

**PENERAPAN BERT UNTUK DETEKSI ULASAN PALSU
DI TIKTOK SHOP BY TOKOPEDIA PADA KATEGORI *FASHION***

LAPORAN TUGAS AKHIR

Laporan ini Disusun untuk Memenuhi Salah Satu Syarat Memperoleh Gelar Sarjana Strata 1 (S1) pada Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang



Disusun Oleh :

NAJMA NORA AGUSTINA

NIM 32602100098

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM SULTAN AGUNG
SEMARANG**

2025

***APPLICATION OF BERT FOR DETECTION OF FAKE REVIEWS
ON TIKTOK SHOP BY TOKOPEDIA IN THE FASHION CATEGORY***

FINAL PROJECT

*Proposed to complete the requirement to obtain a bachelor's degree (SI) at
Informatics Engineering Departement of Industrial Technology Faculty
Sultan Agung Islamic University*



NAJMA NORA AGUSTINA

NIM 32602100098

***MAJORING OF INFORMATICS ENGINEERING
INDUSTRIAL TECHNOLOGY FACULTY
SULTAN AGUNG ISLAMIC UNIVERSITY
SEMARANG***

2025

**LEMBAR PENGESAHAN
TUGAS AKHIR**

**PENERAPAN BERT UNTUK DETEKSI ULASAN PALSU DI
TIKTOK SHOP BY TOKOPEDIA PADA KATEGORI *FASHION***

**NAJMA NORA AGUSTINA
32602100098**

Telah dipertahankan di depan tim penguji ujian sarjana tugas akhir
Program Studi Teknik Informatika
Universitas Islam Sultan Agung
Pada tanggal : *16 Desember 2025*

TIM PENGUJI UJIAN SARJANA:

Badie'ah, ST., M.Kom.

NIK. 210615044

(Penguji 1)

Andi Riansyah, ST., M.Kom.

NIK. 210616053

(Penguji 2)

Sam Farisa Chaerul Haviana,

ST., M.Kom.

NIK. 210615046

(Pembimbing)

Semarang, *07-01-2026*

Mengetahui,
Kaprodik Teknik Informatika
Universitas Islam Sultan Agung

Moch. Fauzi, ST., MIT

NIK. 210604034

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Najma Nora Agustina
NIM : 32602100098
Judul Tugas Akhir : **PENERAPAN BERT UNTUK DETEKSI
ULASAN PALSU DI TIKTOK SHOP BY
TOKOPEDIA PADA KATEGORI *FASHION***

Dengan bahwa ini saya menyatakan bahwa judul dan isi Tugas Akhir yang saya buat dalam rangka menyelesaikan Pendidikan Strata Satu (S1) Teknik Informatika tersebut adalah asli dan belum pernah diangkat, ditulis ataupun dipublikasikan oleh siapapun baik keseluruhan maupun sebagian, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka, dan apabila di kemudian hari ternyata terbukti bahwa judul Tugas Akhir tersebut pernah diangkat, ditulis ataupun dipublikasikan, maka saya bersedia dikenakan sanksi akademis. Demikian surat pernyataan ini saya buat dengan sadar dan penuh tanggung jawab.

Semarang, 15 Januari 2020

Yang Menyatakan,



Najma Nora Agustina

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Najma Nora Agustina
NIM : 32602100098
Program Studi : Teknik Informatika
Fakultas : Teknologi Industri
Alamat Asal : Kabupaten Pekalongan, Jawa Tengah

Dengan ini menyatakan Karya Ilmiah berupa Tugas Akhir dengan Judul :
**PENERAPAN BERT UNTUK DETEKSI ULASAN PALSU DI TIKTOK
SHOP BY TOKOPEDIA PADA KATEGORI FASHION**

Menyetujui menjadi hak milik Universitas Islam Sultan Agung serta memberikan Hak bebas Royalti Non-Eksklusif untuk disimpan, dialihmediakan, dikelola dan pangkalan data dan dipublikasikan diinternet dan media lain untuk kepentingan akademis selama tetap menyantumkan nama penulis sebagai pemilik hak cipta. Pernyataan ini saya buat dengan sungguh-sungguh. Apabila dikemudian hari terbukti ada pelanggaran Hak Cipta/Plagiarisme dalam karya ilmiah ini, maka segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa melibatkan Universitas Islam Sultan agung.

Semarang, 15 Januari 2026

Yang menyatakan,



Najma Nora Agustina

KATA PENGANTAR

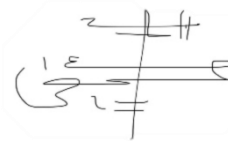
Puji syukur atas kehadiran Allah SWT yang telah memberikan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul ‘Penerapan *Bidirectional Encoder Representations From Transformers* Untuk Deteksi Ulasan Palsu di Tiktok Shop by Tokopedia Pada Kategori *Fashion*’. Tugas akhir ini disusun untuk memenuhi salah satu syarat menyelesaikan studi serta dalam rangka memperoleh gelar sarjana (S-1) pada Program Studi Teknik Informatika, Fakultas Teknologi Industri, Universitas Islam Sultan Agung Semarang.

Tugas Akhir ini tidak lepas dari bantuan, dukungan, dan doa dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. H. Gunarto, S.H., M.H., selaku Rektor Universitas Islam Sultan Agung Semarang.
2. Ibu Dr. Novi Marlyana, S.T., M.T., selaku Dekan Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang.
3. Bapak Moch Taufik, S.T., M.Kom, selaku Ketua Program Studi Teknik Informatika.
4. Bapak Sam Farisa Chaerul Haviana, S.T., M.Kom., selaku Dosen Pembimbing atas bimbingan, ilmu, dan saran selama penyusunan laporan.
5. Kedua orangtua saya, almarhum Bapak M. Fatkhunajib dan almarhumah Ibu Istiqomah. Doa, kasih sayang, dan pengorbanannya selama hidup menjadi sumber semangat dan motivasi terbesar bagi saya.
6. Suami tercinta, M. Saefudin, atas dukungan serta kesabarannya selama proses penyelesaian studi.

Penulis menyadari masih terdapat banyak kekurangan dalam laporan ini, sehingga saran dan kritik yang membangun sangat diharapkan.

Semarang, 15 Januari 2026



Najma Nora Agustina

DAFTAR ISI

COVER	i
LEMBAR PENGESAHAN.....	Error! Bookmark not defined.
SURAT PERNYATAAN KEASLIAN TUGAS AKHIR	Error! Bookmark not defined.
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAHError! Bookmark not defined.
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR.....	x
DAFTAR TABEL.....	xi
DAFTAR LAMPIRAN	xii
ABSTRAK.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah.....	3
1.3 Pembatasan Masalah.....	3
1.4 Tujuan.....	4
1.5 Manfaat.....	5
1.6 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI	7
2.1 Tinjauan Pustaka.....	7
2.2 Dasar Teori	10
2.2.1 Ulasan asli (<i>genuine review</i>)	10
2.2.2 Ulasan palsu (<i>fake review</i>)	11
2.2.3 Tiktok Shop by Tokopedia sebagai platform <i>e-commerce</i>	14
2.2.4 Produk <i>fashion</i> dalam <i>e-commerce</i>	15
2.2.5 NLP (<i>Natural Language Processing</i>)	16
2.2.6 LLM (<i>Large Language Models</i>)	17
2.2.7 BERT (<i>Bidirectional Encoder Representations from</i>	

<i>Transformers</i>).....	20
2.2.8 IAA (<i>Inter-Annotator Agreement</i>).....	22
2.2.9 Evaluasi <i>confusion matrix</i>	24
BAB III METODE PENELITIAN	28
3.1 Metode Penelitian	28
3.1.1 Deskripsi sistem	30
3.1.2 Studi literatur	32
3.1.3 Pengumpulan data	32
3.1.4 <i>Data preprocessing</i>	36
3.1.5 Implementasi model	38
3.1.6 Evaluasi	46
3.2 Analisa Kebutuhan	48
3.3 Penggunaan Sistem.....	53
3.4 Perancangan <i>User Interface</i>	55
3.4.1 Halaman awal sistem.....	55
3.4.2 Tampilan hasil sistem.....	56
BAB IV HASIL DAN ANALISIS PENELITIAN	57
4.1 Persiapan Dataset.....	57
4.2 Pelabelan Data	58
4.2.1 Dataset berlabel	58
4.2.2 Distribusi label	61
4.2.3 Nilai kesepakatan <i>annotator</i>	61
4.3 <i>Text Processing</i>	62
4.4 Pembagian Dataset	64
4.5 Tokenisasi BERT	66
4.6 Konfigurasi Model BERT	67
4.7 <i>Hyparameter</i> Model	68
4.8 Performa <i>Training</i> Model.....	70
4.9 Evaluasi Model	75
4.10 Implementasi <i>User Interface</i>	78
4.10.1 Integrasi model menggunakan <i>streamlit</i>	78

4.10.2	Cara akses aplikasi <i>streamlit</i>	78
4.10.3	Tampilan antarmuka sistem	79
BAB V KESIMPULAN DAN SARAN		87
5.1	Kesimpulan	87
5.2	Saran	88
DAFTAR PUSTAKA		89
LAMPIRAN		93

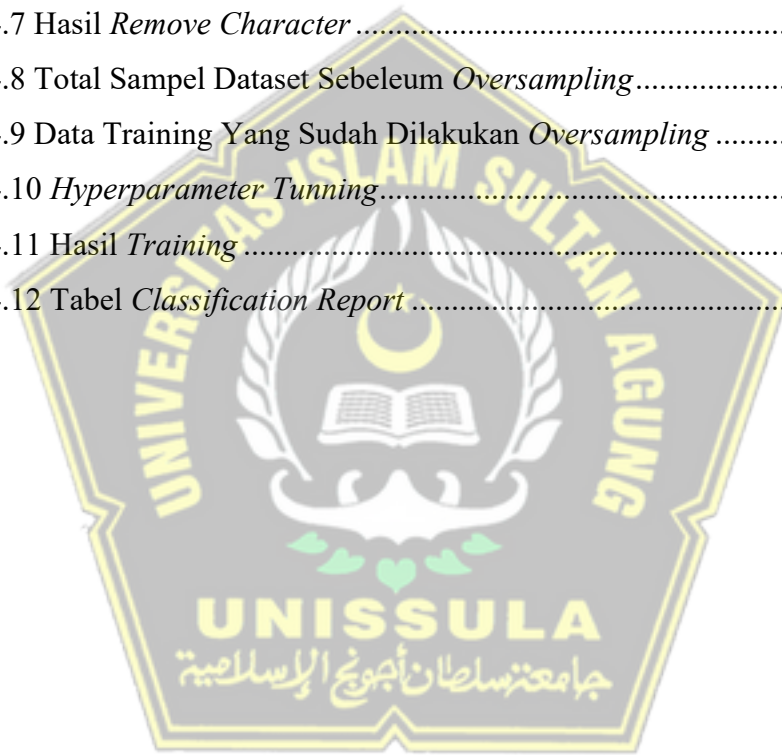


DAFTAR GAMBAR

Gambar 2.1 Arsitektur Model <i>Transformer</i>	18
Gambar 2.2 Model Arsitektur BERT	21
Gambar 3.1 Alur Metode Penelitian	28
Gambar 3.2 <i>Flowchart</i> Kerja Model.....	30
Gambar 3.3 Alur Pengumpulan Data	32
Gambar 3.4 <i>Flowchart</i> Pelabelan Data	34
Gambar 3.4 Alur Model BERT Untuk Klasifikasi Ulasan Palsu.....	41
Gambar 3.5 <i>Flowchart</i> Sistem	53
Gambar 3.6 Tampilan Halaman Awal Sistem.....	55
Gambar 3.7 Tampilan Halaman Hasil Sistem.....	56
Gambar 4.1 Kurva <i>Train Loss</i> dan <i>Validation Loss</i>	71
Gambar 4.2 Kurva <i>Train Accuracy</i> dan <i>Validation Accuracy</i>	72
Gambar 4.3 <i>Heatmap Confusion Matrix</i>	76
Gambar 4.4 Halaman Awal Sistem.....	80
Gambar 4.5 Hasil Prediksi Asli.....	81
Gambar 4.6 Hasil Karakteristik Ulasan Asli.....	82
Gambar 4.7 Hasil Prediksi Palsu.....	83
Gambar 4.8 Hasil Karakteristik Ulasan Palsu.....	84

DAFTAR TABEL

Tabel 4.1 Hasil Pengumpulan Dataset Ulasan	57
Tabel 4.2 Hasil Pelabelan Data Oleh <i>Annotator</i>	58
Tabel 4.3 Hasil Total Distribusi Kelas Seluruh Ulasan	61
Tabel 4.4 Nilai Kesepakatan Antar- <i>annotator</i>	61
Tabel 4.5 Hasil Remove Emoji	63
Tabel 4.6 Hasil <i>Lowercase</i>	63
Tabel 4.7 Hasil <i>Remove Character</i>	63
Tabel 4.8 Total Sampel Dataset Sebelem <i>Oversampling</i>	65
Tabel 4.9 Data Training Yang Sudah Dilakukan <i>Oversampling</i>	65
Tabel 4.10 <i>Hyperparameter Tuning</i>	68
Tabel 4.11 Hasil <i>Training</i>	70
Tabel 4.12 Tabel <i>Classification Report</i>	75



DAFTAR LAMPIRAN

Lampiran 1 Tabel Informasi Identitas <i>Annotator</i>	93
---	----



ABSTRAK

Perkembangan *e-commerce*, khususnya Tiktok Shop by Tokopedia, tidak hanya menghadirkan kemudahan berbelanja, tetapi juga tantangan baru berupa munculnya ulasan palsu (*fake review*) yang dapat menyesatkan konsumen, menurunkan kepercayaan, dan merugikan penjual maupun pembeli. Penelitian ini bertujuan untuk membangun sistem deteksi ulasan palsu pada produk *fashion* di Tiktok Shop by Tokopedia dengan memanfaatkan BERT (*Bidirectional Encoder Representations from Transformers*). Dataset ulasan diperoleh melalui proses *scraping* dari Tiktok Shop by Tokopedia, kemudian diproses dan dilatih menggunakan model BERT dengan penambahan lapisan klasifikasi. Hasil penelitian ini menunjukkan bahwa performa terbaik pada tahap pelatihan dicapai pada *epoch* ke-4 dengan akurasi pelatihan 94,12%, sedangkan evaluasi dengan metrik klasifikasi menunjukkan performa optimal, yaitu *precision* 0,96, *recall* 0,97, *F1-score* 0,96, dan akurasi 0,94. Hal ini membuktikan bahwa BERT mampu mendeteksi ulasan palsu secara efektif, meskipun masih terdapat tantangan berupa *overfitting* pada jumlah *epoch* yang lebih tinggi. Penelitian diharapkan dapat memberikan kontribusi pada pengembangan sistem deteksi *fake review* di platform Tiktok Shop by Tokopedia.

Kata Kunci: BERT, LLM, *E-commerce*, Ulasan Palsu

ABSTRACT

The development of e-commerce, particularly Tiktok Shop by Tokopedia, not only brings convenience to shopping but also new challenges in the form of fake reviews that can mislead consumers, reduce trust, and harm both sellers and buyers. This study aims to build a fake review detection system for fashion products on Tiktok Shop by Tokopedia by utilizing a BERT (Bidirectional Encoder Representations from Transformers). The review dataset was obtained through a scraping process from Tiktok Shop by Tokopedia, then processed and trained using the BERT model with an added classification layer. The results show that the best training performance was achieved at epoch 4 with a training accuracy of 94,12%, while evaluation using classification metrics indicated optimal performance, with precision of 0,96, recall of 0,97, F1-score of 0,96, and accuracy of 0,94. These findings demonstrate that BERT is effective in detecting fake reviews, although challenges remain regarding overfitting at higher epochs. This research is expected to contribute to the development of fake review detection systems on the Tiktok Shop by Tokopedia platform.

Keywords: BERT, LLM, E-commerce, Fake Review

BAB I

PENDAHULUAN

1.1 Latar Belakang

TikTok Shop adalah salah satu contoh bagaimana lanskap e-commerce telah berubah secara dramatis karena perkembangan pesat internet dan teknologi digital. Di Indonesia, TikTok Shop menjadi salah satu platform yang tumbuh paling cepat khususnya pada kategori *fashion*. TikTok Shop awalnya beroperasi sebagai fitur belanja dalam aplikasi TikTok yang mengintegrasikan konten video, interaksi sosial, dan aktivitas transaksi, sehingga mampu memengaruhi keputusan pembelian konsumen secara signifikan (Priyono dan Sari, 2023).

Namun, Peraturan Menteri Perdagangan Nomor 31 Tahun 2023 tentang Perizinan Usaha, Periklanan, Bimbingan, dan Pengawasan Pelaku Usaha dalam Perdagangan melalui Sistem Elektronik yang dikeluarkan oleh Pemerintah Indonesia membatasi penggunaan platform media sosial dalam perdagangan elektronik.

Penyelenggara perdagangan melalui sistem elektronik dengan model bisnis perdagangan sosial tidak diperbolehkan memfasilitasi transaksi pembayaran secara langsung melalui sistem elektronik mereka, sebagaimana dinyatakan dalam Pasal 21 ayat (3) peraturan perundang-undangan. Akibat kebijakan tersebut, operasional TikTok Shop di Indonesia tidak lagi berdiri secara mandiri sebagai *social commerce* dan kemudian berintegrasi dengan platform *e-commerce* Tokopedia. Integrasi ini menghasilkan layanan yang dikenal sebagai TikTok Shop by Tokopedia, di mana aktivitas transaksi, pembayaran, dan pengelolaan *e-commerce* berada di bawah sistem Tokopedia, sementara TikTok berfungsi sebagai media promosi dan penyaluran konten produk. Meskipun demikian, dari sudut pandang pengguna, ulasan produk dan interaksi konsumen tetap menjadi bagian penting dalam proses pengambilan keputusan pembelian pada platform tersebut.

Salah satu faktor penting dalam proses pembelian *online* adalah ulasan konsumen (*review*), yang sering kali menjadi acuan utama dalam menilai kualitas produk (Hayati dkk., 2023). Namun, meskipun banyak manfaat yang dapat diperoleh dari ulasan, keberadaan ulasan palsu menjadi masalah besar yang perlu diatasi dalam platform *e-commerce*. Seiring meningkatnya volume transaksi dan persaingan antar penjual, muncul praktik yang merugikan yaitu ulasan palsu (*fake review*). Maraknya ulasan palsu atau manipulasi ulasan, di mana suatu produk mendapatkan ulasan positif atau negatif yang tidak sesuai dengan pengalaman nyata pengguna, telah menjadi fenomena yang sangat merugikan konsumen, penjual, dan platform *e-commerce*. Tindakan ini sering kali dilakukan oleh pihak-pihak tertentu untuk meningkatkan citra produk atau menjatuhkan reputasi pesaing. Ulasan palsu berpotensi mengubah persepsi konsumen terhadap kualitas produk baik secara positif maupun negatif, yang pada akhirnya memengaruhi keputusan pembelian. Pada kategori *fashion*, yang sangat dipengaruhi oleh opini dan estetika subjektif, keberadaan ulasan palsu dapat berdampak signifikan terhadap persepsi dan keputusan pembelian pengguna (Fauziya dan Isa, 2024). Oleh karena itu, untuk menjaga kepercayaan pada platform *e-commerce*, diperlukan upaya sistematis untuk mengidentifikasi dan menyelesaikan masalah ulasan palsu.

Deteksi ulasan palsu secara manual sangat tidak efisien, terutama mengingat jumlah data yang sangat besar. Oleh karena itu, dibutuhkan pendekatan berbasis *machine learning* untuk membantu proses klasifikasi *review* secara otomatis. Dalam penelitian sebelumnya oleh Alamsyah dkk., (2023), deteksi *fake review* dilakukan dengan menggunakan dataset *review* produk dari platform Tokopedia, serta menerapkan algoritma *Support Vector Machine* dan *Naïve Bayes*. Hasilnya, kedua model tersebut dinilai efektif karena mampu mengidentifikasi pola umum pada *fake review*, seperti penggunaan kata-kata yang berlebihan dan ketidaksesuaian dengan pengalaman pengguna sebenarnya. Namun, penelitian tersebut belum memanfaatkan kemampuan model bahasa yang lebih canggih.

Sebagian besar penelitian sebelumnya masih berfokus pada algoritma

machine learning konvensional dan belum menerapkan model bahasa besar seperti BERT untuk konteks bahasa Indonesia, khususnya pada platform *e-commerce*. Selain itu, sejumlah penelitian terdahulu yang menggunakan BERT umumnya difokuskan pada bidang lain, seperti deteksi ulasan palsu pada hotel dan restoran, identifikasi berita hoaks, analisis sentimen di media sosial, serta klasifikasi opini publik. Meskipun menggunakan pendekatan yang sama, penelitian-penelitian tersebut belum menyoroiti konteks *e-commerce* di Tiktok Shop by Tokopedia, terutama pada kategori *fashion*. Kondisi ini menunjukkan adanya kebutuhan untuk mengembangkan model deteksi ulasan palsu dengan pendekatan modern yang relevan dengan karakteristik bahasa dan konteks pengguna pada platform tersebut.

Berdasarkan keterbatasan tersebut, pada penelitian ini penulis akan mengembangkan sistem deteksi ulasan palsu dengan menggunakan BERT (*Bidirectional Encoder Representations from Transformers*). Model ini mampu memahami konteks dua arah dari sebuah teks, sehingga dapat menginterpretasikan makna kalimat secara lebih mendalam dan mampu memberikan hasil analisis yang lebih akurat. Penelitian ini secara khusus difokuskan untuk mendeteksi *fake review* secara otomatis pada kategori *fashion* di platform Tiktok Shop by Tokopedia, yang hingga saat ini belum banyak menjadi objek penelitian sejenis. Selain itu, diharapkan studi ini akan membantu dalam pembuatan sistem pendeteksi ulasan palsu berbasis bahasa Indonesia, mengingat masih terbatasnya studi yang fokus pada konteks lokal di platform *e-commerce* tersebut.

1.2 Perumusan Masalah

1. Bagaimana merepresentasikan ulasan yang dapat memahami konteks menggunakan BERT ?
2. Bagaimana cara mendeteksi ulasan palsu pada produk *fashion* yang dijual di Tiktok Shop by Tokopedia dengan menggunakan BERT *pre-trained* ?

1.3 Pembatasan Masalah

Pembatasan masalah dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya berfokus pada platform Tiktok Shop by Tokopedia, meskipun deteksi ulasan palsu juga merupakan masalah besar pada platform *e-commerce* lain.
2. Penelitian ini terbatas pada produk *fashion* yang meliputi pakaian, tas, sepatu, dan aksesoris yang dijual di Tiktok Shop by Tokopedia. Meskipun ulasan palsu dapat ditemukan pada berbagai jenis produk, produk *fashion* dipilih karena sektor ini lebih banyak bergantung pada opini dan pengalaman pengguna, sehingga lebih rentan terhadap manipulasi ulasan palsu.
3. Penelitian ini hanya akan mencakup ulasan teks yang ditulis oleh konsumen di Tiktok Shop by Tokopedia. Ulasan berbentuk gambar atau video tidak akan menjadi bagian dari analisis ini.
4. Data yang digunakan dalam penelitian ini hanya berasal dari ulasan yang tersedia secara publik di Tiktok Shop by Tokopedia. Penelitian ini tidak akan melibatkan data historis yang lebih lama atau data yang tidak dapat diakses melalui platform Tiktok Shop by Tokopedia.
5. Penelitian ini membatasi analisis pada perbedaan karakteristik antara ulasan asli dan ulasan palsu berdasarkan aspek linguistik, struktural, emosional, serta gaya dan isi penulisan, sehingga model deteksi hanya akan mengevaluasi ulasan berdasarkan ciri-ciri tersebut.
6. Penelitian ini hanya berfokus pada penggunaan model BERT dalam proses deteksi ulasan palsu. Pengujian dilakukan terhadap dataset yang telah ditentukan tanpa melakukan analisis terhadap model lain maupun penerapan sistem deteksi dalam skala besar di lingkungan operasional Tiktok Shop by Tokopedia.

1.4 Tujuan

Adapun tujuan penelitian ini sebagai berikut:

1. Merepresentasikan ulasan produk *fashion* secara kontekstual menggunakan BERT

2. Mengimplementasikan serta mengembangkan metode deteksi ulasan palsu menggunakan BERT pada produk *fashion* di Tiktok Shop by Tokopedia.
3. Mengevaluasi dan menguji akurasi model deteksi ulasan palsu pada produk *fashion* di Tiktok Shop by Tokopedia menggunakan model BERT.

1.5 Manfaat

Manfaat dari penelitian ini diharapkan dapat diperoleh baik di bidang akademis maupun praktis, khususnya di bidang-bidang berikut:

1. Bagi Akademisi

Studi ini dapat dijadikan referensi untuk kemajuan pembelajaran mesin dan pemrosesan bahasa alami (NLP), khususnya dalam penggunaan model bahasa besar berbasis BERT untuk mengidentifikasi ulasan berbahasa Indonesia yang curang.

2. Bagi Platform *E-commerce*

Penelitian ini dapat membantu meningkatkan kepercayaan konsumen terhadap platform *e-commerce*, khususnya Tiktok Shop by Tokopedia, dengan menyediakan pendekatan otomatis untuk mendeteksi dan meminimalkan ulasan palsu.

3. Bagi Penjual

Penelitian ini dapat membantu penjual dalam menjaga reputasi toko dari ulasan manipulatif, serta memberikan gambaran yang lebih objektif mengenai kualitas produk yang ditawarkan.

4. Bagi Konsumen

Berdasarkan ulasan yang valid dan terpercaya, penelitian ini dapat membantu konsumen membuat keputusan pembelian yang lebih tepat.

1.6 Sistematika Penulisan

Untuk mempermudah penulisan tugas akhir ini, penulis membuat suatu sistematika yang terdiri dari:

BAB I : PENDAHULUAN

Latar belakang perumusan masalah, tujuan penelitian, manfaat penelitian, ruang lingkup masalah, dan batasan masalah semuanya akan dijelaskan dalam bab ini.

BAB II : TINJAUAN PUSTAKA DAN DASAR TEORI

Bab ini mencakup landasan teoritis dan penelitian sebelumnya yang dapat digunakan sebagai sumber atau alat untuk memahami isu-isu seputar penggunaan model BERT.

BAB III : METODOLOGI PENELITIAN

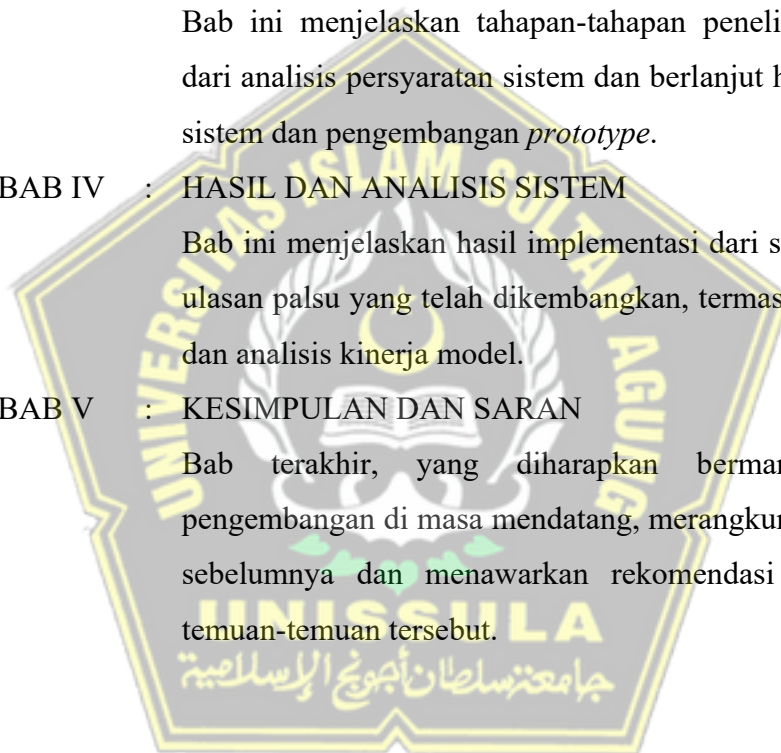
Bab ini menjelaskan tahapan-tahapan penelitian, dimulai dari analisis persyaratan sistem dan berlanjut hingga desain sistem dan pengembangan *prototype*.

BAB IV : HASIL DAN ANALISIS SISTEM

Bab ini menjelaskan hasil implementasi dari sistem deteksi ulasan palsu yang telah dikembangkan, termasuk pengujian dan analisis kinerja model.

BAB V : KESIMPULAN DAN SARAN

Bab terakhir, yang diharapkan bermanfaat untuk pengembangan di masa mendatang, merangkum isi bab-bab sebelumnya dan menawarkan rekomendasi berdasarkan temuan-temuan tersebut.



BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Fake review atau ulasan palsu merupakan salah satu sumber informasi penting dalam proses pengambilan keputusan pembelian secara *online* (Kirom, 2024). Meskipun ulasan negatif dapat membuat pelanggan memutuskan untuk tidak membeli suatu produk, ulasan positif dapat meningkatkan kepercayaan konsumen dan mendorong pembelian. Namun, dalam praktiknya, terdapat ulasan palsu (*fake review*) yang ditulis bukan berdasarkan pengalaman nyata, melainkan bertujuan untuk menipu atau memanipulasi opini publik. Menurut Fauziya dan Isa, (2024) *fake review* sering kali ditandai dengan penggunaan kata-kata yang *hiperbolik*, tidak konsisten dengan konteks produk, dan ditulis secara tidak alami.

Sebaliknya, ulasan asli ditulis oleh pelanggan yang benar-benar menggunakan barang atau jasa tersebut dan bersifat otentik dan informatif, berbeda dengan ulasan palsu yang lebih dangkal. Ulasan asli lebih banyak mengandung informasi yang mendalam dan substansial daripada ulasan palsu, yang memiliki dampak lebih besar pada keputusan pembeli mereka (Kim dkk., 2023). *Fake review* dapat membahayakan ekosistem *e-commerce* karena memengaruhi pandangan pelanggan dan menciptakan persaingan yang tidak sehat.

Tiktok Shop by Tokopedia merupakan fitur *e-commerce* yang terintegrasi langsung dalam aplikasi TikTok. *Platform* ini memadukan konten video pendek dan promosi produk secara langsung, sehingga mampu memengaruhi keputusan konsumen secara *real time* (Priyono dan Dian Permata Sari, 2023). Kategori *fashion* menjadi salah satu kategori paling populer di Tiktok Shop by Tokopedia, dengan tingginya volume ulasan yang diberikan oleh pengguna. Hal ini menjadikan Tiktok Shop by Tokopedia sebagai *platform* yang rawan terhadap manipulasi ulasan, terutama karena sifatnya yang sangat visual dan mudah menjadi *viral*.

Pada penelitian sebelumnya oleh Elmogy dkk., (2021), beberapa algoritma klasifikasi dibandingkan, yaitu (*K-Nearest Neighbor*) KNN, (*Naive Bayes*) NB, (*Support Vector Machine*) SVM, *Logistic Regression*, dan *Random Forest*. Penelitian ini juga mempertimbangkan model bahasa berbasis *n-gram*, khususnya *bigram* dan *trigram* dalam proses evaluasi. Hasil penelitian menunjukkan bahwa KNN ($K = 7$) memberikan performa terbaik dengan nilai *F-score* sebesar 82,40%. Selain itu, ketika fitur perilaku pengguna dimasukkan, terjadi peningkatan nilai *F-score* sebesar 3,80%, yang menunjukkan bahwa faktor perilaku pengguna memainkan peran penting dalam meningkatkan akurasi deteksi *fake review*.

Penelitian yang dilakukan oleh Vidya Chandradev dkk., (2023), menggunakan model *deep learning* BERT (*Bidirectional Encoder Representations from Transformers*) untuk membahas analisis sentimen ulasan hotel di Indonesia. Sentimen positif dan negatif dari ulasan pengunjung hotel dikategorikan menggunakan model ini sebagai bentuk evaluasi layanan dan strategi pengembangan bisnis perhotelan. Dataset yang digunakan mencakup sekitar 515.000 ulasan hotel, dan model SmallBERT digunakan setelah dilatih dan disesuaikan selama 5 epoch. Studi menunjukkan bahwa model SmallBERT memiliki akurasi 0,91, presisi 0,90, *recall* 0,90, dan skor F1 0,90. Hasil ini menunjukkan bahwa model BERT dapat memahami konteks bahasa alami dengan baik dan memberikan hasil klasifikasi sentimen yang akurat untuk teks berbahasa Indonesia.

Penelitian yang dilakukan oleh Khoirotulmuadiba Purifyregalia dkk., (2025) penelitian ini menggunakan algoritma (*Support Vector Machine*) SVM dan *Random Forest*. Dengan menggunakan 3.686 data ulasan produk yang telah dilakukan *preprocessing* seperti pembersihan data, normalisasi, tokenisasi, dan pembobotan menggunakan TF-IDF. Hasil evaluasi menunjukkan bahwa algoritma SVM memiliki performa terbaik dengan akurasi sebesar 88,84%, mengungguli *Random Forest* dengan 80,39%. Penelitian ini juga menunjukkan bahwa algoritma SVM efektif dalam menangani data teks yang sangat besar, serta membantu penerapan LDA

(pemodelan topik otomatis) untuk pelabelan otomatis ulasan dalam konteks e-commerce berbahasa Indonesia.

Penelitian yang dilakukan oleh Putu dkk., (2023) membahas analisis sentimen terhadap kasus perundungan siber (*cyberbullying*) dalam bahasa Indonesia dengan menggunakan model (*Bidirectional Encoder Representations from Transformers*) BERT. Penelitian ini memanfaatkan dataset teks berbahasa Indonesia yang telah diklasifikasikan ke dalam tiga kategori, yaitu netral, bahasa kasar, dan ujaran kebencian. Hasil pengujian menunjukkan bahwa model BERT menghasilkan performa yang baik dengan tingkat akurasi sebesar 0,81, presisi 0,80, *recall* 0,80, serta *F1-score* sebesar 80,67%. Temuan ini menunjukkan bahwa algoritma BERT memiliki kemampuan yang cukup tinggi dalam memahami konteks bahasa alami dan mendeteksi ujaran kebencian atau perilaku perundungan siber secara akurat dalam teks berbahasa Indonesia.

Sementara itu, penelitian oleh Karimah dan Baita, (2024) membahas penerapan analisis sentimen multi-aspek pada ulasan film menggunakan metode BERT. Penelitian ini bertujuan untuk mengidentifikasi polaritas sentimen (positif, negatif, atau netral) berdasarkan beberapa aspek penilaian film, yaitu *acting*, *plot*, *cast*, *animation*, dan *music*. Model yang digunakan adalah BERT *Base Uncased*, dengan beberapa percobaan *hyperparameter tuning* menggunakan nilai *learning rate* sebesar $1e-4$, $5e-5$, $3e-5$, dan $2e-5$. Dataset yang digunakan terdiri atas kumpulan ulasan film berbahasa Indonesia yang dianalisis berdasarkan tiap aspek tersebut. Hasil penelitian menunjukkan bahwa konfigurasi model terbaik diperoleh pada *learning rate* $3e-5$, dengan tingkat akurasi sebesar 82,32%. Evaluasi model dengan nilai *precision* 0,86, *recall* 0,85, dan *F1-score* 0,85, sedangkan kinerja terendah terdapat pada aspek *plot*. Penelitian ini juga mencatat bahwa ketidakseimbangan distribusi data antar-aspek berpengaruh terhadap performa model. Namun, secara keseluruhan, model BERT dinilai cukup efektif dalam melakukan analisis sentimen multi-aspek pada ulasan film.

Dari hasil penelitian-penelitian tersebut, penulis dapat menyimpulkan

bahwa model BERT memiliki potensi yang kuat untuk digunakan dalam membangun sistem deteksi ulasan palsu (*fake review*), khususnya pada platform *e-commerce* modern seperti Tiktok Shop by Tokopedia. Beberapa studi terdahulu menunjukkan bahwa pendekatan berbasis *machine learning* klasik seperti SVM, *Naive Bayes*, dan *Random Forest* mampu memberikan hasil yang cukup baik dalam klasifikasi ulasan, namun memiliki keterbatasan dalam memahami konteks bahasa secara mendalam. Sementara itu, BERT mampu menangkap makna kontekstual dari kata-kata dalam kalimat secara dua arah, sehingga lebih efektif dalam memahami struktur kalimat yang kompleks dan mendeteksi pola-pola tersembunyi dalam teks ulasan. Hal ini sangat relevan untuk konteks Tiktok Shop by Tokopedia, di mana ulasan pengguna sering kali singkat, informal, dan menggunakan bahasa gaul atau lokal. Dengan demikian, penerapan model BERT dalam konteks ulasan berbahasa Indonesia pada platform seperti Tiktok Shop by Tokopedia merupakan langkah logis untuk meningkatkan akurasi dan ketepatan dalam mendeteksi ulasan palsu.

2.2 Dasar Teori

2.2.1 Ulasan asli (*genuine review*)

Ulasan asli (*genuine review*) merupakan ulasan yang ditulis langsung oleh manusia sebagai konsumen berdasarkan pengalaman nyata terhadap suatu produk atau layanan. Ulasan ini dihasilkan dari interaksi langsung dengan produk, bukan melalui bot. Ciri ulasan ini biasanya memiliki konten yang kredibel, relevan dengan produk, serta disampaikan secara natural tanpa adanya indikasi manipulasi. Menurut Tran dkk., (2022), ulasan yang kredibel atau asli mampu meningkatkan kepercayaan konsumen terhadap merek karena dianggap sebagai representasi autentik dari pengalaman pengguna. Dengan demikian, ulasan asli tidak hanya berfungsi sebagai sumber informasi, tetapi juga sebagai faktor penting dalam membentuk kepercayaan merek dan memengaruhi niat beli konsumen.

Menurut Pooja and Upadhyaya, (2024), ulasan asli biasanya ditandai

dengan beberapa karakteristik penting:

1. Detail dan spesifik, berisi pengalaman nyata yang dijelaskan dengan rinci, misalnya kualitas produk, pelayanan, atau kondisi pengiriman.
2. Konsistensi dengan produk/jasa, sehingga isi ulasan sesuai dengan konteks produk yang dibeli, tidak melebar ke hal yang tidak relevan.
3. Bahasa yang digunakan alami, seperti gaya bahasa ulasan mengalir wajar, tanpa terlihat dibuat-buat atau terlalu hiperbolis.
4. Keberimbangan opini, umumnya konsumen mencantumkan kelebihan sekaligus kekurangan, sehingga lebih objektif.
5. Mengandung informasi verifikatif, seperti mencantumkan bukti pengalaman, seperti foto produk, waktu penggunaan, atau kondisi barang saat diterima.

Dengan demikian, ulasan ini tidak hanya berfungsi sebagai sumber informasi, tetapi juga sebagai bahan pertimbangan yang rasional dalam proses pengambilan keputusan pembelian.

2.2.2 Ulasan palsu (*fake review*)

Dalam konteks *e-commerce* dan layanan digital, ulasan atau *review* konsumen merupakan salah satu elemen penting yang memengaruhi persepsi calon pembeli terhadap suatu produk atau layanan. Ulasan yang diberikan oleh pengguna sering dijadikan acuan dalam pengambilan keputusan pembelian. Namun, tidak semua ulasan bersifat jujur dan autentik. Seiring dengan meningkatnya persaingan bisnis secara *online*, praktik ulasan palsu (*fake review*) semakin sering terjadi dan menjadi ancaman serius bagi kepercayaan konsumen serta integritas platform *e-commerce*.

Ulasan palsu yang dimaksud di sini adalah ulasan yang diberikan dengan maksud memanipulasi opini konsumen, baik untuk meningkatkan citra produk sendiri maupun menurunkan reputasi pesaing. Ulasan tersebut umumnya dihasilkan oleh manusia, misalnya penjual atau pihak ketiga yang dibayar untuk menulis pujian berlebihan atau kritik negatif. Tujuan utamanya adalah untuk memengaruhi persepsi calon pembeli tanpa mencerminkan pengalaman nyata pengguna. Meskipun dalam praktiknya

ada juga bot yang digunakan untuk memperbanyak jumlah ulasan, penelitian ini ingin menganalisis pola perilaku manusia yang sengaja menulis ulasan untuk memengaruhi opini konsumen. Pola manipulasi yang kompleks seperti penggunaan gaya bahasa tertentu, hiperbola, deskripsi generik, dan orientasi promosi lebih sering muncul pada ulasan yang dibuat oleh manusia dibandingkan bot. Dengan demikian, manusia tetap menjadi aktor utama dalam praktik manipulasi ulasan yang dianalisis pada penelitian ini.

Menurut Awalina dkk., (2022), praktik ini dikenal dengan istilah *opinion spam* atau opini palsu, yaitu aktivitas ilegal atau tidak etis yang bertujuan untuk menggiring opini publik melalui pemberian ulasan yang tidak sesuai dengan kenyataan. Ulasan tersebut bisa berupa pujian berlebihan untuk menaikkan citra produk atau justru kritik negatif yang ditujukan untuk menjatuhkan reputasi pesaing. Tujuan utamanya adalah untuk memanipulasi persepsi konsumen agar menguntungkan pihak tertentu.

Sejumlah penelitian telah mengidentifikasi ciri-ciri ulasan palsu dari berbagai aspek, termasuk linguistik, struktural, emosional, serta gaya dan isi penulisan. Berdasarkan temuan dari beberapa studi, ciri-ciri tersebut dijelaskan sebagai berikut:

1. Aspek linguistik – bentuk dan pola bahasa

Menurut Abri dkk., (2020) dalam penelitiannya ‘*Fake reviews Detection through Analysis of Linguistic Features*’ ulasan palsu cenderung memiliki ciri-ciri linguistik sebagai berikut:

- Pengulangan kata atau frasa (*redundant terms*): penggunaan kata atau frasa yang sama secara berlebihan sehingga teks tampak tidak natural dan terkesan dibuat-buat.
- Adanya tanda jeda atau *filler* yang tidak wajar: misalnya kata pengisi yang tidak memberikan makna signifikan, sehingga mengganggu alur baca dan keaslian ulasan.
- Struktur kalimat panjang dan kompleks: berbeda dari ulasan asli yang biasanya lebih ringkas, jelas, dan mudah dipahami.

2. Aspek *linguistic features* - gaya dan isi penulisan

Penelitian oleh Abri dkk., (2020) dengan fokus pada '*Linguistic Cues in Online Reviews*' menekankan bahwa fitur linguistik lain juga penting dalam mendeteksi ulasan palsu, antara lain:

- Satu sisi (*one-sidedness*): ulasan hanya menyoroti sisi positif atau negatif tanpa memberikan penilaian seimbang.
- Hiperbola (*exaggeration*): penggunaan klaim berlebihan, misalnya 'terbaik sepanjang masa' atau 'sangat buruk, jangan beli!', menunjukkan kecenderungan melebih-lebihkan kondisi produk.
- Deskripsi generik: kalimat umum yang tidak menyebutkan detail spesifik produk, sehingga informasi sulit diverifikasi

3. Aspek emosional dan orientasi promosi

Penelitian oleh Salminen dkk., (2025) menekankan bahwa ulasan palsu sering menunjukkan orientasi promosi dan emosi yang ekstrem:

- Berorientasi pada promosi atau harga: menekankan diskon, ajakan membeli, atau keunggulan produk secara berlebihan.
- Mengandung emosi ekstrem: baik pujian berlebihan maupun kritik keras, sehingga menimbulkan kesan manipulatif dan tidak proporsional.

4. Aspek struktural

Penelitian oleh Kumar dkk., (2024) mengungkapkan bahwa ulasan palsu dapat dikenali melalui ciri-ciri struktural:

- Jumlah kata dan kalimat lebih sedikit: ulasan terkesan dangkal dan minim detail pengalaman nyata.
- Rasio kata bermakna (*content words*) rendah: ulasan palsu hanya mencapai rasio sekitar 0,96 dibanding ulasan asli yang rata-rata 1,22, menunjukkan informasi yang disampaikan kurang substantif dan kurang dapat diverifikasi.

5. Aspek relevansi konten

Penelitian oleh Mohawesh dkk., (2021) menegaskan adanya kategori *non-reviews* atau *irrelevant reviews* sebagai ciri ulasan palsu:

- *Brand-only reviews*: membahas merek atau penjual tanpa

menyinggung produk spesifik.

- *Irrelevant/off-topic reviews*: ulasan membicarakan hal lain di luar produk (misalnya layanan, pengalaman di toko, atau topik yang sama sekali tidak terkait fashion/barang yang dibeli).

Dari kelima jurnal tersebut, dapat disimpulkan bahwa ulasan palsu ditandai oleh pola bahasa dan konten yang tidak natural, minim informasi, cenderung manipulatif, serta tidak selalu relevan dengan produk. Untuk mendeteksinya, beberapa metode umum memanfaatkan analisis teks dan pola perilaku pengguna, dengan dua pendekatan utama yang sering diterapkan, yaitu:

1. Analisis Sentimen

Menilai sikap, opini, atau emosi dalam teks, termasuk intensitas berlebihan dan pola ekspresi yang tidak wajar, untuk mengungkap ulasan manipulatif.

2. Pengenalan Pola (*Pattern Recognition*)

Mengidentifikasi struktur, gaya bahasa, dan perilaku posting, seperti pengulangan kata, panjang kalimat, rasio kata bermakna, serta frekuensi posting yang tidak wajar.

Kombinasi kedua metode ini, sering dipadukan dengan *machine learning* atau LLM seperti BERT, sehingga memungkinkan sistem deteksi ulasan palsu yang lebih akurat dan responsif, sekaligus menjaga kepercayaan konsumen, integritas sistem rekomendasi, serta validitas informasi di platform *e-commerce*.

2.2.3 Tiktok Shop by Tokopedia sebagai platform *e-commerce*

Tiktok Shop by Tokopedia merupakan platform *e-commerce* berbasis media sosial yang berkembang pesat di Indonesia, yang menggabungkan pengalaman berbelanja dengan konten video. Integrasi ini memungkinkan pelanggan untuk melihat produk secara langsung melalui video singkat maupun *live streaming*, dan melakukan pembelian produk secara langsung

melalui aplikasi Tiktok Shop by Tokopedia. Keunikan Tiktok Shop by Tokopedia terletak pada interaksi yang intens antara *influencer*, penjual, dan konsumen dalam format visual dan interaktif, yang dapat memperkuat promosi produk dan pengalaman berbelanja pengguna (Katampanan dkk., 2023).

Salah satu karakteristik penting dari platform ini adalah konten yang dipandu oleh tren dan pengaruh pengguna, di mana banyak konsumen melakukan keputusan pembelian berdasarkan rekomendasi *influencer* atau pengalaman pengguna lain yang diposting di platform. Karakteristik ini sekaligus menimbulkan potensi manipulasi ulasan oleh pihak yang tidak bertanggung jawab. Oleh karena itu, penelitian ini memfokuskan pada Tiktok Shop by Tokopedia sebagai kasus studi khusus untuk mengidentifikasi dan menganalisis ulasan palsu pada produk kategori *fashion*, dengan tujuan memahami pola manipulasi dan dampaknya terhadap persepsi konsumen (Priyono dan Dian Permata Sari, 2023).

2.2.4 Produk *fashion* dalam *e-commerce*

Kata "fashion" berasal dari etimologi kata Latin "factio," yang berarti "melakukan." Kata Latin ini akhirnya masuk ke dalam bahasa Inggris sebagai "fashion," yang sekarang didefinisikan sebagai cara berpakaian populer dalam suatu masyarakat. Fashion, gaya, tata krama, pakaian, dan busana semuanya dianggap sebagai aspek dari fashion. Akibatnya, pakaian, jeans, tas, sepatu, dan aksesoris lain yang meningkatkan penampilan pemakainya adalah contoh produk fashion yang dibuat melalui proses produksi. Pelanggan sering mempertimbangkan citra diri yang ingin mereka tampilkan dan kritik sosial sebelum melakukan pembelian produk fashion.

Salah satu kategori produk yang sangat bergantung pada rekomendasi dan visualisasi media sosial adalah fashion. Penelitian (Mulyati and Gesitera, 2020) menunjukkan bahwa ulasan konsumen *online* berpengaruh signifikan terhadap niat pembelian, dengan kepercayaan sebagai faktor yang memediasi hubungan tersebut. Hal ini menegaskan bahwa kepercayaan konsumen terhadap ulasan yang mereka baca menjadi faktor utama dalam menentukan

minat pembelian produk *fashion* melalui *e-commerce*. Produk *fashion* sangat rentan terhadap manipulasi ulasan karena konsumen lebih cenderung membeli produk yang dikenalkan oleh *influencer* atau teman sebaya. Ulasan positif yang palsu dapat membuat produk yang tidak berkualitas terlihat lebih menarik di mata konsumen, sedangkan ulasan negatif palsu dapat merusak reputasi suatu merek. Hal ini menjadi tantangan besar bagi platform *e-commerce* yang menjual produk *fashion*, terutama dalam menjaga kepercayaan konsumen dan mencegah praktik manipulatif.

2.2.5 NLP (*Natural Language Processing*)

NLP (*Natural Language Processing*) merupakan salah satu bidang kecerdasan buatan yang berfokus pada cara komputer berinteraksi dengan Bahasa manusia. Penelitian Chang, (2023) menyatakan bahwa NLP berada di persimpangan antara kecerdasan buatan dan linguistik, dan mencakup kemampuan mesin untuk memahami, memproses, mereproduksi, dan menghasilkan bahasa alami secara otomatis. Dalam perkembangannya, NLP telah digunakan untuk berbagai aplikasi seperti klasifikasi teks, pemrosesan bagian kata (*part-of-speech tagging*), analisis semantik, penerjemahan mesin, hingga pengorganisasian teks (*text categorization*).

Teknik NLP yang digunakan dalam penelitian ini yaitu:

1. *Lowercasing dan Cleaning*: Mengubah seluruh teks menjadi huruf kecil serta menghapus tanda baca, angka, dan karakter khusus yang tidak relevan.
2. *Emoji Removal*: Menghapus emoji dari teks agar tidak mengganggu analisis semantik.
3. *Tokenization*: Memecah teks menjadi kata-kata atau frasa untuk analisis lebih lanjut.
4. *Padding*: Menyamakan panjang setiap token agar sesuai dengan batas maksimum *input* model BERT (*max_length*).
5. *Attention Masking*: Memberi penanda pada token yang asli dan hasil *padding*, sehingga model dapat memusatkan perhatian pada bagian teks yang relevan.

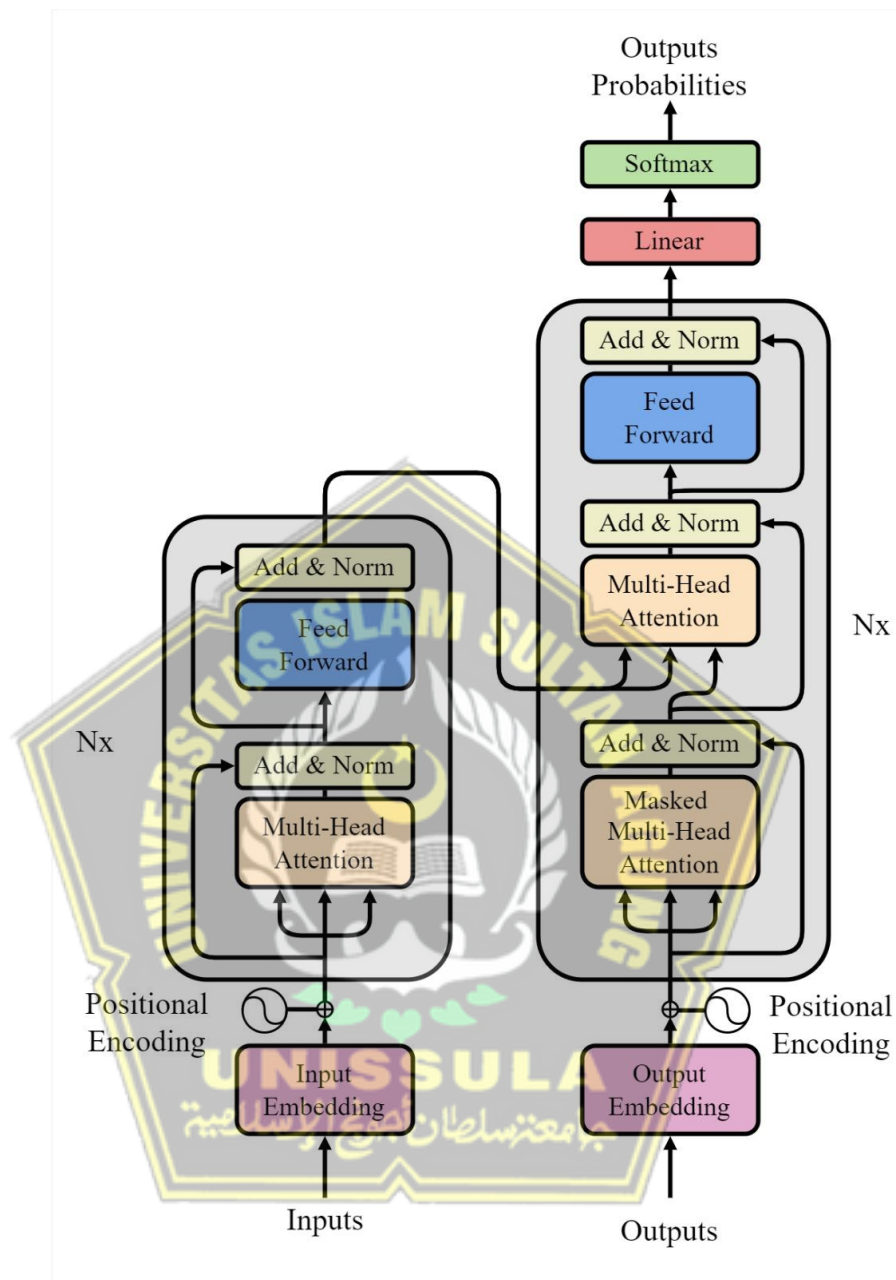
6. *Word Embedding* (BERT *Embedding*): proses mengubah setiap token (kata atau *sub*-kata) menjadi representasi vektor numerik yang memiliki makna kontekstual.
7. *Fine-Tuning* BERT: Melatih ulang model BERT pada dataset ulasan untuk mendeteksi ulasan asli dan ulasan palsu sesuai konteks penelitian.

Menurut Jurafsky & Martin (2021), kombinasi dari teknik tersebut memungkinkan model NLP untuk lebih efektif dalam memahami nuansa dalam teks ulasan dan mendeteksi ketidaknormalan atau manipulasi yang mungkin terjadi.

2.2.6 LLM (*Large Language Models*)

LLM (*Large Language Models*) adalah jenis model pembelajaran mesin yang dilatih menggunakan jumlah data teks yang sangat besar, umumnya menggunakan arsitektur *transformer*. Model ini memiliki kemampuan luar biasa dalam memahami dan menghasilkan teks yang menyerupai tulisan manusia, sehingga dapat digunakan untuk berbagai tugas pemrosesan bahasa alami (NLP) dan aplikasi lainnya (Naveed dkk., 2025).

Beberapa contoh LLM yang terkenal adalah BERT (*Bidirectional Encoder Representations from Transformers*) dan GPT (*Generative Pre-trained Transformer*), yang keduanya menunjukkan kemampuan tinggi dalam memahami konteks dan menghasilkan teks mirip tulisan manusia. GPT-3, yang dikembangkan oleh OpenAI, menonjol karena kemampuannya menghasilkan teks alami dan melakukan berbagai tugas NLP, seperti analisis sentimen dan penerjemahan bahasa. Model ini memiliki lebih dari 175 miliar parameter, memungkinkan pemahaman konteks yang luas bahkan dengan sedikit data pelatihan tambahan (Brown dkk., 2020).



Gambar 2.1 Arsitektur Model *Transformer*

Gambar 2.1 adalah arsitektur model *transformer*, yaitu model dasar yang digunakan untuk membangun LLM seperti GPT, BERT, T5, dan lainnya. Gambar ini menggambarkan arsitektur *transformer encoder-decoder*, yang menjadi fondasi utama dari berbagai model bahasa besar. Bagian kiri merupakan *encoder* yang berfungsi memproses input dan membangun representasi konteks, sedangkan bagian kanan merupakan

decoder yang menghasilkan keluaran berdasarkan representasi tersebut. Model seperti T5 (*Text-to-Text Transfer Transformer*) dan BART (*Bidirectional and Auto-Regressive Transformers*) menggunakan arsitektur *encoder-decoder* secara lengkap, sementara GPT hanya menggunakan bagian *decoder*, dan BERT hanya menggunakan bagian *encoder*.

Pada bagian *encoder*, teks terlebih dahulu diubah menjadi vektor melalui proses *input embedding*, kemudian ditambahkan informasi urutan kata menggunakan *positional encoding* agar model memahami posisi relatif antar kata. Setelah itu, data diproses melalui sejumlah lapisan (Nx layers), yang masing-masing terdiri dari *feed-forward network* dan *multi-head attention*. *feed-forward network* meningkatkan representasi fitur akhir, dan komponen *multi-head attention* memungkinkan model untuk secara simultan memantau hubungan antar kata dalam kalimat dari konteks yang berbeda. Setiap lapisan dilengkapi dengan mekanisme *Add & Norm* untuk menjaga kestabilan pelatihan melalui *residual connection* dan normalisasi.

Sementara itu, bagian *decoder* menghasilkan output berdasarkan informasi dari *encoder*. Prosesnya dimulai dari *output embedding* dan *positional encoding*, kemudian dilanjutkan ke *masked multi-head attention* yang memastikan model hanya memperhatikan token sebelumnya agar urutan keluaran tetap logis. Setelah itu, lapisan *multi-head attention* tambahan menghubungkan hasil *encoder* agar konteks penuh dapat dimanfaatkan. Hasil akhirnya diproses melalui *feed-forward network* dan *Add & Norm*, lalu diteruskan ke lapisan *linear* dan *softmax* untuk menghasilkan *output probabilities*, yaitu probabilitas token keluaran yang paling sesuai.

Penggunaan LLM dalam penelitian deteksi ulasan palsu sangat bermanfaat karena model ini mampu mengenali pola semantik yang kompleks dan halus dalam teks ulasan, bahkan ketika ulasan tersebut tidak secara eksplisit menunjukkan tanda-tanda manipulasi. Dengan pemahaman konteks yang dalam dan kemampuan representasi bahasa yang kuat, model berbasis *transformer* dapat memberikan hasil deteksi yang lebih akurat dan kontekstual.

2.2.7 BERT (*Bidirectional Encoder Representations from Transformers*)

BERT (*Bidirectional Encoder Representations from Transformers*) adalah model *deep learning* berbasis arsitektur *transformer* yang dikembangkan oleh *Google*, yang dirancang untuk memahami representasi kontekstual dari teks secara mendalam. BERT memiliki kemampuan untuk memahami konteks suatu kata dengan melihat kedua arah (*bidirectional*), yaitu dari kiri ke kanan dan dari kanan ke kiri, sehingga dapat menangkap makna kalimat secara lebih menyeluruh (Hanum dkk., 2024).

Setiap layer dalam *encoder* BERT menggunakan mekanisme *multi-head self-attention* untuk mengidentifikasi keterkaitan antar token dalam konteks kalimat. Fungsi *attention* dapat digambarkan sebagai proses pemetaan antara sebuah *query* dan sekelompok pasangan *key-value*, untuk menghasilkan representasi keluaran. Baik *query* (Q), *key* (K), *value* (V), maupun *output*-nya, semuanya direpresentasikan dalam bentuk vektor berdimensi tetap. *Output attention* diperoleh sebagai jumlah tertimbang dari *value*, dengan bobot yang dihitung berdasarkan skor kesesuaian (*compatibility score*) antara *query* dan *key* (Putu dkk., 2023).

Mekanisme dasar yang digunakan dalam *transformer* adalah *Scaled Dot-product Attention*, yang dirumuskan sebagai berikut :

$$Attention(Q, K, V) = softmax\left(\frac{QK^t}{\sqrt{d_k}}\right)V \quad (1)$$

Dimana:

Q : *Query*, mewakili informasi dari kata yang sedang dianalisis.

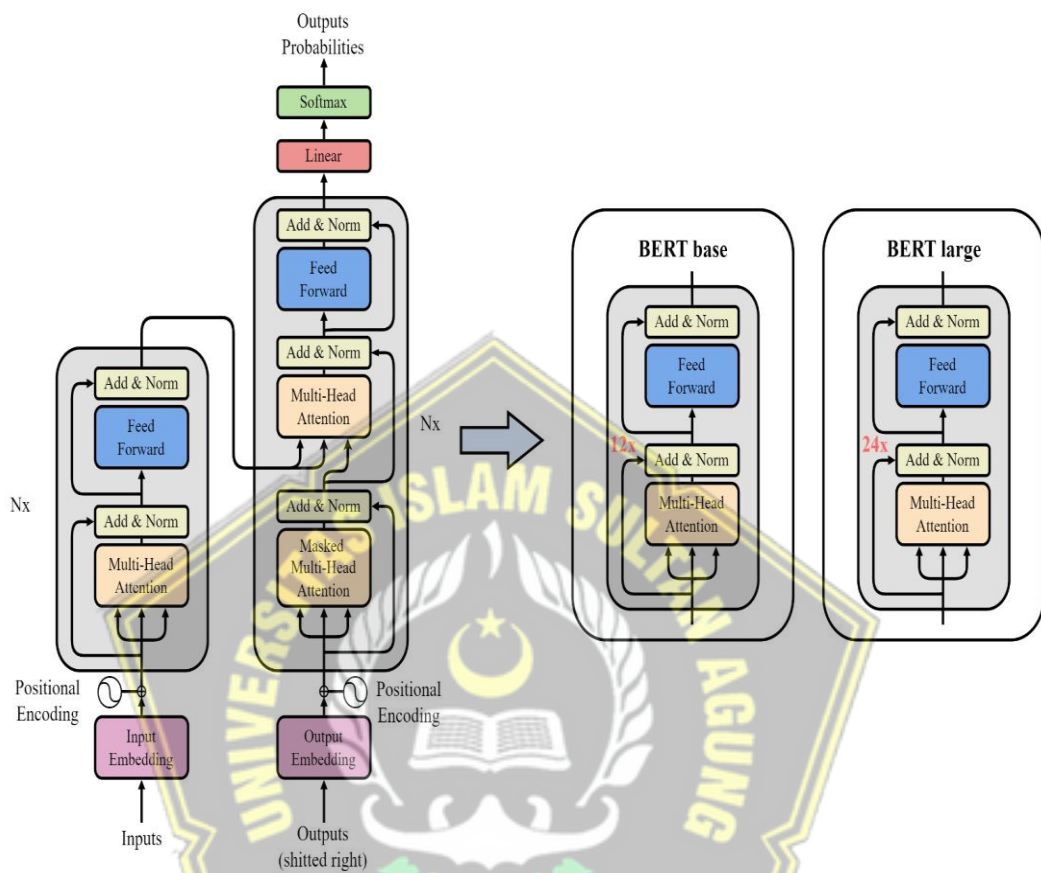
K : *Key*, mewakili konteks dari semua kata lain dalam kalimat.

V : *Value*, digunakan untuk menghasilkan representasi akhir.

d_k : dimensi dari vektor *key*, digunakan untuk normalisasi.

Proses *attention* ini memastikan bahwa setiap kata dipertimbangkan berdasarkan relevansi maknanya terhadap kata lain dalam kalimat. Dengan demikian, model tidak hanya memahami kata secara individual, tetapi juga

hubungan kontekstual antar kata sebelum direpresentasikan menjadi vektor akhir.



Gambar 2.2 Model Arsitektur BERT

Gambar 2.2 menunjukkan arsitektur model BERT yang dibangun berdasarkan arsitektur *transformer encoder*. Secara umum, gambar tersebut memperlihatkan bagaimana struktur dasar *transformer* (seperti yang ditunjukkan di sebelah kiri) digunakan sebagai fondasi bagi model BERT. Pada arsitektur *transformer* penuh, terdapat dua komponen utama, yaitu *encoder* dan *decoder*. Namun, pada BERT hanya bagian *encoder* yang digunakan karena tujuan utamanya adalah memahami konteks kalimat secara dua arah (*bidirectional*).

Setiap encoder pada BERT terdiri dari dua komponen utama, yaitu mekanisme *multi-head self-attention* dan jaringan *feed-forward*, yang masing-masing diikuti oleh *residual connection* dan *layer normalization*

(*Add & Norm*). Komponen-komponen ini berfungsi untuk menstabilkan proses pelatihan dan memperkaya pemahaman kontekstual antar token dalam kalimat.

Bagian kanan gambar memperlihatkan dua konfigurasi utama dari model BERT, yaitu BERT *Base* dan BERT *Large*. Keduanya memiliki struktur *encoder* yang sama, tetapi berbeda pada jumlah lapisan (*layers*) dan parameter yang digunakan. BERT *Base* terdiri dari 12 lapisan *encoder* dengan ukuran *hidden layer* 768 dan 12 *attention heads*, masing-masing berisi komponen *multi-head attention* dan *feed-forward network*, sedangkan BERT *Large* memiliki 24 lapisan *encoder* dengan ukuran *hidden layer* 1024 dan 16 *attention heads*, yang memungkinkan model ini untuk memahami konteks bahasa yang lebih kompleks. Kedua model ini menggunakan tokenisasi *sub-kata* dan posisi *embedding* untuk merepresentasikan kata dan urutan posisi dalam kalimat.

Proses kerja BERT diawali dengan tokenisasi teks, di mana kalimat dipecah menjadi token, yang dapat berupa kata, sub-kata, atau bahkan karakter. Setiap token kemudian diubah menjadi representasi vektor melalui proses *embedding* yang telah dilatih sebelumnya oleh model. Vektor-vektor ini kemudian diproses oleh *encoder* secara berulang, yang terdiri dari beberapa layer *transformer* yang ditumpuk, sehingga menghasilkan representasi kontekstual yang semakin kaya dan bermakna.

Output dari BERT berupa representasi kontekstual dari setiap token. Khusus untuk tugas klasifikasi, digunakan token khusus [CLS] yang representasinya di akhir layer *encoder* mencerminkan pemahaman menyeluruh terhadap kalimat. Representasi ini kemudian dimanfaatkan untuk berbagai tugas NLP, seperti klasifikasi teks, analisis sentimen, maupun deteksi ulasan palsu, karena model mampu mengenali pola semantik yang kompleks serta makna tersirat dari suatu teks (Mudding, 2024).

2.2.8 IAA (*Inter-Annotator Agreement*)

IAA (*Inter-Annotator Agreement*) adalah ukuran tingkat kesepakatan antara dua atau lebih *annotator* dalam memberikan label pada suatu data yang

sama. IAA sangat penting untuk menjamin bahwa proses anotasi tidak hanya bergantung pada pendapat pribadi, tapi konsisten dan dipercaya sebagai dasar dalam penelitian maupun pengembangan sistem berbasis data (Stefanovitch and Piskorski, 2023). Tanpa adanya kesepakatan yang baik antar-annotator, data yang dihasilkan dapat menimbulkan bias sehingga memengaruhi keandalan model yang dibangun.

Beberapa metrik digunakan untuk menghitung IAA, di antaranya:

1. *Cohen's Kappa*

Digunakan untuk mengukur kesepakatan antara 2 *annotator*.

Rumusnya:

$$k = \frac{P_0 - P_e}{1 - P_e} \quad (2)$$

Dengan:

- P_0 = proporsi kesepakatan aktual antara kedua *annotator*
- P_e = proporsi kesepakatan yang diharapkan terjadi secara kebetulan

2. *Fleiss' Kappa*

Digunakan ketika terdapat lebih dari dua *annotator*. *Fleiss' Kappa* merupakan generalisasi *Cohen's Kappa* untuk banyak *annotator*.

Rumusnya:

$$k = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e} \quad (3)$$

Dengan:

- \bar{P} = rata-rata kesepakatan aktual per item
- \bar{P}_e = kesepakatan yang diharapkan secara kebetulan

Kappa menjadi metrik populer karena memperhitungkan kemungkinan kesepakatan yang terjadi secara kebetulan, sehingga lebih akurat dibandingkan sekadar menghitung persentase kesepakatan. Nilai *kappa* yang rendah menandakan kesepakatan kurang konsisten, yang dapat memengaruhi keandalan model yang dibangun.

Nilai *kappa* umumnya diinterpretasikan untuk menunjukkan tingkat

kesepakatan antar-annotator. Landis and Koch, (1977) memberikan panduan interpretasi yang masih banyak digunakan hingga saat ini:

Nilai <i>Kappa</i>	Tingkat Kesepakatan
< 0.00	Poor Agreement
0.00 – 0.20	Slight Agreement
0.21 – 0.40	Fair Agreement
0.41 – 0.60	Moderate Agreement
0.61 – 0.80	Substantial Agreement
0.81 – 1.00	Almost Perfect Agreement

Dalam penelitian terbaru, Fatima dkk., (2023) menegaskan bahwa nilai *kappa* harus dipahami secara kontekstual. Misalnya, pada data yang memiliki distribusi kelas yang tidak seimbang, nilai *kappa* cenderung menurun meskipun persentase kesepakatan antar-annotator sebenarnya tinggi. Oleh karena itu, interpretasi *kappa* harus mempertimbangkan lebih dari tabel di atas, itu harus mempertimbangkan karakteristik data, tujuan penelitian, dan jenis anotasi yang digunakan.

2.2.9 Evaluasi *confusion matrix*

Confusion matrix adalah salah satu metode evaluasi kinerja model klasifikasi dalam *machine learning* yang sangat umum digunakan, khususnya pada klasifikasi biner seperti pendeteksian *review* palsu (*fake review detection*). *Confusion matrix* menyajikan perbandingan antara hasil prediksi oleh model dengan nilai label yang sebenarnya (*ground truth*), dan membantu dalam menghitung berbagai metrik evaluasi seperti *Accuracy*, *Precision*, *Recall*, dan *F1-score* (Vidya Chandradev et al., 2023).

Dalam implementasinya, baik hasil prediksi maupun label sebenarnya direpresentasikan dalam bentuk vektor, yaitu sekumpulan nilai numerik yang menunjukkan kelas atau probabilitas setiap kategori. Vektor prediksi ini kemudian dibandingkan dengan vektor label sebenarnya untuk menghitung

jumlah *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)* yang menjadi dasar pembentukan *confusion matrix*. Representasi berbentuk vektor ini memungkinkan perbandingan hasil klasifikasi dilakukan secara matematis dan objektif.

1. Accuracy

Digunakan untuk menentukan seberapa banyak prediksi model yang benar dibandingkan dengan jumlah data secara keseluruhan. Nilai ini menunjukkan proporsi data yang diklasifikasikan dengan benar, baik yang termasuk kelas positif maupun negatif. Namun, pada dataset yang tidak seimbang (misalnya, jumlah ulasan asli jauh lebih banyak dari ulasan palsu), akurasi saja tidak cukup untuk menilai kinerja model karena nilai akurasi yang tinggi belum tentu mencerminkan performa sebenarnya. Model bisa tampak 'baik' hanya karena sering menebak kelas mayoritas.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Dimana:

TP (*True Positive*): ulasan palsu yang diprediksi secara benar.

TN (*True Negative*): ulasan asli yang diprediksi secara benar.

FP (*False Positive*): ulasan asli yang salah diprediksi sebagai palsu.

FN (*False Negative*): ulasan palsu yang salah diprediksi sebagai asli.

2. Precision

Digunakan untuk mengukur "akurasi" model dalam memprediksi kelas positif (ulasan palsu). Artinya, berapa banyak ulasan yang diidentifikasi model sebagai ulasan palsu yang sebenarnya memang palsu. Angka presisi yang tinggi berarti bahwa model tersebut jarang salah mengklasifikasikan ulasan asli sebagai ulasan palsu (*False Positive* rendah). Dalam konteks deteksi ulasan palsu, *precision* yang tinggi berarti hasil model dapat dipercaya untuk menandai ulasan yang benar-benar palsu.

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

Dimana:

TP (*True Positive*): ulasan palsu yang diprediksi secara benar.

FP (*False Positive*): ulasan asli yang salah diprediksi sebagai palsu

3. Recall

Digunakan untuk mengevaluasi kemampuan model untuk mengidentifikasi semua data positif (ulasan palsu) yang sebenarnya ada. Dengan kata lain, *recall* menunjukkan seberapa banyak ulasan palsu yang berhasil dideteksi dari seluruh ulasan palsu yang ada. Nilai *recall* yang tinggi menandakan bahwa model tidak banyak melewatkan ulasan palsu (*False Negative* rendah). Dalam aplikasi nyata, *recall* penting ketika tujuan utama adalah untuk tidak melewatkan kasus positif (seperti mendeteksi semua ulasan palsu agar tidak menyesatkan konsumen).

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

Dimana:

TP (*True Positive*): ulasan palsu yang diprediksi secara benar.

FN (*False Negative*): ulasan palsu yang salah diprediksi sebagai asli.

4. F1-score

Kombinasi *precision* dan *recall* membuat kinerja model lebih seimbang, terutama dalam kasus ketidakseimbangan kelas. (misalnya, lebih banyak ulasan asli dibandingkan palsu). Jika hanya *precision* atau *recall* yang tinggi saja, *F1-score* tidak akan maksimal. Nilai ini sangat berguna ketika terdapat ketidakseimbangan antara kelas (misalnya, jumlah ulasan asli lebih banyak dibandingkan ulasan palsu).

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

Penelitian oleh Chicco & Jurman (2020) menunjukkan bahwa *F1-score*

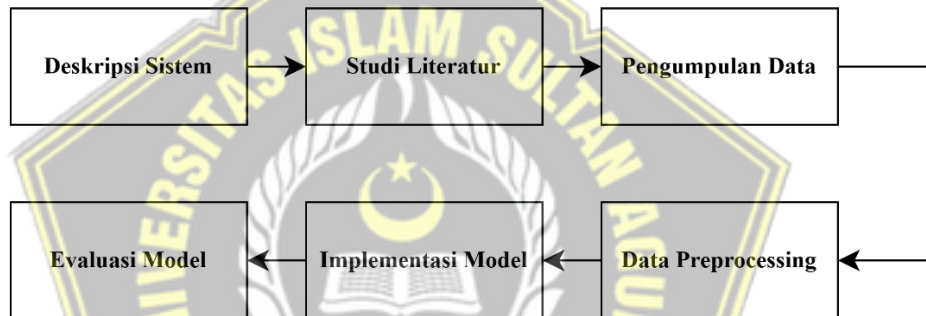
sering kali menjadi metrik terbaik dalam evaluasi model klasifikasi biner seperti deteksi ulasan palsu, karena memberikan informasi yang lebih holistik dibandingkan dengan hanya mengandalkan akurasi semata.



BAB III METODE PENELITIAN

3.1 Metode Penelitian

Dalam penelitian ini metode yang digunakan adalah LLM (*Large Language Models*) dengan menggunakan model BERT (*Bidirectional Encoder Representations from Transformers*), dimana Model ini akan mengidentifikasi ulasan yang palsu dan yang asli. Gambar berikut menunjukkan langkah-langkah yang harus dilakukan dalam penelitian ini:



Gambar 3.1 Alur Metode Penelitian

Pada gambar 3.1 adalah tampilan diagram alur tahap penelitian dalam pengembangan model LLM berbasis BERT untuk mendeteksi ulasan. Berikut adalah penjelasan setiap tahap yang terdapat pada diagram:

1. Deskripsi Sistem

Tahap awal bertujuan untuk merumuskan permasalahan, tujuan penelitian, serta batasan sistem yang akan dikembangkan. Dalam tahap ini akan dilakukan analisis kebutuhan sistem dan indentifikasi fitur-fitur yang relevan dengan topik penelitian ini.

2. Studi Literatur

Pada tahap ini dilakukan pengumpulan informasi melalui studi pustaka dari jurnal, artikel ilmiah, dan penelitian terdahulu yang relevan dengan topik. Tujuannya adalah untuk memahami konsep dasar,

algoritma, dan pendekatan yang umum digunakan, khususnya dalam pemrosesan bahasa alami, model BERT, dan metode deteksi ulasan palsu.

3. Pengumpulan Data

Tahap ini bertujuan memperoleh data ulasan produk dari Tiktok Shop by Tokopedia pada kategori *fashion*. Setiap ulasan mencakup teks dan label asli atau palsu yang ditentukan secara manual oleh *annotator* berdasarkan kriteria tertentu seperti kejelasan isi, kesesuaian dengan produk, dan penggunaan bahasa alami. Data yang terkumpul disimpan dalam format excel untuk digunakan pada tahap berikutnya.

4. Data Preprocessing

Data ulasan yang sudah diperoleh dan dilabeli selanjutnya akan diproses agar siap digunakan dalam pelatihan model. Dalam proses ini akan meliputi *cleaning* (menghapus spasi dan duplikasi), *lowercasing*, *remove* emoji dan karakter tidak relevan, serta pembagian data menjadi data latih dan data uji. Proses ini sangat penting untuk meningkatkan kualitas input yang digunakan oleh model.

5. Implementasi Model

Model dikembangkan menggunakan algoritma yang telah dipilih berdasarkan studi literatur yaitu BERT untuk mendeteksi ulasan palsu. Prosesnya meliputi *tokenization*, *padding*, dan *attention masking* untuk menyiapkan data *input*, kemudian *embedding* untuk mengubah token menjadi representasi vektor kontekstual. Selanjutnya dilakukan *fine-tuning* model menggunakan data berlabel agar mampu membedakan ulasan asli dan palsu. Hasil akhirnya berupa model terlatih yang siap dievaluasi.

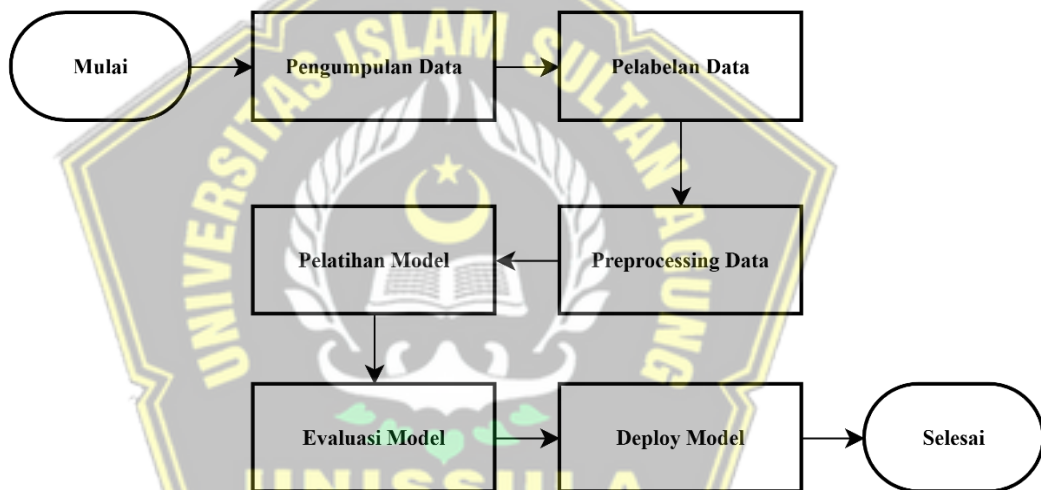
6. Evaluasi Model

Setelah model diimplementasikan, dilakukan evaluasi untuk mengukur kinerja model menggunakan metrik evaluasi (*confusion matrix*) seperti akurasi, *precision*, *recall*, dan *F1-score*. Evaluasi ini bertujuan untuk memastikan bahwa model yang dikembangkan dapat memberikan

hasil yang optimal, mampu membedakan ulasan asli dan palsu secara akurat dan dapat digunakan sebagai dasar pengembangan sistem deteksi ulasan pada platform *e-commerce*.

3.1.1 Deskripsi sistem

Sistem yang dikembangkan bertujuan untuk mendeteksi ulasan palsu pada aplikasi Tiktok Shop by Tokopedia dengan menggunakan model LLM berbasis BERT. Sistem ini dirancang untuk membantu konsumen atau pemilik toko di Tiktok Shop by Tokopedia dalam mengidentifikasi keaslian ulasan pada produk *fashion* secara otomatis dan akurat.



Gambar 3.2 *Flowchart* Kerja Model

Gambar 3.2 adalah *flowchart* untuk proses kerja model yang akan digunakan dalam penelitian ini. Proses dimulai dari pengumpulan data hingga *deploy* model yang diimplementasikan ke dalam *website*. Setiap tahap saling terhubung untuk membentuk sistem deteksi ulasan palsu secara menyeluruh. Berikut penjelasan dari setiap tahap *flowchart* diatas:

1. Pengumpulan Data: Data ulasan produk *fashion* di Tiktok Shop by Tokopedia akan dikumpulkan menggunakan teknik *web scraping*. Tujuannya adalah memperoleh dataset yang representatif untuk pelatihan model.

2. Pelabelan Data: Setelah data dikumpulkan, setiap ulasan diberi label sesuai kategori yaitu asli (*real review*) atau palsu (*fake review*). Proses pelabelan ini dapat dilakukan secara manual oleh beberapa *annotator* menggunakan referensi data yang sudah memiliki label bawaan. Data berlabel ini akan menjadi dasar untuk melatih model klasifikasi.
3. *Preprocessing* Data: Data yang diperoleh diproses untuk menghapus kata-kata yang tidak relevan, mengubah huruf kapital menjadi huruf kecil (*lowercase*), *remove* emoji dan karakter khusus sehingga menghasilkan *clean data*. Proses ini bertujuan meningkatkan kualitas *input* dan performa model.
4. Pelatihan Model: Data yang sudah diproses akan dilatih dengan menggunakan model BERT. Pada langkah awal Teks akan diubah menjadi token, diseragamkan panjangnya melalui *padding*, dan diberi *attention mask* untuk menandai token yang relevan. Selanjutnya, token direpresentasikan ke dalam bentuk vektor melalui proses *embedding* di dalam *encoder* BERT yang memanfaatkan mekanisme *self-attention* dua arah (*bidirectional*). Model dilatih menggunakan data berlabel, untuk mengenali ciri-ciri ulasan palsu. Representasi akhir dari token [CLS] digunakan pada lapisan klasifikasi untuk menentukan apakah ulasan termasuk asli atau palsu.
5. Evaluasi Model: Model diuji menggunakan data testing untuk mengukur kinerja berdasarkan metrik akurasi, *precision*, *recall*, dan *F1-score*. Evaluasi ini memastikan model dapat membedakan ulasan asli dan palsu secara akurat.
6. *Deploy* Model: Setelah model dilatih akan dilakukan *deploy* model ke *website* menggunakan *streamlit* dengan desain antarmuka pengguna sederhana yang akan dikembangkan untuk memungkinkan pengguna untuk memverifikasi ulasan produk secara *real-time*.

Sistem ini dirancang untuk menerima input ulasan dari pengguna, kemudian memprosesnya menggunakan model yang telah dilatih untuk menampilkan hasil klasifikasi secara otomatis. Hasil analisis ditampilkan

dalam bentuk kategori ulasan asli atau palsu beserta penjelasan umum mengenai ciri-ciri ulasan palsu agar pengguna dapat memahami alasan di balik hasil prediksi. Dengan adanya fitur tersebut, pengguna dapat melakukan verifikasi keaslian ulasan secara praktis dan memperoleh informasi yang lebih transparan dalam proses pengambilan keputusan pembelian.

3.1.2 Studi literatur

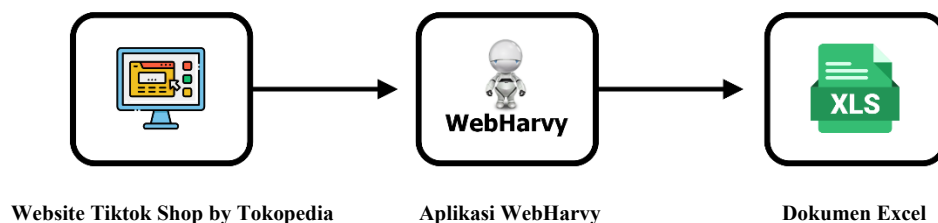
Untuk mengkaji teori dan praktik yang terkait dengan pendekatan LLM, studi ini akan meneliti sejumlah e-book, publikasi ilmiah, jurnal, tesis sebelumnya, dan situs web relevan lainnya. Studi ini bertujuan untuk memahami dasar teori, konsep, serta metode-metode yang telah digunakan dalam penelitian sebelumnya, khususnya yang berkaitan dengan pemrosesan bahasa alami (NLP) dan penerapan LLM dalam deteksi ulasan palsu.

3.1.3 Pengumpulan data

Tahap pertama dari penelitian ini adalah pengumpulan data, yang bertujuan untuk mengumpulkan ulasan produk *fashion* dari Toko TikTok di platform Tokopedia guna menganalisis dan melatih model deteksi ulasan palsu. Data dikumpulkan secara sistematis agar relevan, representatif, dan siap digunakan pada tahap *labeling* serta pemodelan berikutnya. Proses pengumpulan data melalui beberapa tahapan sebagai berikut:

1. Pengumpulan Data

Pada tahap ini, data dikumpulkan dengan menggunakan metode *web scraping* melalui halaman *website* Tiktok Shop by Tokopedia dengan menggunakan aplikasi *WebHarvy*.



Gambar 3.3 Alur Pengumpulan Data

Dalam penelitian ini, terdapat beberapa keterbatasan terkait

pengambilan data ulasan dari platform Tiktok Shop by Tokopedia, antara lain:

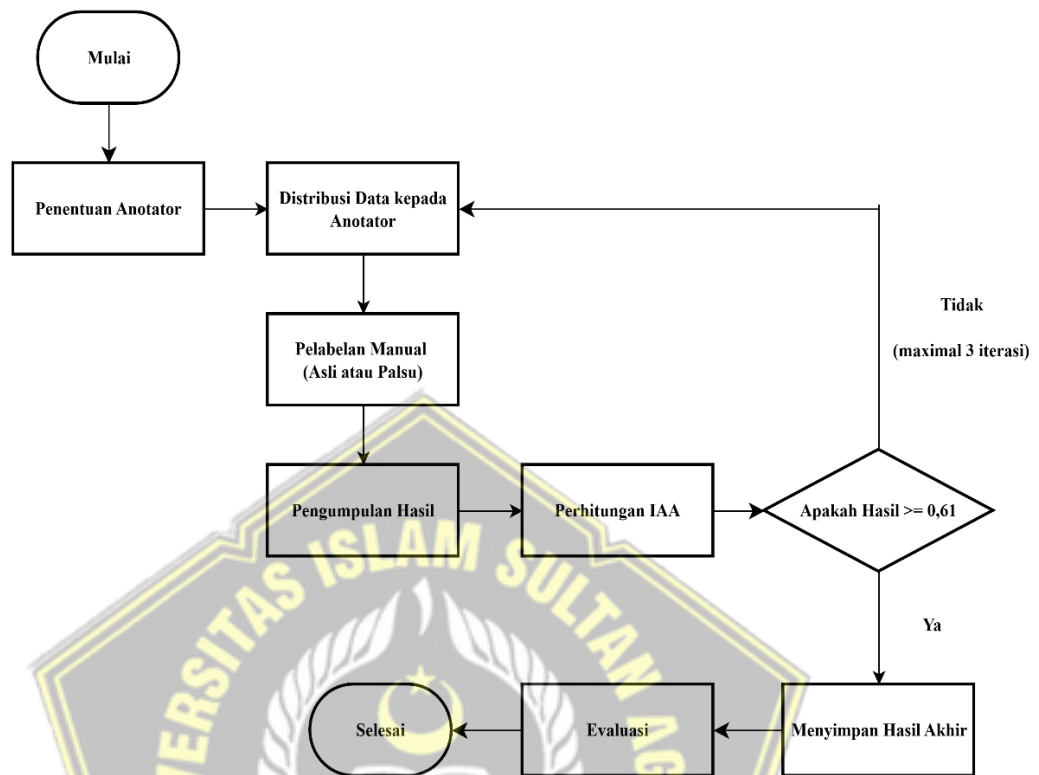
1. Setiap produk hanya diambil 5 ulasan teratas, sesuai dengan rekomendasi yang diberikan oleh platform.
2. Identitas pengguna (*username*) tidak tersedia, sehingga seluruh data ulasan bersifat anonim.
3. Urutan ulasan yang diambil bersifat acak berdasarkan relevansi, termasuk tanggal publikasi dan rating.
4. Kualitas ulasan bervariasi, dapat berupa seluruh ulasan positif, seluruh ulasan negatif, atau kombinasi dari keduanya.
5. Tidak ada atribut pendukung lain yang dikumpulkan, data yang digunakan terbatas pada teks ulasan itu sendiri.

Tahapan pengumpulan data:

1. Mengidentifikasi halaman produk *fashion* yang akan di-*scrape*.
2. Mengambil teks ulasan dari setiap produk yang tersedia (5 ulasan teratas per produk).
3. Menyimpan hasil *scraping* dalam format XLSX untuk memudahkan pemrosesan selanjutnya.

2. Proses *Labeling*

Pelabelan data merupakan tahap penting dalam penelitian ini karena menentukan kualitas dataset yang digunakan pada tahap pemodelan. Proses pelabelan dilakukan secara manual terhadap data ulasan yang diperoleh dari Tiktok Shop by Tokopedia. Alur pelabelan data disusun secara terstruktur dan divisualisasikan dalam bentuk *flowchart* sebagaimana ditunjukkan pada gambar berikut.



Gambar 3. 4 *Flowchart* Pelabelan Data

Dari gambar 3.4 menunjukkan alur proses pelabelan data, dalam proses ini pelabelan dimulai dari penentuan *annotator*. Disini ditentukan pelabelan akan dilakukan oleh 5 orang *annotator* (pemberi label) yang dipilih berdasarkan kriteria tertentu. Pemilihan *annotator* dilakukan dengan mempertimbangkan tingkat pemahaman terhadap produk *fashion* serta pengalaman berbelanja di platform Tiktok Shop by Tokopedia agar dapat memberikan penilaian yang objektif. Selain itu, para *annotator* berasal dari latar belakang yang beragam untuk memperluas sudut pandang dalam proses penentuan label. Keberagaman usia serta peran mereka, baik sebagai penjual, pengelola toko maupun pembeli aktif, memberikan sudut pandang yang seimbang dalam menilai keaslian ulasan produk *fashion*. Variasi pengalaman tersebut berkontribusi dalam meningkatkan reliabilitas proses pelabelan.

Setelah *annotator* ditentukan, dilakukan distribusi data kepada *annotator*. Pada tahap ini, data ulasan yang telah dikumpulkan sebelumnya dibagikan kepada seluruh *annotator* untuk diberi label secara independen. Setiap *annotator* memberikan penilaian tanpa dipengaruhi oleh *annotator* lain guna menjaga objektivitas. Tahap selanjutnya adalah pelabelan manual, di mana setiap ulasan diberi label ke dalam dua kategori, yaitu *Asli* dan *Palsu*. Label *Asli* diberikan pada ulasan yang mencerminkan pengalaman nyata pengguna terhadap produk, sedangkan label *Palsu* diberikan pada ulasan yang tidak sesuai dengan produk atau menunjukkan indikasi manipulasi. Dalam proses ini, *annotator* juga memberikan label dengan menambahkan karakteristik penilaian berdasarkan pemahaman mereka terhadap konteks ulasan dan produk, serta mengacu pada pedoman anotasi yang telah ditetapkan. Pendekatan ini bertujuan untuk menjaga konsistensi penilaian tanpa bergantung pada pendapat pribadi. Hasil pelabelan dari seluruh *annotator* kemudian masuk ke tahap pengumpulan hasil, yaitu proses penggabungan seluruh label yang diberikan *annotator* ke dalam satu dataset. Dataset ini menjadi dasar untuk mengukur tingkat kesepakatan antar-*annotator*.

Untuk validasi label dilakukan setelah seluruh hasil anotasi dari lima *annotator* direkapitulasi ke dalam tabel excel. Tahap ini bertujuan untuk memastikan bahwa label yang dihasilkan layak digunakan sebagai label akhir dalam penelitian. Proses validasi diawali dengan pengecekan kelengkapan data, yaitu memastikan setiap ulasan memperoleh label dari seluruh *annotator* tanpa adanya data kosong. Selanjutnya, pengecekan konsistensi dilakukan pada tahap rekapitulasi data untuk memastikan tidak terjadi kesalahan pencatatan label dari lembar anotasi ke dalam tabel excel.

Penetapan label akhir untuk setiap ulasan dilakukan berdasarkan prinsip keputusan mayoritas. Suatu label dianggap layak digunakan apabila memperoleh persetujuan minimal tiga dari lima *annotator*. Apabila mayoritas *annotator* memberikan label *Asli*, maka ulasan tersebut ditetapkan sebagai ulasan *Asli*, demikian pula sebaliknya untuk label

Palsu. Selain itu, alasan singkat yang diberikan oleh annotator turut diperiksa untuk memastikan bahwa proses penilaian dilakukan sesuai dengan pedoman anotasi yang telah ditetapkan oleh peneliti.

Melalui tahapan validasi ini, label akhir yang digunakan dalam penelitian merupakan hasil kesepakatan kolektif *annotator* yang diperoleh melalui prosedur yang terstruktur dan objektif. Dengan demikian, data berlabel yang dihasilkan memiliki tingkat keandalan yang memadai untuk digunakan pada tahap analisis dan pemodelan selanjutnya.

Selanjutnya, untuk mengukur tingkat konsistensi antar-*annotator*, dilakukan perhitungan *Inter-Annotator Agreement* (IAA) menggunakan metode *fleiss' kappa*. Metode ini dipilih karena sesuai untuk data yang dilabeli oleh lebih dari 2 *annotator*. Proses penghitungan *fleiss' kappa* dimulai dengan membaca dataset yang telah diberi label dari file excel. Selanjutnya, jumlah kesepakatan untuk masing-masing kategori yaitu Asli dan Palsu, dihitung untuk setiap item ulasan. Dari perhitungan ini diperoleh proporsi kesepakatan per item serta nilai *Fleiss' Kappa* secara keseluruhan, yang mencerminkan tingkat konsistensi antar-*annotator*.

Hasil perhitungan *fleiss' kappa* kemudian masuk ke tahap evaluasi. Pada tahap ini, nilai *kappa* yang diperoleh ditafsirkan dengan membandingkannya terhadap ambang batas yang telah ditentukan, yaitu $\geq 0,61$, yang menunjukkan tingkat kesepakatan antar-*annotator* berada pada kategori baik (*substantial agreement*). Berdasarkan hasil evaluasi tersebut, dilakukan pengambilan keputusan apakah dataset hasil anotasi layak digunakan untuk tahap pemodelan. Jika nilai *fleiss' kappa* memenuhi ambang batas, maka data dianggap reliabel dan dilanjutkan ke tahap penyimpanan hasil akhir. Dataset akhir ini selanjutnya digunakan sebagai data masukan pada proses pemodelan sistem klasifikasi ulasan.

3.1.4 Data preprocessing

Tahap *data preprocessing* dilakukan setelah mendapatkan data ulasan yang diperoleh dengan cara web scraping dan diberi label oleh para *annotator*. Tujuan dari langkah ini adalah membersihkan, menstandarisasi,

dan mempersiapkan data untuk pelatihan model. Agar model dapat mempelajari pola dengan lebih sukses, langkah ini meningkatkan kualitas input. Beberapa prosedur pra-pemrosesan meliputi:

1. *Preprocessing* Data

Setelah data diperoleh melalui *web scraping* dan dilakukan *labeling* oleh beberapa *annotator* langkah berikutnya adalah melakukan *preprocessing* agar data siap digunakan dalam pelatihan model. *Preprocessing* bertujuan untuk membersihkan dan menstandarkan teks sehingga model dapat mempelajari pola dengan lebih efektif. Tahapan *preprocessing* meliputi:

a. *Remove* Emoji

Menghapus emoji atau simbol non-teks dari ulasan agar data yang digunakan hanya berisi karakter yang bermakna bagi analisis teks.

b. *Remove* Karakter Khusus

Menghapus tanda baca, angka, atau simbol yang tidak relevan seperti '!', '?', '#', dan sebagainya untuk menjaga kebersihan data teks.

c. *Lowercase*

Untuk memastikan keseragaman dan mencegah perbedaan yang disebabkan oleh penggunaan huruf kapital, ubah semua huruf menjadi huruf kecil.

d. Hasil *Preprocessing*

Setelah melalui seluruh tahapan di atas, diperoleh data teks yang telah bersih (*clean data*) dan siap digunakan untuk pelatihan model BERT.

2. Pembagian Data

Data pelatihan (*training set*), data validasi (*validation set*), dan data uji (*test set*) adalah tiga komponen dari dataset.

a. Data Pelatihan (64%),

Model tersebut akan dilatih menggunakan data ini. Data ini akan mengajarkan model untuk mengidentifikasi hubungan dan pola di dalamnya.

b. Data Validasi (16%),

Data ini digunakan untuk menyempurnakan model selama proses pelatihan. Fungsinya untuk mengevaluasi kinerja model pada data yang belum pernah dilihat sebelumnya.

c. Data Pengujian (20%),

Data ini akan digunakan setelah model selesai dilatih dan disempurnakan. Fungsinya sebagai pengukur kinerja model secara objektif pada data yang benar-benar baru, sehingga memberikan gambaran yang akurat mengenai kemampuan model dalam bekerja.

Pembagian data dilakukan secara acak untuk memastikan distribusi label (ulasan asli dan palsu) seimbang antara data pelatihan dan pengujian. Langkah ini penting agar model tidak bias terhadap salah satu kategori dan mampu melakukan generalisasi dengan baik pada data baru.

3.1.5 Implementasi model

Sebelum model dilatih, data ulasan yang telah dikumpulkan dan melalui tahap preprocessing pada subbab sebelumnya digunakan sebagai input utama. Dataset berupa ulasan produk *fashion* yang telah diberi label ‘Asli’ dan ‘Palsu’ oleh *annotator*, kemudian dibagi menjadi 80% data pelatihan dan 20% data pengujian secara *stratified* agar proporsi label tetap seimbang.

Setelah data siap digunakan, tahap berikutnya adalah *tokenization* dan membangun model yang akan digunakan untuk proses klasifikasi. Model yang diterapkan dalam penelitian ini adalah BERT, yang dikenal efektif dalam memahami konteks semantik teks melalui mekanisme *transformer*. Model ini diadaptasi untuk melakukan klasifikasi biner, yaitu membedakan antara ulasan asli dan ulasan palsu.

1. Arsitektur Model

Model yang digunakan adalah BERT *Base*. BERT dipilih karena kemampuannya dalam memahami konteks semantik teks secara mendalam melalui mekanisme *self-attention* dua arah (*bidirectional*), yang membuatnya efektif dalam mendeteksi makna antar kata.

a. Spesifikasi BERT *Base*

Secara umum, spesifikasi arsitektur BERT *Base* adalah sebagai berikut:

- Jumlah *encoder layer (transformer block)*: 12 *layer*
- Ukuran *hidden state*: 768
- Jumlah *attention head*: 12
- Ukuran *feed-forward network*: 3072
- Total parameter: ±110 juta

Setiap *encoder layer* di dalam model berfungsi untuk mengekstraksi fitur teks melalui kombinasi *multi-head self-attention* dan *feed-forward neural network*. Kombinasi dari 12 *layer* ini memungkinkan BERT untuk menangkap hubungan semantik dan sintaksis antar kata dengan sangat baik, sehingga menjadi alasan utama pemilihan model ini dalam penelitian deteksi ulasan palsu.

b. Representasi Input

Sebelum diproses oleh model, setiap teks ulasan diubah menjadi token menggunakan *WordPiece tokenizer*. *Tokenization* merupakan proses pemecahan teks menjadi unit-unit kecil yang disebut token, yang bisa berupa kata, sub-kata, atau simbol. Dalam konteks BERT, proses ini memungkinkan model untuk memahami kata-kata yang jarang muncul dengan memecahnya menjadi bagian yang lebih kecil (*subwords*).

Proses tokenisasi ini juga melibatkan beberapa token khusus, yaitu:

- [CLS] ditempatkan di awal teks sebagai representasi global dari keseluruhan ulasan.
- [SEP] ditempatkan di akhir teks sebagai penanda akhir kalimat atau pemisah antar segmen teks.

Selanjutnya, setiap token direpresentasikan dalam bentuk vektor numerik melalui proses *embedding*. *Embedding* merupakan teknik untuk mengubah token menjadi representasi vektor berdimensi tinggi sehingga dapat diproses oleh model.

Dalam BERT, terdapat tiga jenis *embedding* yang digunakan:

- *Token Embedding*: merepresentasikan identitas kata atau sub-kata.
- *Segment Embedding*: menandai bagian teks, terutama jika input terdiri dari dua segmen kalimat.
- *Positional Embedding*: memberikan informasi posisi kata dalam kalimat agar model dapat memahami urutan kata secara kontekstual.

c. Proses *Encoding*

Token *embedding* yang telah terbentuk kemudian diproses melalui 12 *encoder layer* pada arsitektur BERT. Setiap *encoder layer* terdiri atas beberapa komponen utama sebagai berikut:

- *Multi-Head Self-Attention* (MHSA)
Berfungsi untuk mempelajari hubungan kontekstual antar kata dalam satu kalimat.
- *Add & Norm* (Normalisasi Pertama)
Tahapan ini menambahkan (*residual connection*) hasil dari proses *self-attention* ke input awal, kemudian dilakukan *layer normalization* untuk menjaga stabilitas distribusi nilai selama pelatihan.
- *Feed-Forward Network* (FFN)
Merupakan jaringan saraf dua lapis yang bertugas memperkuat dan mentransformasikan representasi vektor hasil *self-attention*, sehingga model mampu menangkap hubungan *non-linear* antar fitur teks.
- *Add & Norm* (Normalisasi Kedua)
Setelah melewati FFN, hasilnya kembali dinormalisasi untuk memastikan konsistensi dan stabilitas informasi antar layer.

d. Klasifikasi

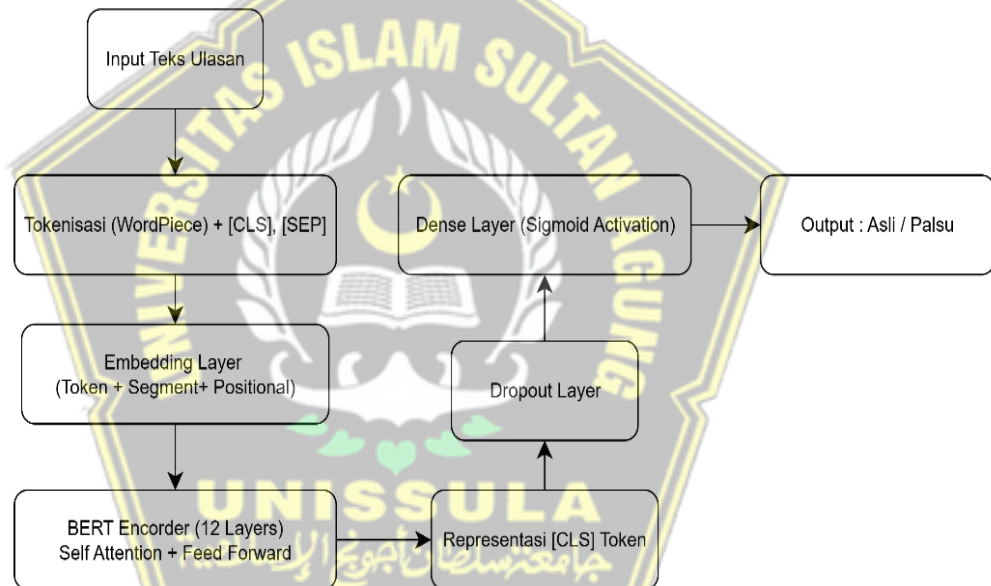
Setelah melalui proses *encoding*, hasil representasi dari setiap token telah memuat informasi konteks dan makna kata dalam kalimat. Dari hasil ini, token khusus [CLS] yang ditempatkan di awal teks digunakan sebagai representasi keseluruhan ulasan. Representasi ini kemudian dilewatkan ke lapisan *dropout* untuk mencegah *overfitting*,

dilanjutkan ke *dense layer* yang berfungsi sebagai lapisan klasifikasi. Selanjutnya, *fully connected layer* mengubah vektor [CLS] menjadi skor probabilitas. Pada lapisan keluaran (*output layer*) menggunakan fungsi aktivasi *sigmoid* untuk klasifikasi biner:

- Label 0 mengindikasikan Ulasan Asli
- Label 1 mengindikasikan Ulasan Palsu

Dengan mekanisme ini, model BERT mampu melakukan klasifikasi biner secara efektif berdasarkan pemahaman konteks dan makna dari setiap teks ulasan.

e. Alur Arsitektur Model BERT



Gambar 3.4 Alur Model BERT Untuk Klasifikasi Ulasan Palsu

Gambar 3.4 menjelaskan alur model BERT untuk klasifikasi ulasan palsu yang dimulai dari *input teks* ulasan yang diubah menjadi token menggunakan *WordPiece tokenizer*, di mana setiap teks ditambahkan token khusus [CLS] dan [SEP]. Selanjutnya, token tersebut dimasukkan ke dalam *embedding layer* yang menggabungkan *token embedding*, *segment embedding*, dan *positional embedding* agar model dapat memahami urutan kata dalam kalimat.

Selanjutnya, *embedding* diproses oleh BERT *Encoder* yang terdiri dari 12 *encoder layer* dengan mekanisme *multi-head self-attention* dan *feed-forward neural network* untuk menangkap konteks semantik secara dua arah (*bidirectional*). Hasil representasi dari token [CLS] digunakan sebagai ringkasan kalimat, lalu dilewatkan ke *dropout layer* untuk mencegah *overfitting*, dan diteruskan ke *dense layer* dengan aktivasi *sigmoid* guna menghasilkan probabilitas klasifikasi biner, yaitu ulasan asli atau palsu.

2. Pelatihan Model

Proses pelatihan dilakukan setelah tahap *preprocessing* dan pembentukan arsitektur model selesai. Model BERT yang digunakan dilatih menggunakan *framework deep learning* populer seperti *PyTorch* dengan memanfaatkan modul *Transformers* dari *Hugging Face*. Pelatihan ini dilakukan dengan *batch processing* dan menggunakan CPU maupun GPU untuk mempercepat proses. Tujuan utama pelatihan ini adalah agar model mampu membedakan ulasan asli dan ulasan palsu berdasarkan pola bahasa.

Beberapa hal penting dalam proses pelatihan dijelaskan sebagai berikut:

a. Konfigurasi Parameter Pelatihan

Parameter utama yang dioptimalkan meliputi:

- *Learning rate*, digunakan untuk mengatur kecepatan pembaruan bobot selama proses pelatihan. Nilainya ditetapkan pada tingkat awal tertentu dan disesuaikan menggunakan *scheduler* (misalnya *warm-up* di awal pelatihan dan *linear decay* setelahnya) agar proses konvergensi model tetap stabil.
- *Batch size*, Jumlah sampel data yang ditangani dalam satu siklus pelatihan ditentukan oleh ukuran batch. Untuk mencapai kompromi antara efisiensi komputasi dan stabilitas pembaruan bobot model, nilai ukuran batch dihitung berdasarkan kemampuan perangkat keras (CPU/GPU).

- *Dropout rate*, digunakan pada beberapa lapisan jaringan untuk mengurangi kemungkinan *overfitting* karena sebagian neuron dinonaktifkan secara acak selama proses pelatihan.

b. *Optimizer* dan Fungsi *Loss*

Model dioptimalkan menggunakan algoritma AdamW, yang merupakan pengembangan dari Adam dengan penambahan *weight decay* sehingga lebih sesuai untuk model berbasis *transformer*. Fungsi *loss* yang digunakan adalah *binary cross-entropy loss*, karena tugas yang dilakukan berupa klasifikasi biner (asli vs palsu).

c. Strategi Pelatihan

Proses pelatihan model dilakukan dalam beberapa *epoch*, di mana setiap *epoch* mencakup iterasi penuh terhadap seluruh data pelatihan. Tujuannya adalah agar model dapat mempelajari pola bahasa yang membedakan antara ulasan asli dan ulasan palsu secara bertahap. Untuk mencegah pelatihan yang berlangsung terlalu lama atau berisiko mengalami *overfitting*, teknik *early stopping* akan digunakan, teknik ini untuk penghentian otomatis ketika kinerja model pada data validasi tidak meningkat dalam beberapa waktu berturut-turut.

Selain itu, untuk memastikan model dapat belajar secara efektif dan memiliki kemampuan generalisasi yang tepat terhadap data baru, metrik evaluasi seperti akurasi dan nilai *loss* pada *training set* dan *validation set* dipantau selama proses pelatihan.

d. Evaluasi Selama Pelatihan

Kemampuan model untuk melakukan generalisasi pada data yang belum pernah dilihat sebelumnya dinilai menggunakan data validasi pada akhir setiap *epoch*. Tujuan dari penilaian ini adalah untuk memastikan model dapat secara akurat mengidentifikasi pola-pola baru selain menghafal pola-pola dari data pelatihan (*overfitting*).

Selama proses evaluasi, metrik seperti akurasi, *precision*, *recall*, dan *F1-score* dapat digunakan untuk mengukur performa model secara menyeluruh. Selain itu, nilai *loss* pada data validasi dibandingkan

dengan *loss* pada data pelatihan guna memantau kestabilan proses pembelajaran dan mendeteksi adanya perbedaan signifikan yang dapat mengindikasikan *overfitting*.

3. Evaluasi Model

Evaluasi model akan dilakukan setelah proses pelatihan selesai dengan menggunakan data uji (*testing set*). Tujuan evaluasi ini adalah untuk mengukur performa model secara objektif dalam membedakan antara ulasan asli dan ulasan palsu.

Tahapan evaluasi meliputi:

a. Pengujian Model

Pada tahap ini, dataset uji dimasukkan ke dalam model BERT yang telah melalui proses pelatihan. Model kemudian menghasilkan prediksi terhadap setiap ulasan, dan hasil prediksi tersebut dibandingkan dengan label asli (*ground truth*) untuk mengukur tingkat akurasi dan performa klasifikasi. Proses pengujian bertujuan untuk mengukur performa generalisasi model, yaitu kemampuan model dalam mengenali pola baru tanpa bergantung pada data pelatihan.

b. Metrik Evaluasi

Evaluasi model dilakukan menggunakan *confusion matrix* yang mencakup empat metrik utama, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. *Confusion matrix* merupakan representasi tabular yang menampilkan distribusi hasil prediksi model ke dalam empat kategori, yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*. Berdasarkan nilai-nilai tersebut, metrik berikut dihitung:

1. Akurasi, yaitu persentase jumlah prediksi yang benar terhadap keseluruhan data uji.
2. *Precision*, yaitu persentase ulasan yang diprediksi sebagai palsu dan benar-benar palsu.
3. *Recall*, yaitu persentase ulasan palsu yang berhasil dideteksi dengan benar oleh model.

4. *F1-score*, yaitu rata-rata harmonis dari *precision* dan *recall* yang relevan terutama pada kondisi data tidak seimbang.

c. Interpretasi Hasil Evaluasi

Nilai metrik yang tinggi menunjukkan kemampuan model dalam mengklasifikasikan ulasan asli dan palsu secara akurat. Sementara itu, *confusion matrix* digunakan untuk memberikan analisis lebih rinci terkait kesalahan prediksi, misalnya kecenderungan model salah mengklasifikasikan ulasan asli sebagai palsu atau sebaliknya. Analisis ini membantu memahami secara menyeluruh kinerja model dan menemukan area yang perlu ditingkatkan.

Selain itu, selama proses pelatihan juga dilakukan pemantauan menggunakan data validasi untuk mencegah *overfitting*. Meskipun demikian, hasil evaluasi pada data uji tetap dijadikan acuan utama untuk menilai kinerja akhir model secara objektif.

4. Deploy Model

Tahap akhir dalam penelitian ini adalah melakukan *deployment* model ke dalam sebuah aplikasi berbasis *web*. Tujuan dari tahap ini adalah agar model yang telah dilatih dapat digunakan secara langsung oleh pengguna akhir. Dengan adanya proses *deployment*, model tidak hanya berhenti pada tahap pengujian di lingkungan pengembangan, tetapi juga dapat memberikan manfaat praktis melalui sistem yang mudah diakses dan interaktif.

a. Platform *Deploy*

Model diimplementasikan menggunakan *Streamlit*, sebuah *framework open-source* berbasis *Python* yang mendukung pembuatan aplikasi *machine learning* secara interaktif dan sederhana. *Streamlit* berfungsi sebagai media integrasi antara model *machine learning* dengan antarmuka pengguna berbasis *web*. Pada tahap ini, model hasil pelatihan diintegrasikan ke dalam aplikasi interaktif agar dapat diakses secara langsung melalui peramban (*web browser*). Penggunaan *Streamlit* dipilih karena mampu menampilkan hasil prediksi secara

real-time dan memudahkan proses pengujian serta evaluasi fungsionalitas sistem setelah model diimplementasikan.

b. Fungsi Sistem

Sistem memungkinkan pengguna untuk memasukkan teks ulasan produk secara manual melalui antarmuka aplikasi. Input tersebut kemudian akan diproses oleh model BERT yang sudah dilatih sebelumnya untuk melakukan klasifikasi ulasan. Hasilnya berupa prediksi kategori termasuk ulasan asli atau ulasan palsu yang ditampilkan secara *real-time*. Dengan adanya fungsi ini, pengguna dapat dengan mudah mengetahui keaslian suatu ulasan tanpa harus melakukan analisis manual.

c. Antarmuka Pengguna

Antarmuka dirancang dengan tampilan sederhana dan interaktif, sehingga mudah digunakan baik oleh pemilik toko maupun calon pembeli di platform *e-commerce*, terutama Tiktok Shop by Tokopedia. Selain itu, sistem juga menyajikan penjelasan atau alasan pendukung berdasarkan ciri-ciri umum dari ulasan palsu, sehingga pengguna dapat memahami faktor-faktor yang memengaruhi keputusan model dalam memberikan hasil analisis.

d. Manfaat *Deployment*

Dengan adanya implementasi berbasis *web* ini, sistem deteksi ulasan palsu tidak hanya bermanfaat dalam ranah penelitian, tetapi juga dapat digunakan secara praktis sebagai alat bantu konsumen dalam pengambilan keputusan sebelum membeli produk sekaligus membantu penjual dalam memantau kredibilitas ulasan pada toko mereka. Dengan demikian, tahap *deployment* berperan penting dalam menjembatani hasil penelitian agar dapat diterapkan secara nyata dan memberikan manfaat langsung bagi pengguna.

3.1.6 Evaluasi

Kemampuan model untuk membedakan antara ulasan produk yang asli dan yang palsu dievaluasi. Tujuan dari evaluasi ini adalah untuk mengukur

seberapa baik model memprediksi data yang belum pernah dilihat sebelumnya (*testing data*). Beberapa kriteria evaluasi yang sering digunakan dalam klasifikasi biner diterapkan pada tahap ini, termasuk:

Evaluasi digunakan untuk menilai kinerja model dalam mengklasifikasikan ulasan produk menjadi ulasan asli atau ulasan palsu. Evaluasi ini bertujuan untuk mengukur sejauh mana model mampu melakukan prediksi secara akurat terhadap data yang belum pernah dilihat sebelumnya (*testing data*). Pada tahap ini digunakan beberapa metrik evaluasi yang umum dipakai dalam permasalahan klasifikasi biner, yaitu:

a. *Accuracy*

Mengukur seberapa besar proporsi prediksi yang benar dibandingkan dengan seluruh data pengujian. Nilai *accuracy* yang tinggi menunjukkan bahwa model memiliki kemampuan yang baik dalam memprediksi kedua kelas secara umum.

b. *Precision*

Menilai seberapa banyak ulasan yang diprediksi sebagai *palsu* benar-benar merupakan ulasan palsu, tanpa terlalu banyak menghasilkan kesalahan prediksi positif (*false positives*). Metrik ini penting untuk menghindari kesalahan dalam menandai ulasan asli sebagai palsu.

c. *Recall*

Mengevaluasi kemampuan model untuk menemukan seluruh ulasan palsu yang sebenarnya ada di dalam dataset. Nilai *recall* yang tinggi menunjukkan bahwa model mampu mendeteksi sebagian besar ulasan palsu dengan baik.

d. *F1-score*

Perspektif yang lebih seimbang tentang kinerja model diberikan oleh rata-rata harmonik dari presisi dan *recall*, terutama ketika data seimbang antara evaluasi asli dan evaluasi palsu.

Hasil visualisasi *confusion matrix* memperlihatkan bahwa sebagian besar prediksi berhasil dilakukan dengan benar, meskipun masih terdapat sejumlah kesalahan, khususnya pada ulasan yang bersifat ambigu. Hal ini

wajar terjadi mengingat keterbatasan jumlah data serta adanya variasi gaya bahasa pengguna dalam memberikan ulasan.

Secara keseluruhan, model yang dibangun telah menunjukkan performa yang cukup baik dan dapat dijadikan sebagai sistem pendukung dalam mendeteksi ulasan palsu pada platform *e-commerce*, khususnya Tiktok Shop by Tokopedia. Namun, masih terdapat ruang untuk peningkatan performa, misalnya dengan menambah jumlah data pelatihan, melakukan *fine-tuning* parameter yang lebih optimal, serta mempertimbangkan penggunaan arsitektur model *transformer* lain yang lebih mutakhir. Dengan pengembangan lanjutan, diharapkan model dapat menghasilkan prediksi yang lebih akurat dan *robust* terhadap berbagai variasi data ulasan.

3.2 Analisa Kebutuhan

Pada tahap ini, penulis melakukan analisis terhadap berbagai kebutuhan yang diperlukan selama proses pembangunan sistem, meliputi bahasa pemrograman, perangkat lunak, serta *library* dan *framework* pendukung. Berikut adalah rincian kebutuhan yang digunakan dalam pengembangan sistem ini:

1. Bahasa Pemrograman

- *Python*

Python merupakan bahasa pemrograman utama yang digunakan dalam pengembangan sistem ini. Bahasa ini dipilih karena memiliki sintaks yang sederhana, mudah dipahami, serta didukung oleh berbagai pustaka *deep learning* yang sangat powerful. Dengan fitur-fitur seperti kemudahan dalam pemrograman berorientasi objek, dukungan untuk pengolahan data, serta kemampuan untuk bekerja dengan data besar, *python* adalah bahasa yang sangat ideal untuk pengembangan sistem berbasis *deep learning*.

2. Perangkat Lunak (*Software*)

- *Google Colab*

Google Colab digunakan sebagai platform pengembangan dan

pelatihan model. Platform ini menyediakan lingkungan *cloud-based* dengan dukungan GPU, sehingga mempercepat proses *training* model. Selain itu, *google colab* memudahkan proses eksperimen, pengujian model, memudahkan kolaborasi dan pengelolaan kode secara daring tanpa memerlukan instalasi perangkat lunak tambahan di komputer lokal.

- *Jupyter Notebook*

Jupyter Notebook digunakan sebagai lingkungan pengembangan untuk menulis dan menjalankan kode secara interaktif. Platform ini memudahkan proses pengujian dan modifikasi kode secara langsung, serta memungkinkan penulis untuk melakukan eksplorasi data, visualisasi, dan pencatatan hasil eksperimen secara terintegrasi.

- *Anaconda*

Anaconda digunakan sebagai *environment manager* membangun aplikasi *streamlit* secara lokal. Dalam penelitian ini, *anaconda* berperan sebagai pengelola *environment* yang memastikan seluruh *library* seperti *pytorch*, *transformers*, *scikit-learn*, *imbalanced-learn*, dan *streamlit* dapat bekerja secara kompatibel. Selain itu, *Anaconda* digunakan untuk menjalankan aplikasi *web-based* berbasis *streamlit* (*app.py*) yang dibuat sebagai antarmuka model untuk mendeteksi ulasan palsu setelah model selesai dilatih di *Jupyter Notebook/google colab*.

- *Draw.io*

Draw.io digunakan sebagai alat bantu untuk membuat berbagai diagram seperti diagram alur sistem, diagram arsitektur, serta rancangan model. *Tool* ini memudahkan penyajian visual terkait desain sistem sehingga struktur dan alur kerja sistem dapat dipahami dengan lebih jelas dan terorganisir.

- *Figma*

Sebuah alat berbasis *web* bernama *Figma* digunakan untuk membuat pengalaman pengguna (*User Experience*) dan antarmuka pengguna (*User Interface*). Dalam penelitian ini, *figma* digunakan

untuk membuat desain antarmuka sistem agar tampilan aplikasi menjadi lebih interaktif, menarik, dan mudah digunakan.

3. *Library dan Framework*

- *PyTorch*

PyTorch merupakan *framework deep learning open-source* yang digunakan untuk membangun, melatih, dan mengimplementasikan model berbasis jaringan saraf tiruan. *Framework* ini dipilih karena memiliki fleksibilitas tinggi, sintaks yang mudah dipahami, serta dukungan GPU yang optimal untuk mempercepat proses pelatihan model. Dalam penelitian ini, *pytorch* digunakan sebagai dasar pengembangan model BERT untuk mendeteksi ulasan palsu, karena mendukung berbagai fungsi *library* NLP dan integrasi yang baik dengan *google colab*.

- *Transformers (HuggingFace)*

Library transformers dari *HuggingFace* digunakan untuk mengakses dan mengimplementasikan berbagai model pralatih (*pre-trained models*) seperti BERT. *Library* ini memudahkan proses tokenisasi teks, pemanggilan model, serta integrasi dengan *framework pytorch*. Dalam penelitian ini, *transformers* digunakan untuk mengelola representasi teks dan menghasilkan *embedding* yang kemudian digunakan dalam proses klasifikasi ulasan.

- *Scikit-learn*

Scikit-learn merupakan *library machine learning* berbasis *Python* yang digunakan untuk mendukung berbagai fungsi dalam proses pengolahan dan evaluasi data. *Library* ini menyediakan beragam fitur seperti pembagian dataset (*train-test split*), *pre-processing* (normalisasi, *encoding*), serta perhitungan metrik evaluasi model seperti *accuracy*, *precision*, *recall*, dan *F1-score* yang didasarkan pada hasil *confusion matrix*. Dalam penelitian ini, *scikit-learn* digunakan untuk membantu tahap evaluasi performa model dan mendukung proses pengolahan data sebelum pelatihan model utama.

- *Imbalanced-Learn (imblearn)*

Imbalanced-Learn merupakan *library* yang digunakan untuk menangani masalah ketidakseimbangan kelas (*class imbalance*) pada dataset. Ketidakseimbangan data dapat menyebabkan model lebih condong memprediksi kelas mayoritas dan mengurangi kemampuan dalam mengenali kelas minoritas. Dalam penelitian ini, *imbalanced-learn* digunakan untuk menerapkan teknik penyeimbangan data seperti *randomoversampling*, sehingga distribusi kelas menjadi lebih seimbang dan performa model dalam mendeteksi ulasan palsu meningkat secara signifikan.

- *NumPy*

NumPy merupakan pustaka *python* yang digunakan untuk melakukan manipulasi *array* dan operasi pada matriks numerik. Pustaka ini menyediakan berbagai fungsi matematis yang efisien dan mendukung komputasi numerik berskala besar. Dalam penelitian ini, *numpy* digunakan untuk mendukung proses pengolahan data numerik, terutama pada tahap *pre-processing* dan pengelolaan data dalam bentuk *batch* sebelum dimasukkan ke dalam model *deep learning*.

- *Pandas*

Pandas merupakan pustaka *python* yang digunakan untuk manipulasi dan analisis data, terutama dalam bentuk tabel (*dataframe*). Pustaka ini menyediakan berbagai fungsi untuk membaca, membersihkan, menggabungkan, dan mentransformasi data. Dalam penelitian ini, *pandas* digunakan untuk mengelola dataset dalam format tabel, melakukan proses *data cleaning*, *transformasi*, menyiapkan data serta analisis data sebelum digunakan dalam model.

- *Matplotlib*

Matplotlib merupakan pustaka visualisasi data yang digunakan untuk menampilkan hasil analisis dan eksperimen model dalam bentuk grafik. *Library* ini berperan penting dalam membantu memahami performa model melalui visualisasi seperti grafik *loss*, *accuracy*, dan

confusion matrix selama proses pelatihan. Dengan visualisasi tersebut, performa dan tren pembelajaran model dapat diamati secara lebih jelas dan informatif.

- TQDM

TQDM merupakan pustaka *Python* yang digunakan untuk menampilkan *progress bar* selama proses iterasi atau pelatihan model. Dengan adanya *progress bar*, pengembang dapat memantau waktu eksekusi dan kemajuan pelatihan secara *real-time*, terutama saat bekerja dengan dataset berukuran besar atau proses pelatihan yang memakan waktu lama.

- *Regex (re) dan String*

Modul *Regex (regular expression)* dan *String* digunakan dalam proses pembersihan teks (*text cleaning*). Keduanya berfungsi untuk menghapus simbol, tanda baca, angka, atau karakter yang tidak relevan sehingga teks menjadi lebih bersih dan siap digunakan dalam tahap pemrosesan berikutnya. Proses ini penting agar model dapat fokus pada konten semantik dari teks tanpa terganggu oleh elemen non-informatif.

- *Gensim*

Gensim merupakan pustaka *python* yang digunakan untuk pemodelan topik dan representasi kata (*word embedding*). Library ini mendukung algoritma seperti *Word2Vec* dan *Doc2Vec* yang mampu mengubah kata menjadi vektor numerik berdasarkan konteks penggunaannya. Dalam penelitian ini, *gensim* dimanfaatkan untuk membantu proses representasi kata dan pembentukan *embedding* yang dapat meningkatkan pemahaman semantik model terhadap teks.

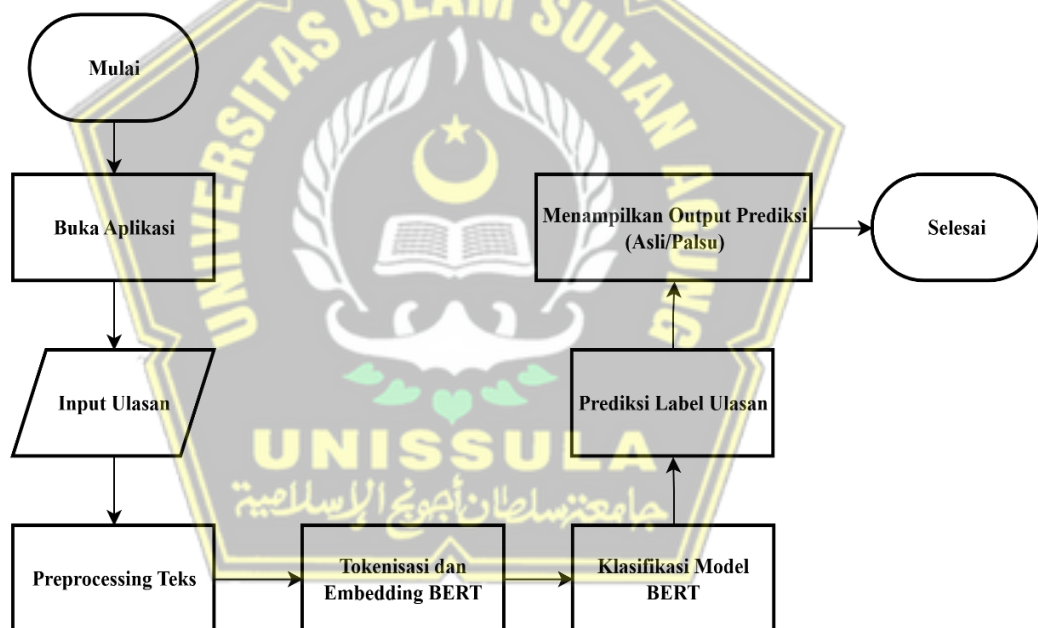
- *Streamlit*

Streamlit merupakan *open-source framework* berbasis *python* yang digunakan untuk membangun aplikasi *web* interaktif, khususnya untuk menampilkan hasil analisis data atau model *machine learning*. *Framework* ini memungkinkan pengembang membuat antarmuka pengguna (*user interface*) secara cepat tanpa memerlukan pengetahuan

mendalam tentang pengembangan *web*. Dalam penelitian ini, *streamlit* digunakan untuk *deploy* model yang telah dilatih ke dalam bentuk aplikasi berbasis *web*, sehingga pengguna dapat berinteraksi langsung dengan sistem untuk melakukan deteksi ulasan palsu secara praktis dan *real-time*.

3.3 Penggunaan Sistem

Pada tahap ini, dilakukan analisis untuk memastikan alur kerja yang dijalankan pengguna untuk sistem deteksi ulasan palsu. Gambar berikut menunjukkan diagram alur yang menjelaskan alur kerja sistem tersebut.



Gambar 3.5 *Flowchart* Sistem

Gambar 3.5 menunjukkan *flowchart* sistem yang menggambarkan proses operasional ketika sistem digunakan oleh pengguna. Secara umum, sistem dirancang agar mudah digunakan dan mampu memberikan hasil klasifikasi secara cepat dan akurat. Adapun tahapan dalam alur sistem adalah sebagai berikut:

1. Pengguna mengakses aplikasi berbasis *web* melalui *browser*. Setelah halaman aplikasi terbuka, pengguna akan melihat tampilan utama sistem deteksi ulasan palsu yang dilengkapi dengan kolom input teks ulasan dan tombol untuk memproses data.
2. Pengguna menyalin atau mengetik teks ulasan dari platform Tiktok Shop by Tokopedia ke dalam kolom input yang tersedia pada halaman utama sistem.
3. Setelah ulasan dimasukkan, sistem melakukan *preprocessing* teks, yang bertujuan untuk membersihkan dan menyiapkan data teks agar dapat diproses oleh model. Tahap ini mencakup pembersihan teks dari karakter yang tidak diperlukan.
4. Teks hasil *preprocessing* selanjutnya melalui tahap tokenisasi dan *embedding* menggunakan BERT, di mana teks diubah menjadi representasi vektor yang mampu menangkap konteks dan makna kata dalam kalimat.
5. Model klasifikasi menganalisis pola kata, konteks kalimat, serta hubungan semantik dalam teks untuk menentukan apakah ulasan tersebut tergolong asli (*genuine review*) atau palsu (*fake review*).
6. Berdasarkan hasil klasifikasi, sistem melakukan prediksi label ulasan, yaitu menentukan apakah ulasan yang dianalisis termasuk kategori asli atau palsu.
7. Sistem kemudian menampilkan hasil klasifikasi kepada pengguna berupa label 'Ulasan Asli' atau 'Ulasan Palsu' sesuai dengan hasil analisis model terhadap teks yang dimasukkan. Selain itu, sistem juga menampilkan ciri-ciri ulasan berdasarkan hasil deteksi untuk memberikan pemahaman lebih kepada pengguna.

Dengan alur kerja ini, pengguna dapat dengan mudah memanfaatkan sistem untuk memeriksa keaslian ulasan produk sebelum melakukan keputusan pembelian. Selain itu, sistem juga dapat membantu pelaku bisnis dalam memantau kredibilitas ulasan pada produk yang dijual.

3.4 Perancangan *User Interface*

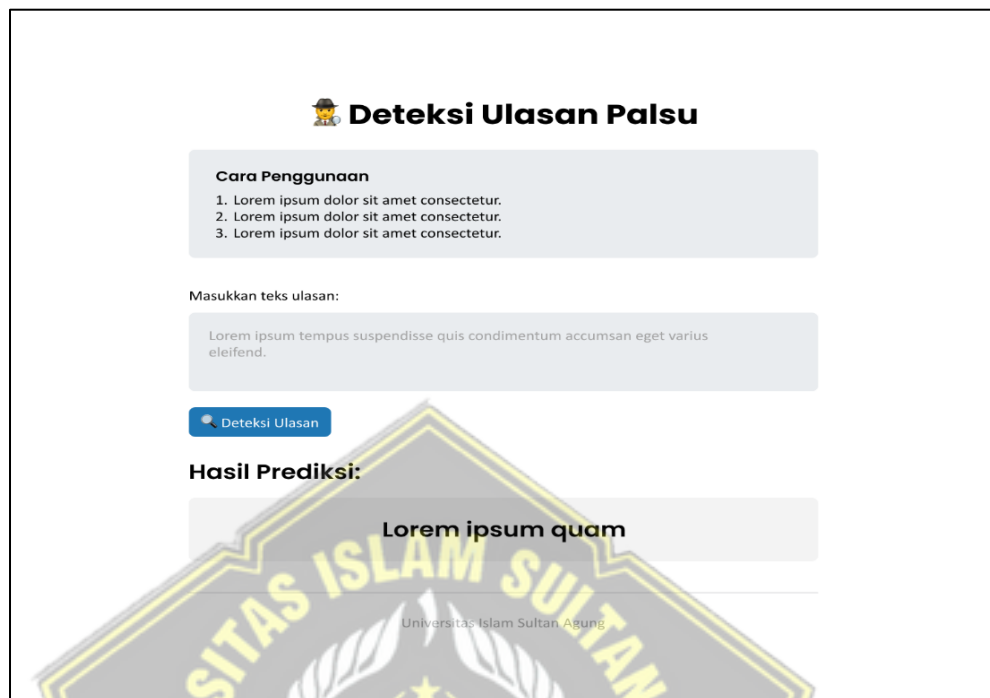
3.4.1 Halaman awal sistem



Gambar 3.6 Tampilan Halaman Awal Sistem

Gambar 3.6 menunjukkan tampilan rancangan antarmuka pada halaman awal sistem ‘Deteksi Ulasan Palsu’. Pada halaman ini terdapat bagian ‘Cara Penggunaan’ yang berisi panduan langkah-langkah dalam menggunakan sistem, mulai dari memasukkan teks ulasan hingga proses deteksi dilakukan. Pengguna diminta untuk mengetik atau menyalin teks ulasan ke dalam kolom yang telah disediakan. Setelah itu, pengguna dapat menekan tombol ‘Deteksi Ulasan’ berwarna biru untuk memulai proses analisis. Sistem kemudian memproses teks yang dimasukkan menggunakan model BERT yang telah dilatih sebelumnya untuk mendeteksi keaslian ulasan.

3.4.2 Tampilan hasil sistem



Gambar 3.7 Tampilan Halaman Hasil Sistem

Gambar 3.7 merupakan tampilan antarmuka yang muncul setelah pengguna berhasil memasukkan dan memproses teks ulasan melalui sistem ‘Deteksi Ulasan Palsu’. Pada tahap ini, sistem akan menampilkan hasil prediksi secara otomatis berdasarkan teks yang telah dimasukkan sebelumnya. Hasil prediksi ditunjukkan dalam bentuk label ‘Ulasan Asli’ atau ‘Ulasan Palsu’, sesuai dengan klasifikasi yang dilakukan oleh model BERT yang telah dilatih menggunakan data ulasan produk. Selain menampilkan label hasil deteksi, sistem juga menampilkan ciri-ciri ulasan berdasarkan analisis model, seperti karakteristik umum yang sering ditemukan pada ulasan asli maupun ulasan palsu. Melalui tampilan ini, pengguna dapat mengetahui tidak hanya hasil klasifikasi, tetapi juga memahami alasan dan pola bahasa yang memengaruhi keputusan sistem dalam menentukan keaslian sebuah ulasan.

BAB IV


HASIL DAN ANALISIS PENELITIAN

4.1 Persiapan Dataset

Data ulasan yang digunakan dalam penelitian ini diperoleh dari platform Tiktok Shop by Tokopedia pada kategori *fashion*. Proses pengambilan data dilakukan dengan menggunakan perangkat lunak *WebHarvy* yang berfungsi untuk mengekstraksi teks ulasan secara otomatis dari halaman *web* Tiktok Shop by Tokopedia. Data yang sudah di peroleh hanya data ulasan *fashion* saja dengan jumlah 1.175 data ulasan, dan disimpan dalam format Excel untuk memudahkan proses pelabelan dan *preprocessing*.

Tabel 4.1 Hasil Pengumpulan Dataset Ulasan

ULASAN
Kain nya halus, nyaman, adem, ga nerawang benr bner rekomendasi,
cakepppp polllll mau co wrna lain ❤️❤️❤️❤️ ,
tb 157 bb 51kg ngepass longgar dikitt di m, bagusss bett aseliii 😊😊👉 ,
Bahan nya strechy , kualitas bagus jahitan rapih..,
ukuran s,
Maaf ada koyak dibagikan ketiak,lalu benang nya keluar', sbnernya mau ngasi 🌟 3 tp krna celana nya Mayan bagus bajunya sih agak tipis hrus pke dalam 🙌. Celana nya si oke,dingin,karetnya enak masih longgar di BB 95 tinggi 165 dan celana nya gede bgtttt kaya gembor/baju nya bagus tp kebesaran tp,
Alhamdulillah paket udh sampai dg selamat,sesuai dengan keterangan real.bagus banget,gk nyesel langsung CO 2,nyaman juga bahan adem enak buat santai sehari²,kulotnya juga nyaman longgar 😊Makasih seller 🙌 lancar terus usahanya,
Ukuranya pas warna sesuai.tp saku bajunya terlalu besar,

ULASAN	AN 1	AN 2	AN 3	AN 4	AN 5	HASIL AKHIR
Alhamdulillah paket udh sampai dg selamat,sesuai dengan keterangan real.bagus banget,gk nyesel langsung CO 2,nyaman juga bahan adem enak buat santai sehari ² ,kulotnya juga nyaman longgar 😊Makasih seller 🙏 lancar terus usahanya,	Asli	Asli	Asli	Asli	Asli	Asli
Ukuranya pas warna sesuai.tp saku bajunya terlalu besar,	Asli	Asli	Asli	Asli	Asli	Asli
bagus barang nya bintang bnyk 	Palsu	Palsu	Palsu	Palsu	Palsu	Palsu
Alhamdulillah sdh nyampk udah ditunggu Semoga sesuai keinginan,	Asli	Asli	Asli	Asli	Asli	Asli

Berdasarkan hasil pelabelan tersebut, seluruh ulasan berhasil memperoleh label akhir sesuai dengan aturan penetapan label yang telah ditentukan. Dataset berlabel ini selanjutnya digunakan sebagai dasar dalam penghitungan kesepakatan antar-annotator serta sebagai data masukan pada

tahap pemodelan.

4.2.2 Distribusi label

Distribusi label menunjukkan jumlah data ulasan pada masing-masing kategori setelah proses pelabelan dan penetapan label akhir dilakukan. Tabel 4.3 menyajikan distribusi jumlah ulasan berdasarkan kategori label Asli dan Palsu.

Tabel 4.3 Hasil Total Distribusi Kelas Seluruh Ulasan

LABEL	JUMLAH
Asli	1056
Palsu	119
Total	1175

Berdasarkan distribusi label tersebut, seluruh data ulasan berhasil diklasifikasikan ke dalam dua kategori label, yaitu Asli dan Palsu, tanpa adanya data yang dikategorikan sebagai ambigu.

4.2.3 Nilai kesepakatan *annotator*

Untuk mengukur tingkat kesepakatan antar-annotator terhadap hasil pelabelan data, dilakukan perhitungan *Inter-Annotator Agreement* (IAA) menggunakan metode *fleiss' kappa*. Nilai kesepakatan yang diperoleh dari perhitungan tersebut disajikan pada Tabel 4.4.

Tabel 4.4 Nilai Kesepakatan Antar-annotator

METODE	NILAI
<i>Fleiss' Kappa</i>	0,644

Berdasarkan nilai *fleiss' kappa* tersebut, terlihat bahwa tingkat kesepakatan antar-annotator relatif tinggi. Nilai tersebut menunjukkan bahwa tingkat kesepakatan antar-annotator berada pada kategori *substantial agreement* menurut interpretasi Landis dan Koch (1977).

Proses pelabelan data dilakukan sebanyak 2 iterasi untuk memastikan konsistensi dan keandalan hasil anotasi. Tingkat kesepakatan yang *substansial* menunjukkan bahwa sebagian besar *annotator* memberikan label

secara konsisten terhadap data ulasan. Hal ini menandakan bahwa pedoman pelabelan telah dipahami dengan baik, sehingga perbedaan subjektivitas dalam penilaian dapat diminimalkan. Meskipun terdapat beberapa perbedaan label, perbedaan tersebut masih berada dalam batas wajar pada proses pelabelan teks berbasis opini.

Distribusi label menunjukkan bahwa seluruh data ulasan berhasil diklasifikasikan ke dalam dua kategori, yaitu Asli dan Palsu, tanpa adanya data ambigu. Kondisi ini menegaskan bahwa metode *majority voting* dalam penetapan label akhir berjalan efektif. Dengan demikian, kombinasi antara tingkat kesepakatan *substansial* antar-annotator dan distribusi label yang jelas menunjukkan bahwa dataset hasil pelabelan layak dan valid untuk digunakan pada tahap *preprocessing* dan pelatihan model BERT.

4.3 Text Processing

- a. Membersihkan nilai kosong
 - Seluruh baris yang memiliki nilai kosong (*missing values*) diperiksa menggunakan perintah `df.isnull().sum()`.
 - Hasil pemeriksaan menunjukkan tidak ada nilai kosong pada dataset.
 - Dengan demikian, semua 1.175 ulasan tetap utuh dan siap digunakan untuk tahap pemrosesan selanjutnya.
- b. Membersihkan *duplicate*
 - Dataset diperiksa adanya duplikasi menggunakan perintah `df.duplicated().sum()`.
 - Ditemukan 19 baris duplikat, yang kemudian dihapus dengan perintah `df = df[~df.duplicated()]`.
 - Setelah penghapusan duplikat, dataset menjadi 1.156 ulasan unik.
- c. *Remove Emoji*

Tabel berikut menyajikan hasil perubahan teks sebelum dan sesudah proses penghapusan emoji. Tabel menunjukkan bagaimana seluruh emoji dalam teks ulasan telah dihapus, sehingga hanya tersisa isi verbal dari setiap ulasan.

Tabel 4. 5 Hasil Remove Emoji

SEBELUM	SESUDAH
TB 157, BB 51 Kg. ngepass longgar dikitt di 'm' bagusss banget asli 🤔 😊 🤔	TB 157, BB 51 Kg. ngepass longgar dikitt di 'm' bagusss banget asli

d. *Lowercase* (mengubah semua huruf kapital menjadi huruf kecil)

Tabel berikut menyajikan hasil perubahan teks sebelum dan sesudah tahap *lowercase*. Tabel menunjukkan bahwa seluruh huruf kapital dalam teks ulasan telah diubah menjadi huruf kecil, sehingga teks menjadi seragam.

Tabel 4. 6 Hasil *Lowercase*

SEBELUM	SESUDAH
TB 157, BB 51 Kg. ngepass longgar dikitt di 'm' bagusss banget asli	tb 157, bb 51 kg. ngepass longgar dikitt di 'm' bagusss banget asli

e. *Remove Character*

Tabel berikut menyajikan hasil perubahan teks sebelum dan sesudah tahap *remove character*. Tabel menunjukkan bahwa seluruh karakter yang tidak termasuk huruf kecil, angka, dan spasi telah dihapus, sehingga teks hanya tersisa elemen verbal yang relevan.

Tabel 4. 7 Hasil *Remove Character*

SEBELUM	SESUDAH
tb 157, bb 51 kg. ngepass longgar dikitt di 'm' bagusss banget asli	tb 157 bb 51 kg ngepass longgar dikitt di m bagusss banget asli

Setelah dilakukan proses *text processing*, dataset ulasan menjadi lebih bersih dan konsisten. Pemeriksaan nilai kosong menunjukkan tidak ada data yang hilang, sehingga seluruh 1.175 ulasan tetap utuh, sedangkan

penghapusan 19 baris duplikat menghasilkan 1.156 ulasan unik. Seluruh emoji yang terdapat dalam teks dihapus, huruf kapital diubah menjadi huruf kecil, dan karakter yang tidak termasuk huruf kecil, angka, atau spasi juga dihilangkan. Dengan langkah-langkah ini, teks ulasan hanya tersisa elemen verbal yang relevan, sehingga dataset menjadi seragam dan lebih mudah diproses pada tahap tokenisasi, ekstraksi fitur, dan pelatihan model berbasis BERT. Secara keseluruhan, proses ini meningkatkan kualitas data, meminimalkan kebisingan, dan menyiapkan dataset yang siap digunakan untuk analisis lebih lanjut.

4.4 Pembagian Dataset

Dataset ulasan disiapkan untuk digunakan dalam tahap pemodelan setelah semua prosedur pemrosesan teks selesai. Dataset tersebut kemudian dibagi menjadi beberapa bagian untuk pengujian, validasi, dan pelatihan model. Untuk mengumpulkan tiga jenis data yang berbeda, pembagian ini dilakukan dalam tiga langkah:

- data latih (*training set*),
- data validasi (*validation set*),
- data uji (*test set*).

Metode `'train_test_split'` dari pustaka *scikit-learn* digunakan untuk membagi data. Delapan puluh persen dari dataset digunakan untuk pelatihan awal, dan dua puluh persen sisanya digunakan untuk pengujian. Data pelatihan dan validasi kemudian dipisahkan dari data pelatihan asli masing-masing sebesar 80% dan 20%. Hasilnya, 64% dari dataset akhir terdiri dari data pelatihan, 16% data validasi, dan 20% data pengujian.

Pembagian data dilakukan dengan mempertahankan proporsi kelas ulasan asli dan ulasan palsu pada setiap subset. Selain itu, nilai `random_state` ditetapkan sebesar 2042 untuk memastikan hasil pembagian data bersifat konsisten. Berikut ini tabel yang menunjukkan jumlah sampel masing-masing kelas sebelum dilakukannya *oversampling*.

Tabel 4.8 Total Sampel Dataset Sebelum *Oversampling*

DATASET	KELAS	JUMLAH SAMPEL
Training Set	Asli	665
	Palsu	74
	Total	739
Validation Set	Asli	167
	Palsu	18
	Total	185
Test Set	Asli	209
	Palsu	23
	Total	232

Setelah pembagian data, dilakukan proses penyeimbangan pada *training set* menggunakan metode *oversampling* (*RandomOverSampler*). Proses ini menambahkan sampel dari kelas minoritas (ulasan palsu) sehingga jumlah sampel pada kedua kelas menjadi seimbang dengan jumlah sampel kelas mayoritas (ulasan asli), seperti terlihat pada tabel berikut:

Tabel 4.9 Data Training Yang Sudah Dilakukan *Oversampling*

KELAS	JUMLAH
Asli	665
Palsu	665

Dataset hasil *oversampling* selanjutnya digunakan sebagai data masukan pada tahap pelatihan model.

Penetapan nilai *seed* dilakukan untuk memastikan seluruh proses yang melibatkan pengacakan, seperti pembagian data dan pelatihan model, menghasilkan keluaran yang konsisten setiap kali eksperimen dijalankan. Hal ini penting untuk menjaga keterulangan (*reproducibility*) hasil penelitian. Pembagian dataset ke dalam data latih, validasi, dan uji bertujuan untuk memastikan bahwa model dilatih, dievaluasi, dan diuji menggunakan data

yang berbeda. Data uji digunakan untuk mengukur kemampuan model dalam melakukan generalisasi terhadap data yang belum pernah dilihat sebelumnya, data validasi digunakan untuk meningkatkan kinerja model selama pelatihan, dan data pelatihan digunakan untuk menemukan pola..

Penerapan metode *oversampling* pada data latih dilakukan untuk mengatasi ketidakseimbangan jumlah data antar kelas. Dengan menyeimbangkan kelas mayoritas dan minoritas, model diharapkan tidak bias terhadap satu kelas tertentu dan mampu mempelajari karakteristik ulasan asli maupun palsu secara lebih optimal.

4.5 Tokenisasi BERT

Proses tokenisasi dilakukan menggunakan *BertTokenizer* dari pustaka *transformers* dengan model pralatih *bert-base-uncased*. Agar model BERT dapat memprosesnya, semua teks ulasan diubah menjadi bentuk numerik. Token khusus [CLS] di awal dan [SEP] di akhir setiap kalimat terdapat di setiap teks. Panjang maksimum token ditetapkan menggunakan parameter *max_length* dengan mekanisme *truncation* dan *padding* sehingga seluruh data memiliki panjang yang seragam. Hasil dari proses tokenisasi ini berupa *input_ids* yang merepresentasikan token numerik serta *attention_mask* yang menunjukkan posisi token aktif dan *padding*.

Penggunaan *BertTokenizer* dengan model *bert-base-uncased* bertujuan untuk memastikan bahwa representasi teks yang dihasilkan selaras dengan arsitektur dan pengetahuan linguistik BERT pralatih. Token [CLS] dimanfaatkan sebagai representasi global dari setiap ulasan sehingga informasi kontekstual dapat dirangkum secara menyeluruh. Penyeragaman panjang token membantu menjaga stabilitas proses pelatihan, sementara *attention mask* memungkinkan model memfokuskan perhatian pada token yang relevan. Dengan pendekatan ini, data teks siap digunakan secara optimal pada tahap pemodelan berbasis BERT.

4.6 Konfigurasi Model BERT

Model klasifikasi teks dibangun menggunakan arsitektur *Bert_Classifier* berbasis *PyTorch* dan pustaka *transformers*. Model menggunakan BERT pralatih *bert-base-uncased* sebagai ekstraktor fitur utama.

Konfigurasi model terdiri dari:

- Dimensi keluaran BERT sebesar 768
- *Hidden layer classifier* sebesar 128 neuron
- Jumlah kelas keluaran sebanyak 2 kelas

Representasi token [CLS] dari keluaran BERT digunakan sebagai masukan ke lapisan klasifikasi yang terdiri dari

- `nn.Linear(n_input, n_hidden)` : mengubah dimensi keluaran BERT menjadi ukuran yang lebih kecil.
- `nn.ReLU()` : fungsi aktivasi non-linear yang membantu jaringan mengenali pola kompleks.
- `nn.Dropout(p=0.5)` : mencegah *overfitting* dengan cara menonaktifkan sebagian neuron secara acak saat pelatihan.
- `nn.Linear(n_hidden, n_output)` : menghasilkan prediksi akhir berupa dua kelas.

Model menghasilkan nilai logits sebagai keluaran akhir.

Arsitektur *Bert_Classifier* dirancang untuk memanfaatkan representasi kontekstual yang dihasilkan oleh BERT dalam memahami makna ulasan secara menyeluruh. Penggunaan lapisan klasifikasi sederhana bertujuan untuk mengolah fitur hasil ekstraksi BERT secara efisien tanpa meningkatkan kompleksitas model secara berlebihan. Penerapan fungsi aktivasi non-linear dan *dropout* membantu model dalam mengenali pola yang lebih kompleks sekaligus meningkatkan kemampuan generalisasi. Dengan memanfaatkan representasi global dari token [CLS], model dapat melakukan klasifikasi berdasarkan konteks keseluruhan ulasan, sehingga diharapkan mampu membedakan ulasan asli dan ulasan palsu secara efektif.

4.7 Hyperparameter Model

Pada tahap pemodelan, penelitian ini menetapkan sejumlah *hyperparameter* penting untuk memastikan proses pelatihan model BERT berjalan optimal. *Hyperparameter* ini mencakup pengaturan dasar seperti *batch size*, *learning rate*, jumlah *epoch*, serta panjang maksimum token. Ringkasan parameter yang digunakan dalam penelitian ini disajikan pada tabel berikut:

Tabel 4.10 *Hyperparameter Tunning*

NO	HYPERPARAMETER	NILAI (VALUE)
1	<i>Max Length</i>	100
2	<i>Batch Size</i>	16
3	<i>Learning Rate</i>	1e-5
4	<i>Optimizer</i>	<i>AdamW</i>
5	<i>Epsilon Optimizer</i>	1e-8
6	<i>Epoch</i>	10
7	<i>Dropout Rate</i>	0,5
8	<i>Weight_decay</i>	0,1
9	<i>Scheduler</i>	<i>Linear Warmup</i>
10	<i>Hidden Size (Classifier)</i>	128
11	<i>Freeze BERT</i>	<i>Freeze 6 Layer</i>
12	<i>Loss</i>	<i>CrossEntropyLoss</i>

Berdasarkan Tabel 4.6, proses pelatihan model menggunakan panjang maksimum token ditetapkan sebesar 100, *batch size* sebesar 16, dan *learning rate* sebesar 1e-5. Proses pelatihan model dilakukan selama maksimal 10 *epoch* dengan *optimizer* AdamW dan nilai *epsilon* sebesar 1e-8. Model dilatih menggunakan fungsi *loss* *CrossEntropyLoss* serta *dropout rate* sebesar 0,5 pada lapisan klasifikasi. Untuk mengatur laju pembelajaran selama proses pelatihan, digunakan *scheduler linear warmup*. Ukuran *hidden layer* pada *classifier* ditetapkan sebesar 128 neuron. Selain itu, diterapkan teknik *freeze BERT* sebagian parameter BERT, yaitu dengan membekukan 6 layer awal dan

melatih 6 layer terakhir. Seluruh konfigurasi *hyperparameter* tersebut digunakan secara konsisten pada proses pelatihan model dalam penelitian ini.

Pemilihan *hyperparameter* di atas disesuaikan dengan kebutuhan penelitian untuk menghasilkan performa model yang optimal dalam mendeteksi ulasan palsu. Parameter *max_length* ditetapkan pada nilai 100 karena ulasan terpanjang sejumlah 113 namun rata-rata panjang ulasan pada dataset relatif pendek, sehingga nilai ini cukup mewakili konteks tanpa membebani proses komputasi. *Batch size* sebesar 16 dipilih untuk menjaga keseimbangan antara stabilitas pembaruan bobot dan efisiensi pelatihan. *Learning rate* bernilai $1e-5$, nilai ini dipilih karena dataset yang relatif sedikit supaya dapat stabil dalam pelatihan serta mencegah terjadinya *overfitting*. *Optimizer* AdamW dan nilai *epsilon* $1e-8$ digunakan untuk meningkatkan stabilitas perhitungan gradien, sementara *weight decay* dinaikan menjadi 0.1. Nilai tersebut digunakan pada *fine-tuning* model BERT karena memberikan regularisasi yang lebih kuat untuk mencegah bobot model tumbuh terlalu besar. Dengan penambahan regularisasi ini, performa model menjadi lebih stabil dan *overfitting* berhasil dikurangi.

Model dilatih selama 10 *epoch* untuk memantau secara menyeluruh perkembangan performa model sepanjang proses pembelajaran. Proses pelatihan dilengkapi dengan *early stopping*, yaitu mekanisme penghentian otomatis ketika metrik validasi tidak mengalami peningkatan selama beberapa *epoch* terus-menerus. Teknik ini membantu mencegah *overfitting* serta memastikan bahwa model berhenti pada titik performa optimal. Sedangkan *dropout rate* bernilai 0,5 digunakan pada lapisan klasifikasi untuk meningkatkan generalisasi model. Scheduler *Linear Warmup* diterapkan untuk menyesuaikan laju pembelajaran secara bertahap, sehingga pelatihan menjadi lebih stabil di tahap awal. Ukuran *hidden size* sebesar 128 pada *classifier* dipilih untuk memberikan kapasitas pemrosesan yang cukup tanpa membuat model menjadi terlalu kompleks.

Pada penelitian ini diterapkan teknik *freeze* sebagian parameter BERT, yaitu dengan membekukan 6 layer awal (layer 0–5), sementara 6 layer sisanya

tetap dilatih. Pendekatan *partial freezing* ini digunakan untuk mempertahankan representasi dasar BERT yang sudah stabil, sekaligus memungkinkan bagian atas model beradaptasi secara lebih spesifik terhadap karakteristik dataset ulasan Tiktok Shop by Tokopedia. Terakhir, *loss function* yang digunakan adalah *CrossEntropyLoss* karena penelitian ini merupakan kasus klasifikasi biner. Dengan pengaturan *hyperparameter* tersebut, model diharapkan mampu mempelajari perbedaan pola antara ulasan asli dan ulasan palsu secara efektif.

4.8 Performa Training Model

Sebelum dilakukan evaluasi lebih lanjut, penting untuk melihat performa model BERT selama proses training. Tabel berikut menyajikan ringkasan nilai *train loss*, *train accuracy*, *validation loss*, dan *validation accuracy* pada beberapa *epoch* tertentu. Data ini memberikan gambaran kemampuan model dalam mempelajari pola dari data ulasan serta performa model pada data validasi.

Pada penelitian ini, jumlah *epoch* yang direncanakan adalah 10 *epoch*, namun proses pelatihan berhenti lebih awal pada *epoch* ke-7 sebelum memasuki tahap validasi. Akibatnya, data performa model yang lengkap hanya tersedia hingga *epoch* ke-6.

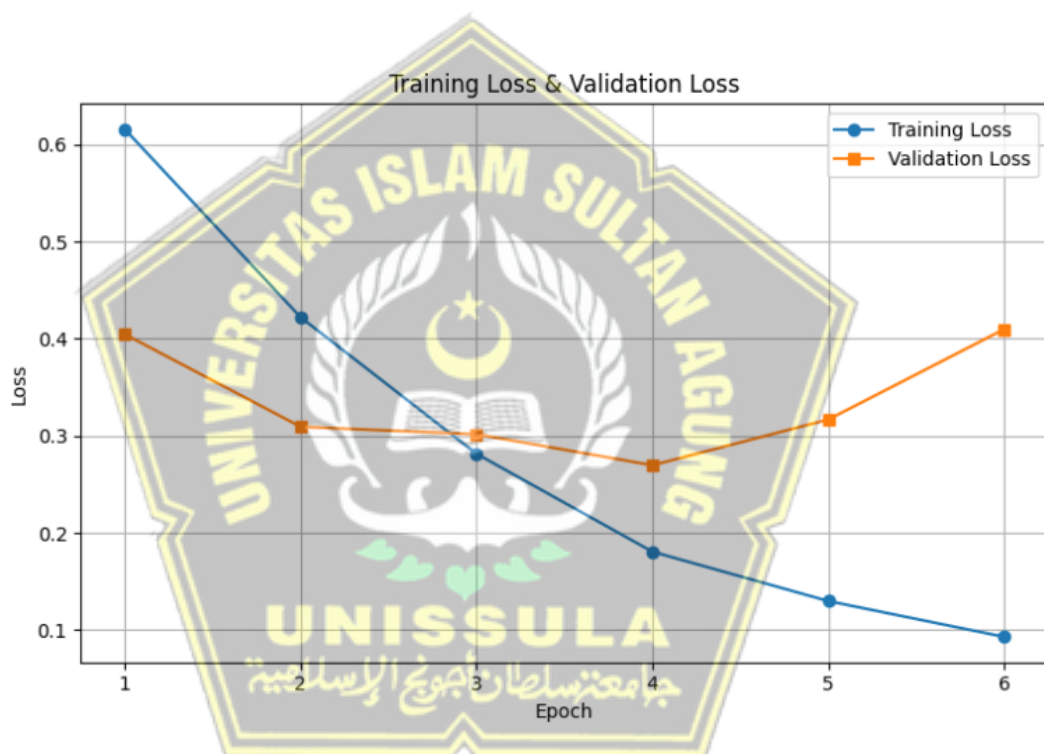
Tabel 4.11 Hasil *Training*

<i>Epoch</i>	<i>Train Loss</i>	<i>Train Acc</i>	<i>Val loss</i>	<i>Val Acc</i>
1	0.615770	69.35	0.404640	86.57
2	0.422098	83.11	0.309306	88.14
3	0.281483	90.48	0.301629	88.14
4	0.180658	94.12	0.269680	91.26
5	0.129861	96.21	0.316788	89.70
6	0.092969	97.47	0.409867	89.70

Tabel ini menunjukkan perkembangan performa model BERT selama 6 *epoch* pertama. Terlihat peningkatan *train accuracy* dari 69,35% menjadi

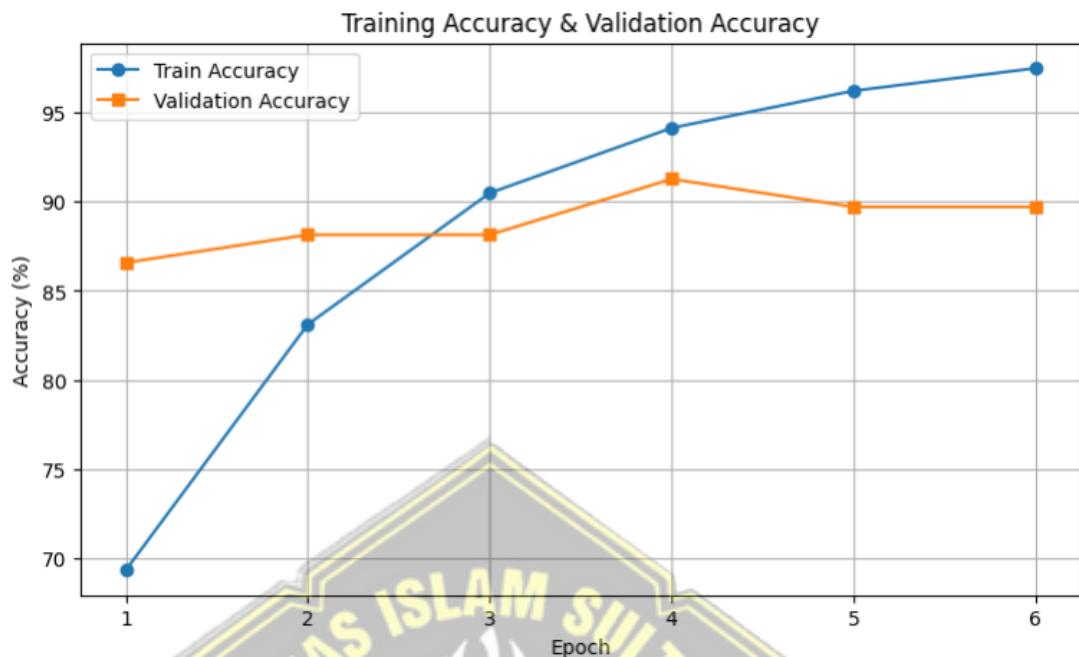
97,47%, sedangkan *validation accuracy* mengalami fluktuasi, menandakan indikasi awal *overfitting* pada *epoch* lebih tinggi. *Epoch* 4 menunjukkan performa optimal dengan *validation loss* terendah 0,269680 dan perbedaan minimal antara *train accuracy* dan *validation accuracy*.

Untuk memudahkan pemahaman terhadap perkembangan performa model selama proses *training*, dilakukan visualisasi *curve loss* dan *accuracy*. Visualisasi ini menampilkan perubahan nilai *train loss*, *validation loss*, *train accuracy*, dan *validation accuracy* pada setiap *epoch*.



Gambar 4.1 Kurva *Training Loss* dan *Validation Loss*

Kurva ini memperlihatkan penurunan *training loss* yang konsisten pada setiap *epoch*, menandakan model belajar dengan baik dari data *training*. *Validation loss* juga konsisten menurun namun pada *epoch* 5–6 mengalami kenaikan yang lumayan tinggi.



Gambar 4.2 Kurva *Training Accuracy* dan *Validation Accuracy*

Kurva ini menunjukkan *training accuracy* meningkat secara konsisten, sementara *validation accuracy* stabil pada *epoch* 2-3 dan mengalami kenaikan pada *epoch* 4 kemudian mengalami penurunan di *epoch* 5 dan stabil hingga *epoch* 6. Hal ini mengindikasikan potensi *overfitting*, terutama ketika *train accuracy* meningkat tetapi *validation accuracy* menurun. Titik optimal model berada pada *epoch* 4, di mana *validation accuracy* tertinggi sebelum penurunan berikutnya.

Pada tabel 4.11 terlihat bahwa model menunjukkan peningkatan performa yang signifikan pada awal pelatihan. *Train accuracy* terus naik dari *epoch* ke *epoch*, menandakan model mampu mempelajari pola dari data latih dengan sangat baik. Namun mulai *epoch* 5 dan 6, terlihat adanya gejala awal *overfitting*, yaitu ketika *train accuracy* terus meningkat, tetapi *validation accuracy* mulai turun, dan *validation loss* semakin naik. Hal ini menunjukkan bahwa model mulai terlalu menyesuaikan diri dengan data pelatihan, sehingga kemampuan generalisasi terhadap data validasi mulai melemah.

- Pada *epoch* 1, model sudah menunjukkan performa yang sangat baik

dengan *train accuracy* 69,35% dan *val accuracy* 86,57%. Nilai ini menandakan model langsung mampu mengenali pola dasar pada dataset dengan baik.

- Pada *epoch 2*, *train accuracy* meningkat menjadi 83,11%, dan *val accuracy* naik menjadi 88,14%. Untuk *val loss* dan *train loss* nya juga mengalami penurunan.
- Pada *epoch 3*, *train accuracy* terus meningkat menjadi 90,48% sedangkan *val accuracy* tetap stabil di angka 88,14%. *Val loss* mengalami penurunan sebesar 0,301629.
- Pada *epoch 4*, model mencapai performa terbaik. *train accuracy* meningkat hingga 94,12%, dan *val accuracy* mengalami kenaikan menjadi 91,26%, sedangkan *val loss* mengalami penurunan yang menjadi 0,269680. *Epoch* ini menjadi titik optimal karena perbedaan antara *train accuracy* dan *val accuracy* relatif kecil, menunjukkan keseimbangan antara pelatihan dan validasi. Hal ini ditunjukkan oleh nilai *validation loss* paling rendah diantara *epoch* yang lain.
- Pada *epoch 5*. Meskipun *train accuracy* kembali meningkat menjadi 96,21%, namun *val accuracy* turun menjadi 89,70%. Di sisi lain, *val loss* juga naik menjadi 0,316788, yang menunjukkan bahwa model kembali *overfitting* setelah mengalami performa terbaik.
- Pada *epoch 6*. Gejala *overfitting* semakin kuat karena *train accuracy* terus meningkat menjadi 97,47%, namun *val accuracy* disini tetap stabil di angka 89,70%. Di sisi lain, *val loss* juga semakin naik hingga menjadi 0,409867. Hal ini *early stopping* mendeteksi jika performa model semakin *overfitting* sehingga pada *epoch 7 training* dihentikan.

Dari gambar 4.1 terlihat bahwa *train loss* menurun secara konsisten pada setiap *epoch*, menunjukkan bahwa model semakin baik dalam mempelajari pola dari data *training*. Sementara itu, *validation loss* tidak menunjukkan penurunan yang stabil. Pola ini mengindikasikan bahwa model mulai menunjukkan gejala *overfitting* setelah *epoch* ke 4. Tanda-tanda

overfitting terlihat jelas pada *epoch* 5 dan 6 yaitu ketika *train loss* terus menurun tetapi *validation loss* justru meningkat. Kondisi ini menandakan bahwa model terlalu menyesuaikan diri dengan data *training* dan mulai kehilangan kemampuan untuk menggeneralisasi dengan baik pada data validasi. Dengan memperhatikan pola tersebut, *epoch* optimal berada pada titik ketika *validation loss* berada pada nilai terendah sebelum terjadi peningkatan kembali.

Dari gambar 4.2 memperlihatkan perubahan *train accuracy* dan *validation accuracy* selama proses pelatihan. *Train accuracy* tampak meningkat secara konsisten dari *epoch* ke *epoch*, menunjukkan bahwa model mampu mempelajari pola pada data *training* dengan sangat baik. Sementara itu, *validation accuracy* mengalami fluktuasi yaitu naik pada *epoch* 2, stabil pada *epoch* 3, naik lagi pada *epoch* 4, dan kembali turun pada *epoch* 5 dan 6. Pola yang tidak stabil ini menunjukkan bahwa performa model pada data validasi tidak meningkat secara konsisten meskipun akurasi pada data *training* terus naik. Perbedaan tren ini mengindikasikan potensi *overfitting*, terutama ketika *train accuracy* terus meningkat tetapi *validation accuracy* justru menurun atau berfluktuasi. Dengan melihat pola tersebut, *epoch* optimal adalah titik di mana *validation accuracy* mencapai nilai terbaik sebelum kembali menurun atau tidak stabil pada *epoch* berikutnya.

Berdasarkan visualisasi *curve loss* dan *accuracy*, dapat disimpulkan bahwa model BERT mampu belajar dengan sangat baik pada data *training*, terlihat dari penurunan *train loss* yang konsisten serta peningkatan *train accuracy* yang stabil. Namun, *validation loss* yang berfluktuasi dan *validation accuracy* yang tidak menunjukkan peningkatan stabil mengindikasikan adanya gejala *overfitting* pada *epoch* lebih tinggi. Kondisi ini menunjukkan bahwa meskipun model semakin akurat pada data *training*, kemampuannya dalam menggeneralisasi pada data baru tidak meningkat secara sejalan. Dengan mempertimbangkan kedua grafik, titik pelatihan yang paling optimal berada pada *epoch* ketika *validation loss* berada pada nilai paling rendah dan *validation accuracy* mencapai nilai terbaik sebelum

mengalami fluktuasi kembali. Titik ini menjadi dasar untuk strategi pelatihan berikutnya, seperti menerapkan *early stopping* atau menambahkan regularisasi agar model tetap stabil dan tidak *overfitting*.

4.9 Evaluasi Model

Setelah proses pelatihan selesai, model BERT dievaluasi untuk mengukur performanya dalam mengklasifikasikan ulasan asli dan ulasan palsu. Evaluasi dilakukan menggunakan data validasi dengan beberapa metrik utama, yaitu *Precision*, *Recall*, *F1-score*, dan *Accuracy*. Penghitungan metrik ini dilakukan menggunakan fungsi *classification report* dari pustaka `sklearn.metrics`.

Setelah pelatihan, kemampuan model BERT untuk membedakan antara ulasan asli dan ulasan palsu dinilai. *Precision*, *Recall*, *F1-score*, dan *Accuracy* adalah metrik utama yang digunakan untuk validasi data dalam evaluasi. Fungsi laporan klasifikasi dari pustaka `sklearn.metrics` digunakan untuk menghitung metrik-metrik ini.

Tabel 4.12 Tabel *Classification Report*

KELAS	ACCURACY	PRECISION	RECALL	F1-SCORE	SUPPORT
Asli	0,94	0,96	0,97	0,96	209
Palsu	0,94	0,68	0,65	0,67	23

Tabel 4.12 adalah hasil evaluasi model menggunakan laporan klasifikasi yang mencakup metrik seperti presisi, recall, F1-score, dan kuantitas data per kelas (dukungan) ditampilkan pada Tabel 4.12. Dengan presisi 0,96, recall 0,97, dan F1-score 0,96, model menunjukkan kinerja yang sangat baik untuk kelas asli. Ini menunjukkan bahwa model dapat secara akurat dan konsisten mengidentifikasi ulasan berlabel otentik. Di sisi lain, kelas palsu berkinerja lebih buruk, dengan F1-score 0,67, presisi 0,68, dan recall 0,65. Algoritma masih sering gagal mengidentifikasi data yang sebenarnya diberi label palsu, seperti yang dibuktikan oleh nilai recall yang

rendah untuk kelas ini. Ketidakseimbangan jumlah data uji, yaitu 209 untuk kelas asli dan hanya 23 untuk kelas palsu, turut mempengaruhi hasil ini.

Secara keseluruhan, model memperoleh *accuracy* sebesar 0.94 yang menunjukkan bahwa 94% prediksi pada data uji sudah tepat. Metrik *weighted average* juga menunjukkan nilai *precision* 0,93, *recall* 0,94, dan *F1-score* 0,93, yang menggambarkan performa model secara keseluruhan dengan mempertimbangkan proporsi setiap kelas. Sementara itu, nilai *macro average* lebih rendah karena menghitung rata-rata antar kelas tanpa memperhatikan jumlah sampel, sehingga terpengaruh oleh performa yang lebih rendah pada kelas palsu.

Selain itu, dilakukan visualisasi confusion matrix untuk memahami distribusi prediksi:



Gambar 4.3 Heatmap Confusion Matrix

Pada gambar 4.3 menunjukkan *heatmap confusion matrix* model BERT

pada data uji. Dari matriks diatas dapat dilihat bahwa :

1. Kelas asli memiliki total data uji sebesar 209, dengan 202 ulasan yang diklasifikasikan secara benar dan terdapat 7 ulasan data yang salah diklasifikasikan sebagai ulasan palsu
2. Kelas palsu memiliki total data uji sebesar 23, dengan 15 ulasan yang diklasifikasikan secara benar sebagai ulasan palsu dan terdapat 8 ulasan data yang salah diklasifikasikan sebagai ulasan asli

Terlihat bahwa jumlah prediksi pada kelas asli jauh lebih dominan dibandingkan kelas palsu, yang menggambarkan distribusi data validasi yang memang tidak seimbang. Meskipun proses *training* menggunakan *oversampling* untuk menyeimbangkan data, hasil prediksi pada data uji tetap mengikuti proporsi asli, sehingga jumlah prediksi benar untuk kelas palsu tampak lebih sedikit. Pada diagonal utama terlihat bahwa model memiliki kemampuan yang baik dalam mengenali kelas asli, sementara performa untuk kelas palsu masih terbatas karena sedikitnya sampel pada kelas tersebut. *Confusion matrix* ini memberikan gambaran visual mengenai distribusi prediksi per kelas dan menegaskan bahwa ketimpangan pada data validasi memengaruhi jumlah prediksi benar pada kelas palsu, meskipun model tetap menunjukkan performa yang cukup stabil.

Berdasarkan hasil evaluasi, model BERT berkinerja sangat baik dalam mengkategorikan ulasan asli, namun masih berkinerja buruk dalam kelas ulasan palsu. Model ini dapat mengidentifikasi ulasan autentik dengan tingkat akurasi dan konsistensi yang tinggi, sebagaimana dibuktikan oleh nilai *precision*, *recall*, dan *F1-score* yang tinggi untuk kelas autentik. Di sisi lain, nilai *recall* dan *F1-score* yang relatif rendah menunjukkan bahwa kelas palsu berkinerja lebih buruk. Hal ini menunjukkan bahwa algoritma masih sering gagal mengidentifikasi ulasan palsu, kemungkinan besar sebagai akibat dari kuantitas data uji yang tidak merata. *Confusion matrix* memperkuat temuan ini dengan menunjukkan bahwa prediksi untuk kelas palsu jauh lebih sedikit dibandingkan kelas asli, meskipun proses *training* telah menggunakan *oversampling* untuk menyeimbangkan data. Ketidakseimbangan ini

menyebabkan model lebih cenderung memprediksi kelas mayoritas, sehingga kemampuan generalisasi untuk kelas minoritas (ulasan palsu) terbatas. Analisis ini menunjukkan bahwa meskipun model memiliki akurasi keseluruhan tinggi (0,94), peningkatan performa untuk kelas minoritas masih diperlukan.

4.10 Implementasi *User Interface*

4.10.1 Integrasi model menggunakan *streamlit*

Tahap integrasi dilakukan dengan menghubungkan model BERT dengan antarmuka pengguna melalui *framework streamlit*. Komponen antarmuka ini berfungsi untuk mempermudah proses interaksi sehingga pengguna tidak perlu memahami detail teknis model. Pengguna cukup memasukkan teks ulasan, lalu sistem akan memanggil fungsi prediksi secara otomatis. Model, *tokenizer*, dan *label encoder* yang telah disimpan dimuat kembali setiap kali tombol “Deteksi” ditekan.

Hasil prediksi kemudian ditampilkan secara visual melalui indikator warna yang informatif, yaitu hijau untuk label “ASLI” dan merah untuk label “PALSU”. Penyajian visual ini membantu pengguna dalam memahami hasil klasifikasi dengan cepat tanpa meninjau detail teknis yang berada di balik proses prediksi. Integrasi model BERT dengan *streamlit* memungkinkan sistem berjalan secara *real-time*, responsif, dan mudah digunakan sehingga meningkatkan pengalaman pengguna dalam menguji keaslian teks ulasan pada aplikasi yang dikembangkan.

4.10.2 Cara akses aplikasi *streamlit*

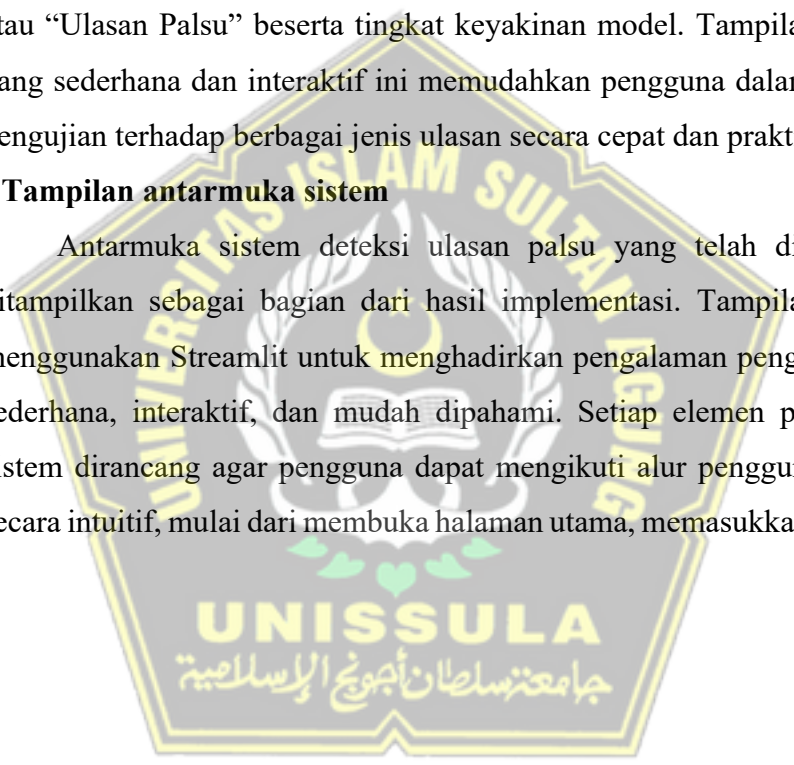
Aplikasi deteksi ulasan palsu yang dikembangkan menggunakan *framework streamlit* dapat dijalankan secara lokal melalui *anaconda*. Untuk memulai aplikasi, pengguna perlu membuka *anaconda prompt* dan menavigasi ke direktori tempat file *app.py* berada. Setelah itu, aplikasi dapat dijalankan menggunakan perintah `streamlit run app.py`. Ketika perintah dijalankan, *streamlit* akan melakukan proses inialisasi dan menyiapkan server lokal yang berfungsi sebagai antarmuka aplikasi. Proses ini

memungkinkan pengguna untuk menjalankan aplikasi secara interaktif tanpa perlu melakukan konfigurasi server secara manual.

Setelah aplikasi berhasil dijalankan, *streamlit* biasanya akan membuka browser secara otomatis. Namun, jika tidak terbuka, pengguna dapat mengakses aplikasi melalui URL default `http://localhost:8501`. Pada halaman utama aplikasi, pengguna dapat memasukkan teks ulasan yang ingin dianalisis, kemudian menekan tombol proses untuk menjalankan prediksi. Sistem akan menampilkan hasil klasifikasi berupa kategori “Ulasan Asli” atau “Ulasan Palsu” beserta tingkat keyakinan model. Tampilan antarmuka yang sederhana dan interaktif ini memudahkan pengguna dalam melakukan pengujian terhadap berbagai jenis ulasan secara cepat dan praktis.

4.10.3 Tampilan antarmuka sistem

Antarmuka sistem deteksi ulasan palsu yang telah dikembangkan ditampilkan sebagai bagian dari hasil implementasi. Tampilan ini dibuat menggunakan Streamlit untuk menghadirkan pengalaman penggunaan yang sederhana, interaktif, dan mudah dipahami. Setiap elemen pada halaman sistem dirancang agar pengguna dapat mengikuti alur penggunaan aplikasi secara intuitif, mulai dari membuka halaman utama, memasukkan teks ulasan,



hingga memperoleh hasil prediksi dari model.



Deteksi Ulasan Palsu

Cara Penggunaan:

1. Masukkan teks ulasan di bawah ini
2. Klik tombol "Deteksi Ulasan"
3. Aplikasi akan menampilkan apakah ulasan tersebut ASLI atau PALSU

Masukkan teks ulasan: ?

Contoh: 'Barang ini kece abis, pengiriman cepat banget!'

Deteksi Ulasan

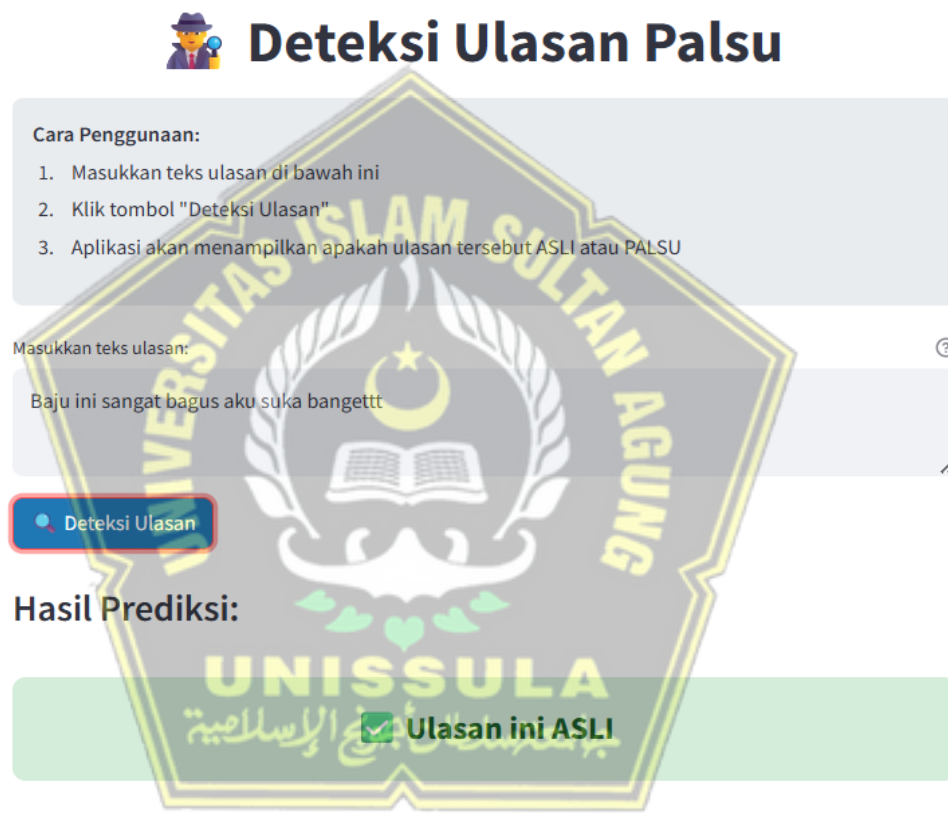


Gambar 4.4 Halaman Awal Sistem

Pada gambar 4.4 ditampilkan halaman awal dari aplikasi Deteksi Ulasan Palsu. Di bagian atas terdapat judul 'Deteksi Ulasan Palsu' yang menegaskan fungsi utama aplikasi. Tepat di bawah judul, terdapat kotak panduan penggunaan aplikasi yang berisi langkah-langkah sederhana untuk membantu pengguna memahami cara menjalankan sistem. Langkah-langkahnya dijelaskan secara singkat, yaitu pengguna diminta untuk memasukkan teks ulasan pada kolom yang tersedia, kemudian menekan tombol 'Deteksi Ulasan', dan setelah itu aplikasi akan menampilkan hasil apakah ulasan tersebut ASLI atau PALSU.

Selanjutnya, di bagian tengah halaman tersedia kolom input teks tempat pengguna dapat mengetikkan ulasan yang ingin diuji. Pada kolom tersebut juga ditampilkan contoh teks ulasan sebagai panduan bagi pengguna,

sehingga memudahkan dalam memahami format input yang diharapkan. Tepat di bawah kolom input terdapat sebuah tombol deteksi berwarna biru dengan ikon kaca pembesar, yang berfungsi untuk menjalankan proses analisis ulasan menggunakan model BERT yang telah dilatih sebelumnya. Setelah tombol ditekan, sistem akan melakukan pemrosesan teks dan menampilkan hasil prediksi.

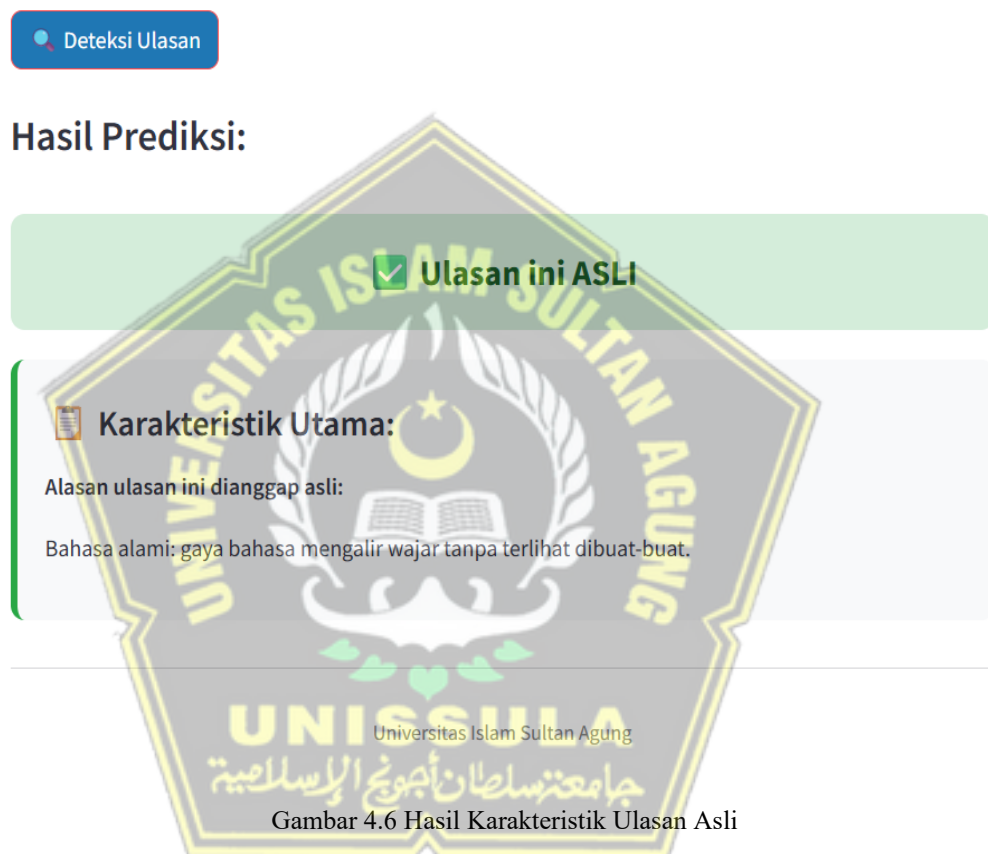


Universitas Islam Sultan Agung

Gambar 4.5 Hasil Prediksi Asli

Pada gambar 4.5 menampilkan hasil setelah pengguna memasukkan teks ulasan ke dalam sistem. Dalam contoh ini, pengguna mengetikkan ulasan 'Baju ini sangat bagus aku suka bangettt' pada kolom input teks yang tersedia. Setelah menekan tombol 'Deteksi Ulasan', sistem secara otomatis memproses masukan teks menggunakan model BERT yang telah dilatih untuk

mendeteksi ulasan palsu. Hasil prediksi kemudian ditampilkan pada bagian bawah dengan label ‘Hasil Prediksi’. Sistem memberikan keluaran berupa indikator visual berwarna hijau dengan ikon tanda centang dan keterangan ‘Ulasan ini ASLI’, yang menunjukkan bahwa teks ulasan yang dimasukkan terdeteksi sebagai ulasan asli.



Gambar 4.6 Hasil Karakteristik Ulasan Asli

Selain memberikan hasil prediksi, pada gambar 4.6 sistem juga menampilkan karakteristik ulasan asli yang sesuai dengan ulasan teks yang di input untuk membantu pengguna memahami alasan klasifikasi tersebut. Beberapa karakteristik ulasan asli yang bisa saja muncul antara lain:

- Detail dan spesifik, berisi pengalaman nyata seperti kualitas produk, pelayanan, atau kondisi pengiriman.
- Konsistensi dengan produk atau jasa, di mana isi ulasan sesuai dengan konteks produk yang dibeli.

- Bahasa alami, dengan gaya penulisan yang mengalir wajar tanpa terlihat dibuat-buat.
- Keberimbangan opini, menyebutkan kelebihan sekaligus kekurangan produk secara objektif.
- Mengandung informasi verifikatif, seperti foto produk, waktu penggunaan, atau kondisi barang saat diterima.

Namun karakteristik diatas akan muncul salah satunya saja sesuai dengan ulasan yang di input kedalam sistem.



Gambar 4.7 Hasil Prediksi Palsu

Pada gambar 4.7 merupakan hasil tampilan setelah pengguna memasukkan teks ulasan ke dalam sistem. Dalam contoh ini, pengguna mengetikkan ulasan ‘Baju bagus’ pada kolom input teks yang tersedia. Setelah menekan tombol ‘Deteksi Ulasan’, sistem secara otomatis memproses

masukan teks menggunakan model BERT yang telah dilatih untuk mendeteksi ulasan palsu. Hasil prediksi kemudian ditampilkan pada bagian bawah dengan label 'Hasil Prediksi'. Sistem memberikan keluaran berupa indikator visual berwarna merah dengan ikon tanda silang dan keterangan 'Ulasan ini PALSU', yang menunjukkan bahwa teks ulasan yang dimasukkan terdeteksi sebagai ulasan palsu.



Gambar 4.8 Hasil Karakteristik Ulasan Palsu

Selain menampilkan hasil klasifikasi, pada gambar 4.8 sistem juga menampilkan karakteristik dari ulasan palsu sebagai bentuk penjelasan tambahan sesuai dengan ulasan yang di masukan kedalam sistem. Karakteristik yang bisa saja muncul umumnya meliputi:

- Pengulangan kata atau frasa berlebihan sehingga teks tampak tidak natural.
- Adanya *filler* atau jeda yang tidak wajar sehingga mengganggu alur baca.
- Struktur kalimat panjang dan kompleks, berbeda dari ulasan asli yang umumnya lebih ringkas.

- Bersifat *one-sided*, yaitu hanya menyoroti sisi positif atau negatif tanpa penilaian seimbang.
- Mengandung *hiperbola*, seperti klaim ‘terbaik sepanjang masa’ atau ‘sangat buruk’.
- Menggunakan deskripsi generik tanpa menyertakan detail spesifik mengenai produk.
- Berorientasi promosi atau harga, seperti menekankan diskon dan ajakan membeli secara berlebihan.
- Mengandung emosi ekstrem, seperti pujian berlebihan atau kritik keras yang manipulatif.
- Memiliki jumlah kata atau kalimat yang sedikit sehingga minim detail pengalaman nyata.
- Menunjukkan rasio kata bermakna rendah, sehingga informasi yang disampaikan kurang substantif.
- Hanya membahas merek atau penjual (*brand-only reviews*) tanpa menyinggung produk secara spesifik.
- Mengandung informasi yang tidak relevan (*off-topic*), yaitu membicarakan hal di luar konteks produk yang diulas.

Namun karakteristik diatas akan muncul salah satunya saja sesuai dengan ulasan yang di input kedalam sistem.

Berdasarkan hasil implementasi yang ditunjukkan pada gambar 4.4 hingga gambar 4.8, sistem deteksi ulasan palsu berbasis BERT telah berhasil diintegrasikan ke dalam antarmuka pengguna yang sederhana dan mudah digunakan. Alur penggunaan sistem disusun secara jelas, mulai dari proses memasukkan teks ulasan hingga penampilan hasil prediksi, sehingga sistem dapat dioperasikan oleh pengguna tanpa memerlukan pemahaman teknis yang mendalam.

Penggunaan indikator warna hijau untuk ulasan asli dan warna merah untuk ulasan palsu memberikan penegasan visual yang jelas. Penyajian visual ini membantu pengguna memahami hasil klasifikasi dengan cepat dan intuitif. Selain itu, penampilan karakteristik ulasan asli dan ulasan palsu berfungsi sebagai bentuk penjelasan terhadap hasil klasifikasi, sehingga pengguna dapat mengetahui alasan

di balik keputusan model dalam menentukan kategori ulasan.

Keberadaan penjelasan karakteristik tersebut tidak hanya meningkatkan kejelasan dan keterbukaan sistem, tetapi juga mendorong pengguna untuk lebih kritis dalam menilai keaslian ulasan pada platform *e-commerce*. Dengan demikian, sistem yang dikembangkan tidak hanya berfungsi sebagai alat klasifikasi otomatis, tetapi juga sebagai sarana edukatif yang membantu meningkatkan pemahaman pengguna terhadap ciri-ciri ulasan palsu.



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, telah berhasil dibangun sebuah sistem *Deteksi Ulasan Palsu* pada produk kategori *fashion* di platform Tiktok Shop by Tokopedia menggunakan model BERT (*Bidirectional Encoder Representations from Transformers*). Dataset diperoleh melalui proses *web scraping* dari platform Tiktok Shop by Tokopedia dan kemudian melalui tahap pelabelan serta pra-pemrosesan sebelum digunakan dalam pelatihan model. Penelitian ini berhasil memanfaatkan representasi kontekstual BERT untuk memahami ulasan secara lebih mendalam. Dengan kemampuan BERT menangkap konteks dua arah dalam teks, sistem dapat mengenali nuansa bahasa dan makna kalimat pada ulasan produk, sehingga mendeteksi ulasan palsu tidak hanya berdasarkan kata kunci, tetapi juga pola bahasa dan konteks yang terkandung dalam ulasan.

Hasil pelatihan menunjukkan performa terbaik pada epoch ke-4 dengan nilai akurasi *training* sebesar 94,12%, meskipun pada *epoch* yang lebih tinggi mulai muncul indikasi *overfitting*. Evaluasi menggunakan metrik klasifikasi juga menunjukkan hasil yang konsisten, yaitu *precision* 0,96, *recall* 0,97, *F1-score* 0,96, dan akurasi 0,94. Secara keseluruhan, penelitian ini berhasil menyelesaikan permasalahan yang diidentifikasi, yaitu mendeteksi ulasan palsu pada kategori *fashion* menggunakan model BERT. Diharapkan, sistem ini dapat membantu konsumen dalam membuat keputusan pembelian yang lebih tepat serta membantu penjual menjaga reputasi dan kepercayaan terhadap produk mereka di platform *e-commerce*.

Dengan demikian, sistem ini memberikan manfaat praktis bagi pengguna dan pengelola *e-commerce* dalam meningkatkan transparansi dan integritas ulasan produk. Meskipun demikian, selama pelatihan model muncul indikasi *overfitting* pada *epoch* tinggi, yang menunjukkan bahwa

model cenderung menyesuaikan diri terlalu kuat dengan data pelatihan. Pengelolaan *overfitting* melalui teknik regulasi, peningkatan jumlah data, atau metode lain menjadi aspek penting yang perlu diperhatikan pada penelitian selanjutnya.

5.2 Saran

Penelitian ini masih memiliki beberapa keterbatasan, terutama terkait masalah *overfitting* pada jumlah *epoch* yang lebih tinggi. Oleh karena itu, penelitian selanjutnya disarankan untuk melakukan pengembangan pada beberapa aspek. Seperti ukuran dan keragaman dataset perlu ditingkatkan dengan mengumpulkan lebih banyak ulasan dari berbagai kategori produk, sehingga model dapat belajar dengan lebih baik dan mampu melakukan generalisasi pada data baru. Kemudian penggunaan model LLM yang lebih baik, seperti RoBERTa, ALBERT, atau *GPT-based models*, dapat menjadi alternatif untuk membandingkan performa dengan BERT. Selain itu, integrasi sistem ini ke dalam aplikasi berbasis *web* atau *mobile* juga dapat dikembangkan agar hasil penelitian dapat lebih bermanfaat secara praktis bagi konsumen maupun penjual di platform *e-commerce*. Dengan adanya pengembangan pada aspek-aspek tersebut, diharapkan sistem deteksi ulasan palsu ini dapat semakin akurat, efisien, dan memberikan manfaat nyata dalam mendukung transparansi informasi di ekosistem perdagangan digital di Indonesia.

DAFTAR PUSTAKA

- , D. M. U., -, D. A. U., and -, D. Y. K. (2025). A Comprehensive Overview of Large Language Models (LLMs). *International Journal For Multidisciplinary Research*, 7(1). <https://doi.org/10.36948/ijfmr.2025.v07i01.34609>
- Abri, F., Gutierrez, L. F., Namin, A. S., Jones, K. S., and Sears, D. R. W. (2020a). *Fake Reviews Detection through Analysis of Linguistic Features*. 1–11. <http://arxiv.org/abs/2010.04260>
- Abri, F., Gutierrez, L. F., Namin, A. S., Jones, K. S., and Sears, D. R. W. (2020b). Linguistic Features for Detecting Fake Reviews. *Proceedings - 19th IEEE International Conference on Machine Learning and Applications, ICMLA 2020*, 352–359. <https://doi.org/10.1109/ICMLA51294.2020.00063>
- Alamsyah, H., Cahyana, Y., and Pratama, A. R. (2023). Deteksi Fake Review Menggunakan Metode Support Vector Machine dan Naïve Bayes Di Tokopedia. *Jutisi : Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi*, 12(2), 585. <https://doi.org/10.35889/jutisi.v12i2.1222>
- Awalina, A., Bachtiar, F. A., and Indriati, I. (2022). Klasifikasi Ulasan Palsu Menggunakan Borderline Over Sampling (BOS) dan Support Vector Machine (SVM) (Studi Kasus : Ulasan Tempat Makan). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 9(2), 419–426. <https://doi.org/10.25126/jtiik.2022925692>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners -- special version. *Conference on Neural Information Processing Systems (NeurIPS 2020)*, *NeurIPS*, 1–25.
- Chang, K. H. (2023). Natural Language Processing: Recent Development and Applications. *Applied Sciences (Switzerland)*, 13(20), 10–11. <https://doi.org/10.3390/app132011395>
- Elmogy, A. M., Tariq, U., Ibrahim, A., and Mohammed, A. (2021). Fake Reviews

- Detection using Supervised Machine Learning. *International Journal of Advanced Computer Science and Applications*, 12(1), 601–606. <https://doi.org/10.14569/IJACSA.2021.0120169>
- Fatima, N., Mento, F., Zanforlin, A., Smargiassi, A., Torri, E., Perrone, T., and Demi, L. (2023). Human-to-AI Interrater Agreement for Lung Ultrasound Scoring in COVID-19 Patients. *Journal of Ultrasound in Medicine*, 42(4), 843–851. <https://doi.org/10.1002/jum.16052>
- Fauziya, R. A. H., and Isa, M. (2024). Pengaruh Fake Buyer Dan Fake Review Terhadap Purchase Intention Produk Fashion Pada Marketplace Shopee Dengan Trust Sebagai Variabel Intervening. *Journal of Management : Small and Medium Enterprises (SMEs)*, 17(3), 1225–1242. <https://doi.org/10.35508/jom.v17i3.19288>
- Hanum, A. R., Zetha, I. A., Putri, S. C., Wulandari, R. A., Andina, S. P., Fajrina, J. N., and Yudistira, N. (2024). Analisis Kinerja Algoritma Klasifikasi Teks Bert dalam Mendeteksi Berita Hoaks. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 11(3), 537–546. <https://doi.org/10.25126/jtiik.938093>
- Hayati, A. N., Marzuki, A. A. A., and Firmanditya, N. (2023). Ulasan Palsu Di Platform Digital: Perlindungan Hukum Bagi Konsumen Dan Pelaku Usaha. *Jurnal Masyarakat Indonesia*, 49(1), 123–133.
- Karimah, N., and Baita, A. (2024). *Multi-Aspect Sentiment Analysis Pada Review Film Menggunakan Metode Bidirectional Encoder Representations From Transformers (BERT) Multi-Aspect Sentiment Analysis of Film Review Using Bidirectional Encoder Representations from Transformers (BERT)*. 13(148). <https://doi.org/10.34010/komputika.v13i1.11098>
- Katampanan, S. E., Althon Pongtuluran, and Astriwati Biringkanae. (2023). Pengaruh Pengalaman Belanja Online Dan Kepercayaan Terhadap Minat Beli Ulang Di Tiktok Shop by Tokopedia (Studi Kasus Mahasiswa Manajemen Universitas Kristen Indonesia Toraja). *Jurnal Bisnis Kolega*, 9(2), 102–118. <https://doi.org/10.57249/jbk.v9i2.132>
- Khoirotulmuadiba Purifyregalia, Khothibul Umam, Nur Cahyo Hendro Wibowo, and Maya Rini Handayani. (2025). Detecting Fake Reviews in E-Commerce:

- A Case Study on Shopee Using Support Vector Machine and Random Forest. *Journal of Applied Informatics and Computing*, 9(3), 955–965. <https://doi.org/10.30871/jaic.v9i3.9514>
- Kim, J. M., Park, K. K. cheon, and Mariani, M. M. (2023). Do online review readers react differently when exposed to credible versus fake online reviews? *Journal of Business Research*, 154(January 2022), 113377. <https://doi.org/10.1016/j.jbusres.2022.113377>
- Kirom, M. I. (2024). Hukum Fake Review Dalam Transaksi Jual Beli Online (Studi Toko Online Shopee Arina_Fashion24). *Jurnal Hukum*.
- Kumar, R., Mukherjee, S., and Rana, N. P. (2024). Exploring Latent Characteristics of Fake Reviews and Their Intermediary Role in Persuading Buying Decisions. *Information Systems Frontiers*, 26(3), 1091–1108. <https://doi.org/10.1007/s10796-023-10401-w>
- Landis, J. R., and Koch, G. G. (2020). *The Measurement of Observer Agreement for Categorical Data Published by: International Biometric Society Stable URL: <https://www.jstor.org/stable/2529310>*. 33(1), 159–174.
- Mohawesh, R., Xu, S., Tran, S. N., Ollington, R., Springer, M., Jararweh, Y., and Maqsood, S. (2021). Fake Reviews Detection: A Survey. *IEEE Access*, 9, 65771–65802. <https://doi.org/10.1109/ACCESS.2021.3075573>
- Mudding, A. A. (2024). Mengungkap Opini Publik: Pendekatan BERT-based-caused untuk Analisis Sentimen pada Komentar Film. *Journal of System and Computer Engineering (JSCE)*, 5(1), 36–43. <https://doi.org/10.61628/jsce.v5i1.1060>
- Mulyati, Y., and Gesitera, G. (2020). Pengaruh Online Customer Review terhadap Purchase Intention dengan Trust sebagai Intervening pada Toko Online Bukalapak di Kota Padang. *Jurnal Maksipreneur: Manajemen, Koperasi, Dan Entrepreneurship*, 9(2), 173. <https://doi.org/10.30588/jmp.v9i2.538>
- Pooja, K., and Upadhyaya, P. (2024). What makes an online review credible? A systematic review of the literature and future research directions. In *Management Review Quarterly* (Vol. 74, Issue 2). Springer International Publishing. <https://doi.org/10.1007/s11301-022-00312-6>

- Priyono, M. B., and Dian Permata Sari. (2023). Dampak Aplikasi Tiktok Dan Tiktok Shop by Tokopedia Terhadap UMKM Di Indonesia. *Jurnal Ilmiah Wahana Pendidikan*, 9(September), 497–506. [https://repository.uin-suska.ac.id/15231/7/7.BAB II_201884KOM.pdf](https://repository.uin-suska.ac.id/15231/7/7.BAB%20II_201884KOM.pdf)
- Putu, N., Saraswati, V. D., Yudistira, N., and Adikara, P. P. (2023). Analisis Sentimen terhadap Perundungan Siber pada Twitter menggunakan Algoritma Bidirectional Encoder Representations from Transformer (BERT). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 7(2), 909–916. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12345>
- Salminen, J., Mustak, M., Jung, S. G., Makkonen, H., and Jansen, B. J. (2025). Decoding deception in the online marketplace: enhancing fake review detection with psycholinguistics and transformer models. *Journal of Marketing Analytics*, 0123456789. <https://doi.org/10.1057/s41270-025-00393-8>
- Stefanovitch, N., and Piskorski, J. (2023). Holistic Inter-Annotator Agreement and Corpus Coherence Estimation in a Large-scale Multilingual Annotation Campaign. *EMNLP 2023 - 2023 Conference on Empirical Methods in Natural Language Processing, Proceedings*, 71–86. <https://doi.org/10.18653/v1/2023.emnlp-main.6>
- Tran, V. D., Nguyen, M. D., and Luong, L. A. (2022). The effects of online credible review on brand trust dimensions and willingness to buy: Evidence from Vietnam consumers. *Cogent Business and Management*, 9(1). <https://doi.org/10.1080/23311975.2022.2038840>
- Vidya Chandradev, I Made Agus Dwi Suarjaya, and I Putu Agung Bayupati. (2023). Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT. *Jurnal Buana Informatika*, 14(02), 107–116. <https://doi.org/10.24002/jbi.v14i02.7244>