sekelompok anotator dengan menggunakan kriteria emosi yang disusun oleh seorang ahli psikologi, sehingga konsistensi dan kualitas label dapat terjaga. Ringkasan kriteria anotasi untuk masing-masing kategori emosi disajikan pada tabel 2.

Tabel 2 Kriteria anotasi emosi pada dataset PRDECT-ID (Sutoyo et al., 2022)

Emosi	Karakteristik Kalimat	Contoh Kalimat
Anger	- Mengandung kata kasar atau makian. - Menunjukkan kemarahan, kekesalan, atau kebencian. - Keluhan atau ketidaksukaan terhadap produk/layanan/pengiriman. - Menggunakan huruf kapital berlebihan atau tanda baca marah (“!!!”, “???”). - Ekspresi jengkel, sebel, atau benci.	barang jelek!!! tiga hari sudah pada lepas pinggirnya, barang mahal tapi kualitasnya jelek banget
Fear	- Kalimat peringatan atau rasa takut. - Kekhawatiran terhadap produk atau penjual. - Keraguan, ketidakpastian, atau menanyakan hal yang membuat cemas. - Ekspresi hati-hati, waspada, curiga, atau ragu terhadap keamanan.	buat agan agan yang mau beli disini, saya cuman bisa saran buat bikin video unboxing, terus hidupin langsung instalin cpu z

Emosi	Karakteristik Kalimat	Contoh Kalimat
Happy	- Mengandung pujian. - Kepuasan terhadap produk/penjual. - Ekspresi senang, gembira, bahagia, atau puas. - Bangga dan memberikan penilaian positif terhadap kualitas produk/penjual.	mantep adminnya selalu merhatiin pembeli. Respect, proses super cepat, sampai jg cepat, barang Sesuai, mantaaaap, thanks
Love	- Ekspresi perasaan sayang, suka, atau rasa cinta. - Sangat puas terhadap produk. - Mengandung ungkapan berlebihan untuk menyanjung. - Mengandung pujian kuat atau kekaguman mendalam. - Ekspresi bangga atau sangat menyukai produk/penjual.	produknyaaa bagus dan sukaaakkk banggettt!!!
Sad	- Menyatakan kekecewaan terhadap produk. - Penyesalan, sedih, atau tidak puas. - Ekspresi kecewa, menyesal, atau merasa dirugikan.	sangat kecewa, phone holder tidak lengkap penyambung nya tidak ada, packing cuman pake keresek hitam doang

2.3 Waktu dan Tempat Penelitian

2.3.1 Waktu penelitian

Penelitian ini dilaksanakan dalam kurun waktu sekitar dua bulan. Rincian jadwal pelaksanaan penelitian disajikan pada tabel 3 berikut.

Tabel 3 Waktu penelitian

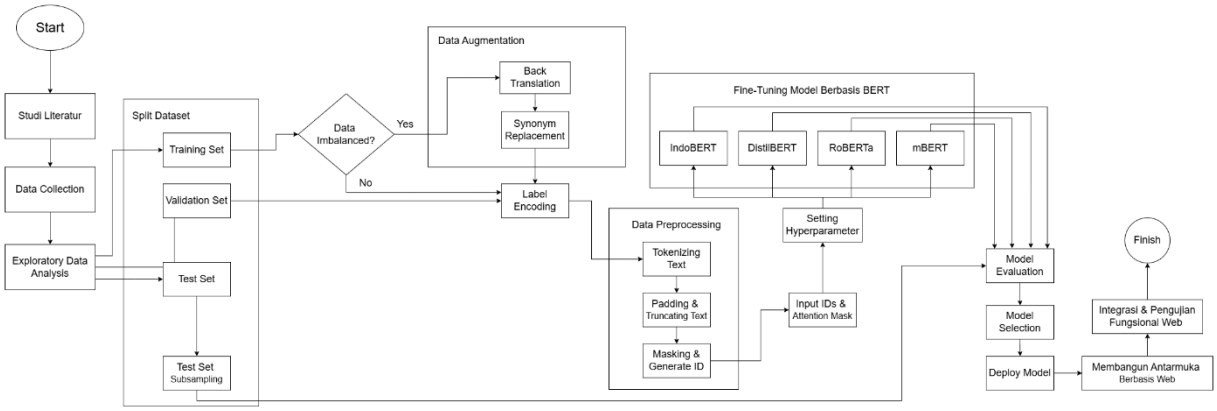
Kegiatan	Oktober				November			
	1	2	3	4	1	2	3	4
Studi Literatur								
Data Collection								
Exploratory Data Analysis								

Kegiatan	Oktober				November			
	1	2	3	4	1	2	3	4
Data Augmentation		■	■					
Split Dataset			■					
Data Preprocessing			■					
Fine-tuning Model Berbasis BERT				■				
Design Prompting					■			
Eksperimen Inferensi Model Generative AI					■			
Model Evaluation						■		
Deploy Model							■	
Membangun Antarmuka Berbasis Web								■
Integrasi & Pengujian Fungsional Web								■

2.4 Tahapan Penelitian

Pada penelitian ini, terdapat dua jalur eksperimen utama yang menerapkan pendekatan pemodelan yang berbeda, sebagaimana dapat dilihat lebih lanjut pada bagian ini.

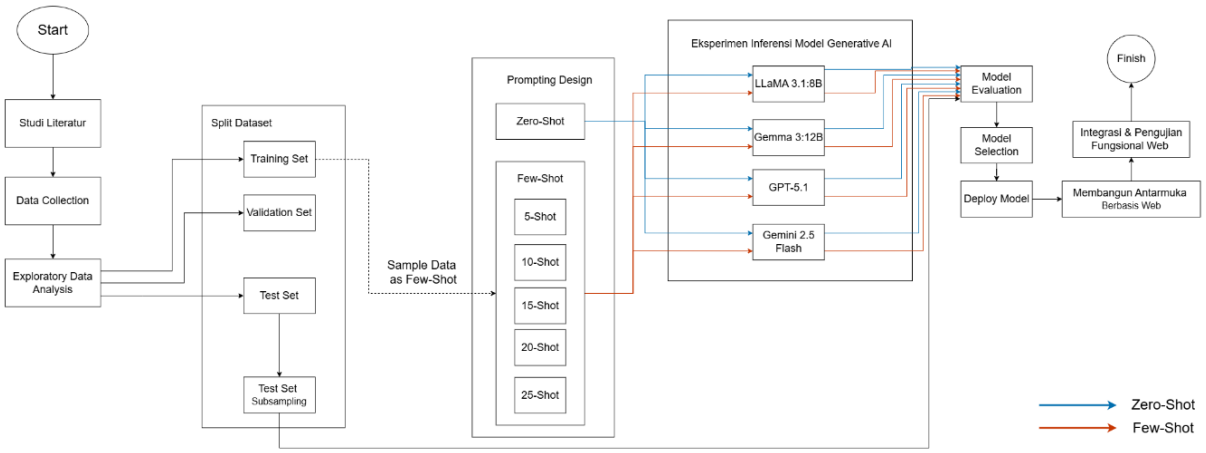
2.4.1 *Fine-tuning* model berbasis BERT



Gambar 7 Tahapan penelitian dengan pendekatan *fine-tuning* model berbasis BERT

Jalur *fine-tuning* model berbasis BERT berfokus pada proses pelatihan arsitektur *pre-trained* model menggunakan data yang telah dikonversi menjadi format input IDs dan *attention mask*. Dalam tahap ini, empat varian model utama yaitu IndoBERT, DistilBERT, RoBERTa, dan mBERT. Model dilatih menggunakan *training set* dan *validation set* dengan penyesuaian pada setting *hyperparameter* untuk mendapatkan performa yang optimal. Hasil pelatihan tersebut kemudian diteruskan ke tahap evaluasi model untuk diuji kinerjanya secara objektif menggunakan *test set* setelah penerapan teknik *subsampling*, yang selanjutnya dipilih model terbaik berdasarkan hasil eksperimen, baik pendekatan *fine-tuning* model berbasis BERT maupun dengan pendekatan *zero-shot* dan *few-shot* untuk dipublikasikan hingga akhirnya membangun antarmuka berbasis *website* sebagai media implementasi model.

2.4.2 *Zero-shot* dan *few-shot learning* model Generative AI



Gambar 8 Tahapan penelitian dengan pendekatan *zero-shot* dan *few-shot* learning model Generative AI

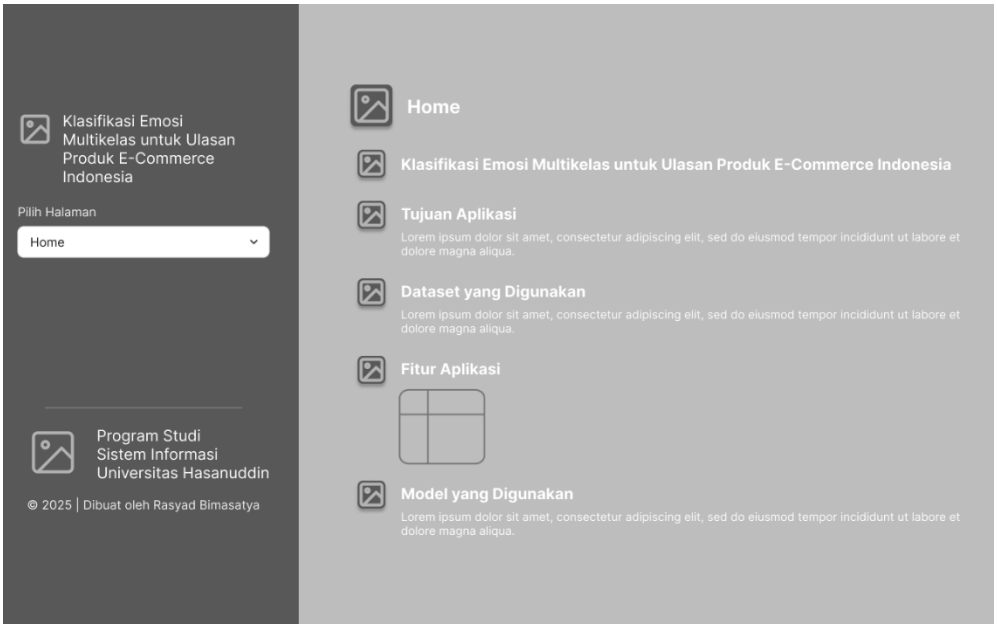
Jalur ini berfokus pada pengujian kemampuan model LLaMA 3.1:8B, Gemma 3:12B, GPT-5.1, dan Gemini 2.5 Flash dalam mengklasifikasikan emosi menggunakan teknik *zero-shot* dan *few-shot learning*. Proses ini melibatkan perancangan *prompting design* yang terstandarisasi, di mana skenario *few-shot* memanfaatkan sampel data dari *training set* dengan variasi pemberian contoh mulai dari 5 hingga 25 *shots*. Pengujian dilakukan dengan menggunakan *test set* setelah penerapan teknik subsampling, yang kemudian dilanjutkan dengan tahap evaluasi model untuk memilih model terbaik melalui perbandingan antara hasil *fine-tuning* berbasis BERT dan inferensi Generative AI yang kemudian diimplementasikan ke dalam antarmuka berbasis *website* sebagai media implementasi model.

2.5 Rancangan Aplikasi *Website*

Pengembangan aplikasi web pada penelitian ini bertujuan untuk mendukung analisis dan visualisasi hasil perbandingan performa pendekatan *zero-shot* dan *few-shot learning* pada model Generative AI dengan metode *fine-tuning* model berbasis BERT dalam klasifikasi emosi ulasan produk *e-commerce* berbahasa Indonesia. Aplikasi web ini juga menyediakan fitur prediksi emosi teks secara interaktif.

2.5.1 Halaman beranda

Dalam proses pengembangan *website*, diperlukan perancangan awal yang dikenal sebagai *wireframe*. Halaman beranda dirancang untuk menampilkan berbagai informasi umum yang menjelaskan tujuan dan fungsi *website* yang dikembangkan. Rancangan *wireframe* pada halaman beranda ditunjukkan pada gambar berikut.

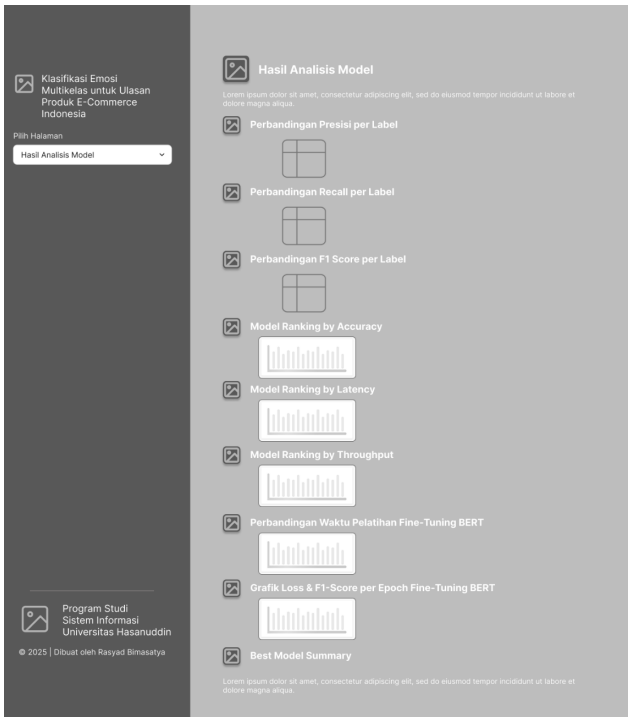


Gambar 9 *Wireframe* – halaman beranda

Halaman beranda dirancang sebagai tampilan utama aplikasi yang berfungsi sebagai pusat informasi awal bagi pengguna sebelum mengakses fitur-fitur lainnya. Tata letak halaman ini dibagi ke dalam dua area utama, yaitu panel navigasi di sisi kiri serta area konten utama yang menempati bagian tengah hingga kanan halaman. Panel navigasi memuat judul proyek, menu navigasi berbentuk *dropdown* dengan indikator halaman aktif pada beranda, serta informasi identitas instansi Program Studi Sistem Informasi Universitas Hasanuddin dan keterangan hak cipta pengembang yang ditempatkan pada bagian bawah. Sementara itu, area konten utama disusun secara vertikal yang diawali dengan judul halaman dan nama lengkap proyek klasifikasi emosi, kemudian dilanjutkan dengan empat bagian informasi utama, meliputi tujuan pengembangan aplikasi, deskripsi dataset yang digunakan, fitur-fitur aplikasi yang disajikan dalam bentuk kerangka tabel, serta model yang diterapkan dalam sistem. Setiap bagian informasi dari tujuan, dataset yang digunakan, dan model yang digunakan dilengkapi dengan paragraf penjelasan untuk memberikan pemahaman yang jelas kepada pengguna.

2.5.2 Halaman hasil analisis model

Halaman hasil analisis model dirancang untuk menampilkan hasil evaluasi dan perbandingan performa model yang digunakan dalam penelitian. Pada halaman ini disajikan berbagai informasi terkait metrik evaluasi, seperti metrik efektivitas dan efisiensi, dalam bentuk visualisasi yang mudah dipahami. Rancangan *wireframe* pada halaman hasil analisis model ditunjukkan pada gambar berikut.

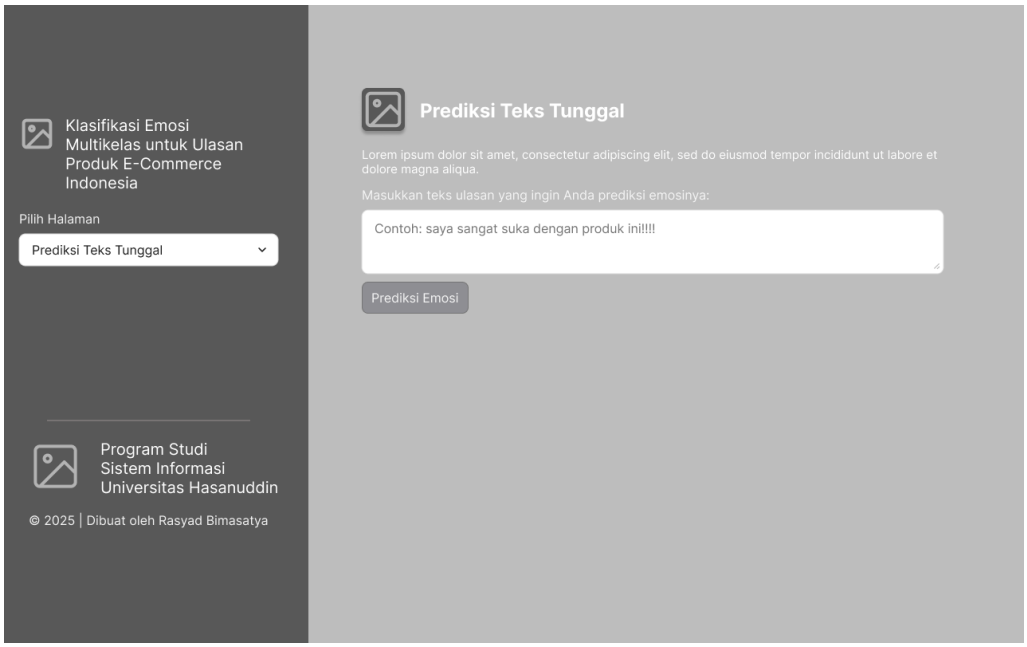


Gambar 10 *Wireframe* – halaman hasil analisis model

Halaman hasil analisis model dirancang untuk menyajikan informasi evaluasi performa model secara menyeluruh dan terstruktur. Struktur halaman ini tetap mempertahankan panel navigasi di sisi kiri yang konsisten dengan halaman sebelumnya, dengan indikator menu *dropdown* menunjukkan bahwa pengguna sedang berada pada halaman hasil analisis model. Area konten utama yang berada di bagian tengah hingga kanan halaman diawali dengan judul halaman serta paragraf deskripsi singkat sebagai pengantar. Selanjutnya, ditampilkan rangkaian visualisasi data yang disusun secara vertikal, meliputi tabel perbandingan metrik evaluasi berupa *precision*, *recall*, dan *F1 score* untuk setiap label emosi, serta grafik pemeringkatan model berdasarkan metrik *accuracy*, *latency*, dan *throughput*. Pada bagian akhir halaman disajikan analisis khusus untuk pendekatan *fine-tuning model* berbasis BERT, yang mencakup grafik waktu pelatihan serta grafik perkembangan *loss* dan *F1 score* pada setiap *epoch*, dan ditutup dengan ringkasan model terbaik.

2.5.3 Halaman prediksi teks tunggal

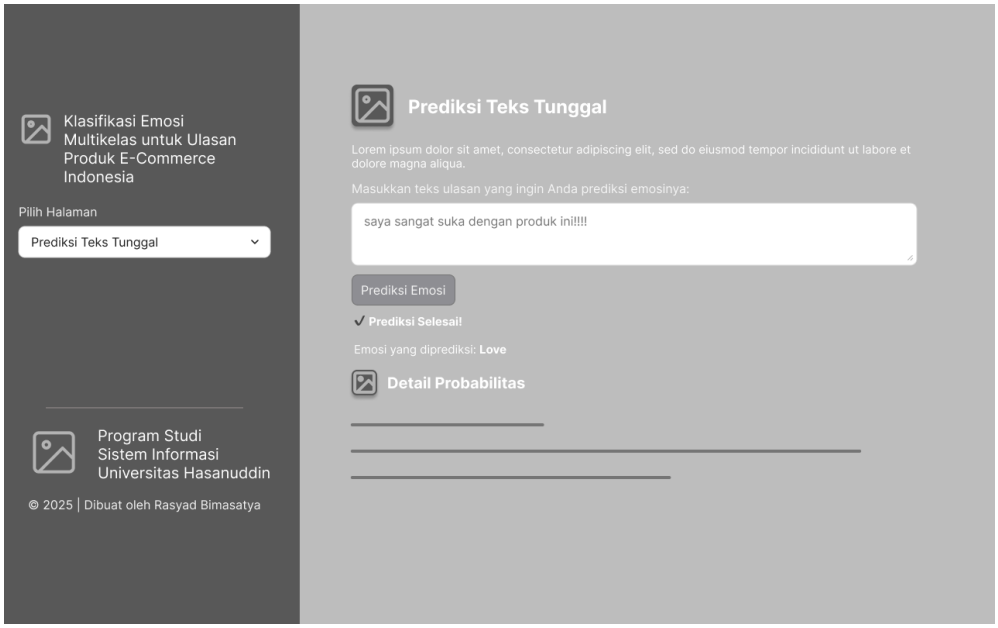
Halaman prediksi teks tunggal dirancang untuk menampilkan antarmuka awal yang memfasilitasi pengguna sebelum melakukan proses klasifikasi emosi terhadap satu teks ulasan produk *e-commerce*. Pada tampilan awal ini, disediakan area input teks yang memungkinkan pengguna memasukkan ulasan secara manual, disertai informasi singkat mengenai fungsi halaman dan cara penggunaan. Rancangan *wireframe* tampilan halaman sebelum proses prediksi teks tunggal ditunjukkan pada gambar berikut.



Gambar 11 *Wireframe* – halaman prediksi teks tunggal sebelum proses prediksi

Setelah proses prediksi dilakukan, halaman ini akan menampilkan hasil klasifikasi emosi yang dihasilkan oleh model beserta rincian probabilitas untuk setiap kelas emosi.

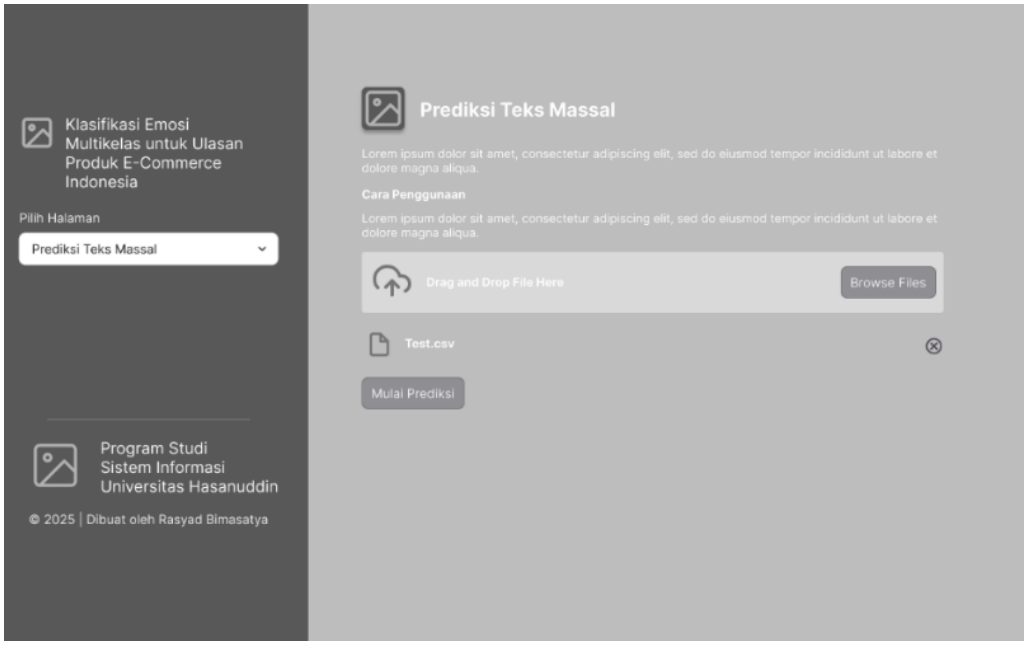
Informasi probabilitas tersebut disajikan dalam bentuk visualisasi yang memudahkan pengguna memahami tingkat keyakinan model terhadap hasil prediksi yang diberikan. Rancangan *wireframe* tampilan halaman setelah proses prediksi ditunjukkan pada gambar berikut.



Gambar 12 *Wireframe* – halaman prediksi teks tunggal setelah proses prediksi

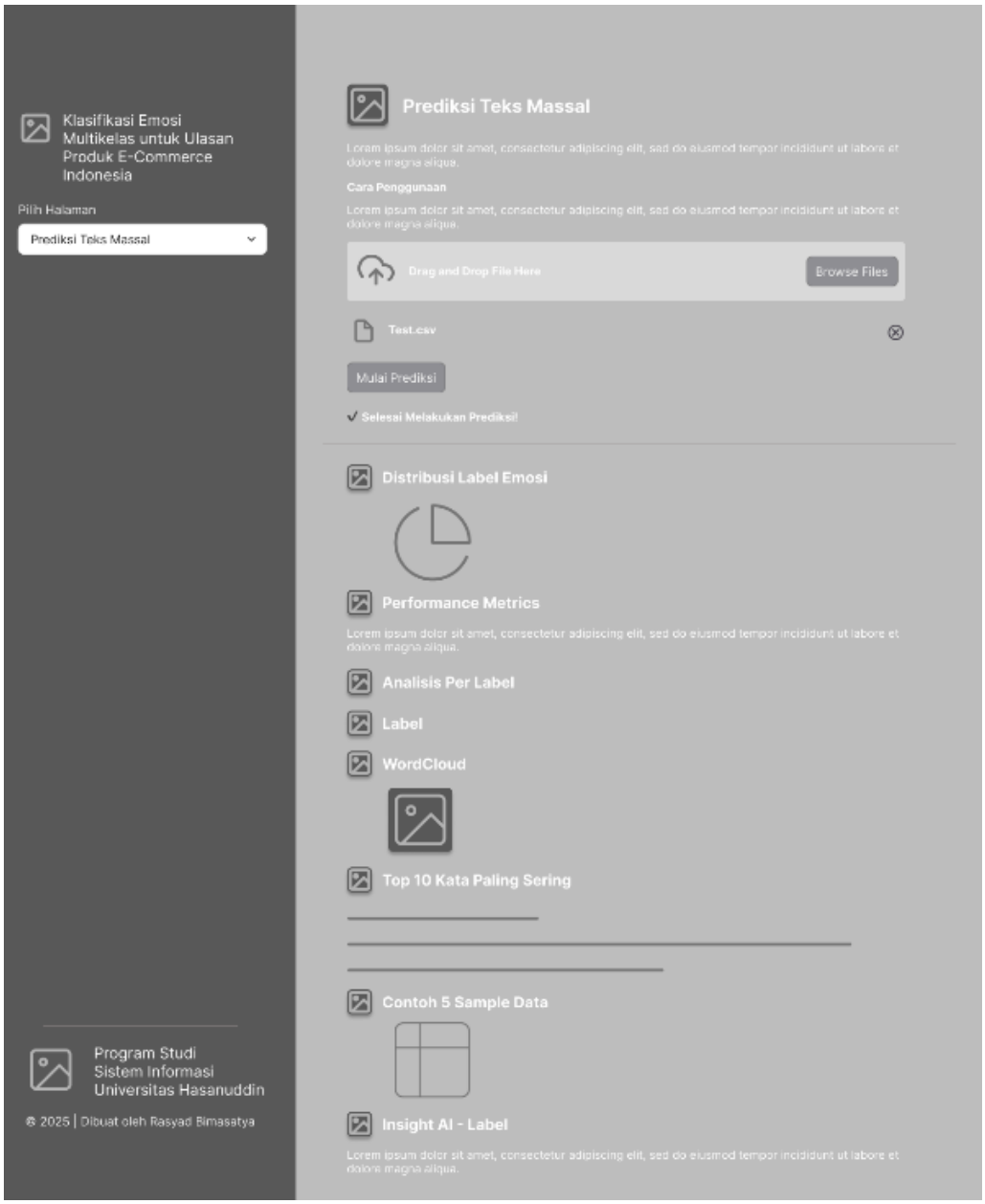
2.5.4 Halaman prediksi teks massal

Halaman prediksi teks massal dirancang untuk menampilkan antarmuka awal yang memfasilitasi pengguna sebelum melakukan proses klasifikasi emosi terhadap sekumpulan teks ulasan produk *e-commerce* secara sekaligus. Pada tampilan awal ini, disediakan area pengunggahan berkas yang memungkinkan pengguna mengunggah dataset ulasan dalam format *Comma Separated Value* (CSV) atau Excel, disertai informasi singkat mengenai cara penggunaan, tombol pemilihan berkas, serta tombol aksi untuk memulai proses prediksi. Rancangan *wireframe* pada halaman sebelum proses prediksi teks massal ditunjukkan pada gambar berikut.



Gambar 13 *Wireframe* – halaman prediksi teks massal sebelum proses prediksi

Setelah proses prediksi selesai dilakukan, halaman prediksi teks massal menampilkan hasil analisis klasifikasi emosi terhadap data yang telah diproses. Pada bagian atas konten utama, area pengunggahan berkas tetap ditampilkan dan dilengkapi dengan indikator status “Selesai Melakukan Prediksi!” di bawah tombol aksi, yang menandakan bahwa sistem telah berhasil memproses data. Selanjutnya, ditampilkan visualisasi hasil prediksi yang diawali dengan diagram *pie chart* pada bagian distribusi label emosi serta ringkasan metrik performa terkhusus untuk metrik *latency* dan *throughput*. Konten kemudian dilanjutkan dengan rincian analisis per label, visualisasi WordCloud, dan daftar top 10 kata paling sering yang muncul dalam ulasan. Pada bagian akhir halaman disajikan tabel pratinjau contoh 5 *sample* data serta bagian *insight* AI – label yang memberikan interpretasi naratif otomatis terhadap hasil klasifikasi emosi. Rancangan *wireframe* pada halaman setelah proses prediksi teks massal ditunjukkan pada gambar berikut.



Gambar 14 *Wireframe* – halaman prediksi teks massal setelah proses prediksi

2.6 Instrumen Penelitian

Dalam penelitian ini digunakan beberapa instrumen untuk mendukung kelancaran pelaksanaan penelitian. Instrumen tersebut dikelompokkan ke dalam dua kategori, yaitu perangkat keras (*hardware*) dan perangkat lunak (*software*).

2.6.1 Perangkat keras (*hardware*)

Penelitian ini memanfaatkan sejumlah perangkat keras yang berfungsi untuk mendukung pelaksanaan penelitian. Adapun perangkat keras yang digunakan adalah sebagai berikut dalam tabel 4.

Tabel 4 Perangkat keras

Perangkat Keras	Spesifikasi
Operation System (OS)	Windows 11
Processor	Intel(R) Core(TM) i7-8850H CPU @ 2.60GHz (12 CPUs), ~ 2.6GHz
RAM	16 GB
GPU	NVIDIA Quadro P1000

2.6.2 Perangkat lunak (*software*)

Penelitian ini juga memanfaatkan sejumlah perangkat lunak yang berfungsi untuk memudahkan dan mendukung proses penelitian. Adapun perangkat lunak yang digunakan adalah sebagai berikut dalam tabel 5.

Tabel 5 Perangkat lunak

Perangkat Lunak	Fungsi
Google Colaboratory	Digunakan sebagai lingkungan komputasi berbasis cloud untuk melakukan *fine-tuning* model berbasis BERT dengan dukungan GPU.
Visual Studio Code	Digunakan sebagai lingkungan pengembangan untuk melakukan *model evaluation* serta eksperimen inferensi model Generative AI berbasis *zero-shot* dan *few-shot learning*.
Ollama	Digunakan untuk menjalankan dan menguji inferensi model Generative AI secara lokal dalam skenario *zero-shot* dan *few-shot learning*.
Streamlit	Digunakan untuk membangun antarmuka berbasis web guna menampilkan hasil analisis model dan melakukan prediksi emosi teks.
Hugging Face	Digunakan sebagai penyedia model *pre-trained*, tokenizer, dan *library* pendukung untuk *fine-tuning* dan evaluasi model berbasis Transformer.
Python	Digunakan sebagai bahasa pemrograman utama dalam seluruh tahapan penelitian.
Google Gen AI SDK	Digunakan untuk mengakses dan memanfaatkan model Gemini 2.5 Flash melalui API.

Perangkat Lunak	Fungsi
OpenAI	Digunakan untuk mengakses model GPT melalui API untuk eksperimen klasifikasi emosi
PyTorch	Digunakan sebagai framework *deep learning* untuk mengimplementasikan dan melakukan *fine-tuning* model berbasis BERT.
TensorFlow	Digunakan sebagai framework pendukung untuk eksperimen dan pengolahan model.
Googletrans	Digunakan untuk melakukan *back translation* sebagai teknik data augmentation pada teks ulasan berbahasa Indonesia.
Pandas	Digunakan untuk pengolahan, manipulasi, dan manajemen dataset ulasan teks selama proses penelitian.
Scikit-learn	Digunakan untuk proses *label encoding*, pembagian dataset, serta perhitungan metrik evaluasi model.
Transformers	Digunakan untuk mengakses arsitektur Transformer, *tokenizer*, dan *pipeline fine-tuning* model berbasis BERT.
SciPy	Digunakan untuk melakukan perhitungan statistik, khususnya analisis korelasi Pearson dalam evaluasi hubungan antar variabel hasil eksperimen.
NumPy	Digunakan untuk operasi numerik dan pengolahan data berbasis array selama proses komputasi.
Matplotlib	Digunakan untuk visualisasi hasil evaluasi model dan analisis performa dalam bentuk grafik.
WordCloud	Digunakan untuk memvisualisasikan distribusi kata yang dominan pada teks ulasan berdasarkan kelas emosi.