

**ANALISIS SENTIMEN PADA KINERJA PELATIH TIM NASIONAL
SEPAKBOLA INDONESIA SHIN TAE YONG DENGAN PERSEPSI
PENGGUNA SOSIAL MEDIA MENGGUNAKAN METODE *BERT*
(*BIDIRECTIONAL ENCODER REPRESENTATIONS FROM
TRANSFORMERS*)**

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:

HIDAYATUL MUAWANAH

32602100052

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNOLOGI INDUSTRI

UNIVERSITAS ISLAM SULTAN AGUNG

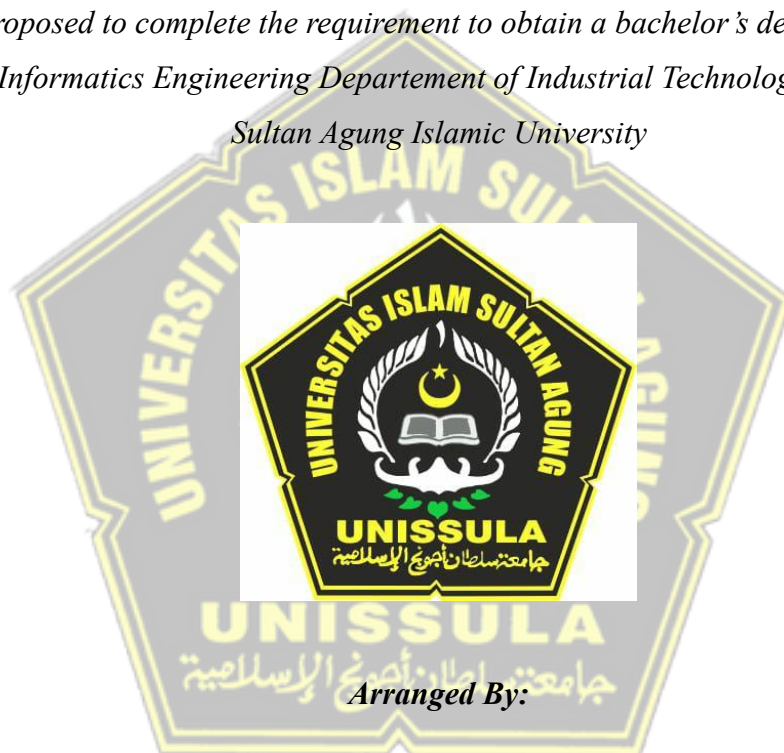
SEMARANG

2025

FINAL PROJECT

**SENTIMENT ANALYSIS ON THE PERFORMANCE OF THE
INDONESIAN NATIONAL FOOTBALL TEAM COACH SHIN TAE YONG
WITH THE PERCEPTION OF SOCIAL MEDIA USERS USING THE BERT
(BIDIRECTIONAL ENCODER REPRESENTATIONS FROM
TRANSFORMERS) METHOD**

*Proposed to complete the requirement to obtain a bachelor's degree (S1)
at Informatics Engineering Departement of Industrial Technology Faculty
Sultan Agung Islamic University*



Arranged By:

HIDAYATUL MUAWANAH

32602100052

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

2025

LEMBAR PENGESAHAN
TUGAS AKHIR

ANALISIS SENTIMEN PADA KINERJA PELATIH TIM NASIONAL
SEPAKBOLA INDONESIA SHIN TAE YONG DENGAN PERSEPSI
PENGGUNA SOSIAL MEDIA MENGGUNAKAN METODE *BERT*
(*BIDIRECTIONAL ENCODER REPRESENTATIONS FROM
TRANSFORMERS*)

HIDAYATUL MUAWANAH
NIM 32602100052

Telah dipertahankan di depan tim penguji ujian sarjana tugas akhir
Program Studi Teknik Informatika
Universitas Islam Sultan Agung
Pada tanggal : 27 Februari 2025

TIM PENGUJI UJIAN SARJANA :

Moch. Taufik, ST., MIT
NIDN. 0622037502
(Ketua Penguji)

11-03-2025

Bagus S.W.P., S.Kom, M.Cs
NIDN. 1027118801
(Anggota Penguji)

11-03-2025

Sam Farisa C.H., ST., M.Kom
NIDN. 0628028602
(Pembimbing)

11-03-2025

Semarang, 13 Maret 2025

Mengetahui,
Kaprodik Teknik Informatika
Universitas Islam Sultan Agung



Moch. Taufik, ST., MIT
NIDN. 0622037502

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Hidayatul Muawanah

NIM : 32602100052

Judul Tugas Akhir : ANALISIS SENTIMEN PADA KINERJA PELATIH TIM NASIONAL SEPAKBOLA INDONESIA SHIN TAE YONG DENGAN PERSEPSI PENGGUNA SOSIAL MEDIA MENGGUNAKAN METODE *BERT (BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS)*

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, 27 Februari 2025

Yang Menyatakan



Hidayatul Muawanah

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Hidayatul Muawanah

NIM : 32602100052

Program Studi : Teknik Informatika

Fakultas : Teknologi industri

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul :
ANALISIS SENTIMEN PADA KINERJA PELATIH TIM NASIONAL
SEPAKBOLA INDONESIA SHIN TAE YONG DENGAN PERSEPSI
PENGGUNA SOSIAL MEDIA MENGGUNAKAN METODE *BERT*
(*BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS*)

Menyctujui 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, 27 Februari 2025

Yang Menyatakan



Hidayatul Muawanah

KATA PENGANTAR

Dengan mengucapkan rasa syukur alhamdulillah atas kehadiran Allah SWT yang telah memberikan Rahmat dan karunianya kepada penulis, sehingga dapat menyelesaikan Laporan Proposal Tugas Akhir dengan judul “Analisis Sentimen pada Kinerja Pelatih Tim Nasional Sepakbola Indonesia Shin Tae Yong dengan Persepsi pengguna Sosial Media menggunakan Metode Bert (*Bidirectional Encoder Representations From Transformers*)” ini 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.

Tugas Akhir ini disusun dan dibuat dengan adanya bantuan dari berbagai pihak, materi maupun teknis, oleh karena itu saya selaku penulis mengucapkan terima kasih kepada:

1. Rektor UNISSULA Bapak Prof. Dr. H. Gunarto, S.H., M.H yang mengizinkan penulis menimba ilmu dikampus ini.
2. Dekan Fakultas Teknologi Industri Ibu Dr. Ir. Novi Marlyana, S.T., M.T., IPU., ASEAN. Eng.
3. Dosen Pembimbing penulis Bapak Sam Farisa Chaerul Haviana, S.T., M.Kom yang memberikan banyak nasehat dan saran.
4. Orang tua dan keluarga penulis yang menjadi *support system* dan mengizinkan untuk menyelesaikan laporan ini.
5. Para sahabat, teman-teman yang telah memberikan begitu banyak bantuan, semangat, inspirasi, pengambilan data dan diskusi progres penyusunan Tugas Akhir.

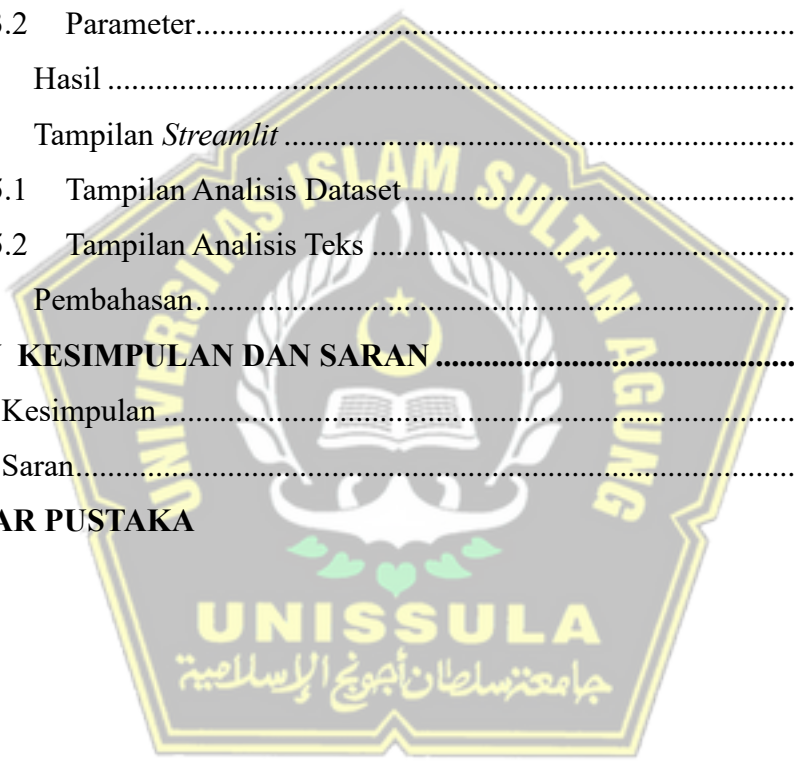
Semarang, 27 Februari 2025

Hidayatul Muawanah

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN TUGAS AKHIR	iii
SURAT PERNYATAAN KEASLIAN TUGAS AKHIR.....	iv
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR.....	ix
DAFTAR TABEL	x
ABSTRAK	xi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	2
1.3 Pembatasan Masalah	2
1.4 Tujuan.....	2
1.5 Manfaat	3
1.6 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI.....	4
2.1 Tinjauan Pustaka	4
2.2 Dasar Teori	14
2.2.1 <i>Natural Language Processing (NLP)</i>	14
2.2.2 Analisis Sentimen	15
2.2.3 Transformers	15
2.2.4 <i>BERT (Bidirectional Encoder Representations from Transformers)</i>	17
2.2.5 IndoBERT	19
BAB III METODE PENELITIAN	20
3.1 Metode Penelitian.....	20
3.2 Studi Literatur	20
3.3 Pengumpulan Data	21
3.4 Perancangan Implementasi Sistem.....	21

3.5	Pembuatan Sistem	24
BAB IV HASIL DAN ANALISIS PENELITIAN.....		25
4.1	Persiapan Data.....	25
4.2	<i>Preprocessing</i> Data	26
4.2.2	<i>Cleaning</i>	26
4.2.2	<i>Oversampling</i>	27
4.3	<i>Optimasi</i> Model.....	28
4.3.1	<i>Split</i> Data.....	28
4.3.2	Parameter.....	28
4.4	Hasil	29
4.5	Tampilan <i>Streamlit</i>	31
4.5.1	Tampilan Analisis Dataset.....	31
4.5.2	Tampilan Analisis Teks	34
4.6	Pembahasan.....	37
BAB V KESIMPULAN DAN SARAN		38
5.1	Kesimpulan	38
5.2	Saran.....	38
DAFTAR PUSTAKA		



DAFTAR GAMBAR

Gambar 2. 1 Arsitektur <i>Transformers</i>	16
Gambar 2. 2 <i>Pre-training</i> dan <i>Fine Tuning</i> BERT.....	18
Gambar 3. 1 <i>Flowchart</i> Alur Penelitian	20
Gambar 3. 2 <i>Flowchart</i> Alur Sistem	21
Gambar 3. 3 <i>Flowchart</i> Alur Kerja Sistem	24
Gambar 4. 1 Hasil gabungan data	26
Gambar 4. 2 Grafik <i>Train Loss</i> dan <i>Validation Loss</i>	29
Gambar 4. 3 <i>Confusion Matrix</i>	30
Gambar 4. 4 Tampilan Awal <i>Streamlit</i>	32
Gambar 4. 5 Hasil Prediksi Dataset	33
Gambar 4. 6 Distribusi Sentimen.....	33
Gambar 4. 7 Tampilan Analisis Teks.....	34
Gambar 4. 8 Tampilan Hasil Sentimen Positif.....	34
Gambar 4. 9 Tampilan Hasil Sentimen Negatif	35
Gambar 4. 10 Tampilan Hasil Sentimen Netral	35
Gambar 4. 11 Tampilan Hasil Sentimen Kalimat non Shin Tae-yong	36
Gambar 4. 12 Tampilan Hasil Sentimen Kalimat non Shin Tae-yong.....	36

DAFTAR TABEL

Tabel 2. 1 Tinjauan Pustaka	5
Tabel 4. 1 hasil data sesudah dan sebelum <i>cleaning</i>	27
Tabel 4. 2 Jumlah data sebelum dan sesudah <i>oversampling</i>	28
Tabel 4. 3 <i>Hyperparameter</i>	28
Tabel 4. 4 Matrik Evaluasi	31



ABSTRAK

Sepak bola memiliki popularitas tinggi di Indonesia, sehingga kinerja tim nasional menjadi sorotan publik. Sejak dilatih oleh Shin Tae Yong pada 2020, tim nasional mencatat berbagai pencapaian yang meningkatkan ekspektasi masyarakat. Hal ini mendorong banyak opini publik yang tersebar di media sosial, baik dalam bentuk apresiasi maupun kritik. Penelitian ini dilakukan untuk memahami sentimen yang berkembang di masyarakat terhadap kinerja Shin Tae Yong menggunakan BERT versi Indo yaitu IndoBERT. Data dari media sosial diklasifikasikan ke dalam tiga kategori sentiment yaitu positif, negatif, dan netral. Hasil analisis mengungkap bahwa sebagian besar masyarakat memberikan respon positif terhadap kinerja Shin Tae Yong, meskipun kritik juga masih cukup banyak. Dari 1.500 data yang dianalisis, model prediksi menghasilkan 583 sentimen positif, 566 negatif, dan 351 netral. Model IndoBERT yang digunakan dalam penelitian ini mencapai akurasi sebesar 93% menunjukkan efektivitasnya dalam memahami konteks bahasa Indonesia. Penelitian ini diimplementasikan dalam bentuk sistem berbasis Streamlit, sehingga memudahkan pengguna dalam menganalisis sentimen secara interaktif. Penelitian ini diharapkan dapat memberikan wawasan mengenai kepuasan publik terhadap kinerja Shin Tae Yong serta menjadi referensi dalam pemanfaatan analisis sentimen berbasis BERT untuk mengevaluasi opini publik di bidang olahraga.

Kata kunci: Analisis Sentimen, BERT, IndoBERT, Media Sosial, Kepuasan Publik

ABSTRACT

Football has high popularity in Indonesia, so the performance of the national team is in the public spotlight. Since being coached by Shin Tae Yong in 2020, the national team has recorded various achievements that have raised people's expectations. This has encouraged a lot of public opinion to spread on social media, both in the form of appreciation and criticism. This research was conducted to understand the growing sentiment in society regarding Shin Tae Yong's performance using the Indo version of BERT, namely IndoBERT. Data from social media is classified into three sentiment categories, namely positive, negative and neutral. The results of the analysis reveal that the majority of people responded positively to Shin Tae Yong's performance, although there was still quite a lot of criticism. Of the 1,500 data analyzed, the prediction model produced 583 positive, 566 negative and 351 neutral sentiments. The IndoBERT model used in this research achieved an accuracy of 93% indicating its effectiveness in understanding the Indonesian language context. This research was implemented in the form of a Streamlit-based system, making it easier for users to analyze sentiment interactively. This research is expected to provide insight into public satisfaction with Shin Tae Yong's performance and become a reference in using BERT-based sentiment analysis to evaluate public opinion in the sports sector.

Keywords: Sentiment Analysis, BERT, IndoBERT, Social Media, Public Satisfaction

BAB I PENDAHULUAN

1.1 Latar Belakang

Sepak bola diminati oleh berbagai kalangan, mulai dari anak-anak hingga kalangan dewasa. Kegiatan nonton bersama atau nobar menjadi bukti kebersamaan lintas generasi mempersatukan Indonesia. Menurut laporan Nielsen berjudul *The 2022 World Football Report*, Indonesia menempati peringkat ketiga di Asia sebagai negara dengan jumlah penggemar sepak bola terbanyak. Berdasarkan survei yang dilakukan, lebih dari setengah penduduk Indonesia, yaitu sekitar 69%, menggemari olahraga tersebut (Nielsen, 2022). Popularitas sepak bola di Indonesia yang begitu tinggi turut mendorong perhatian besar terhadap perkembangan tim nasional. Berdasarkan berita Tempo sejak Shin Tae-yong dipercaya menangani tim nasional Indonesia pada 2020 sampai 2024 telah meraih beberapa prestasi yang menaikkan ranking FIFA Indonesia sampai posisi 134. Setelah membawa Indonesia lolos ke babak ketiga kualifikasi Piala Dunia 2026 zona Asia, kemudian melakukan perpanjangan kontrak pada Juni 2024 (Bagja, 2024), Shin Tae Yong semakin mejadi perhatian publik.

Ekspektasi masyarakat semakin besar, ditambah dengan adanya pemberitaan dari beberapa media tentang Shin Tae Yong yang dibungkus secara berbeda menimbulkan beragam sudut pandang dan pendapat berujung kontroversial (Nugraha *dkk.*, 2024). Hal ini membuat masyarakat menjadikan Media sosial digunakan sebagai wadah untuk mengekspresikan opini. Pada Datareportal.com tercatat Indonesia memiliki sebanyak 139 juta orang, setara dengan sekitar 49,9% dari total populasi aktif menggunakan media sosial pada awal tahun 2024 (Kemp, 2024). Banyaknya masyarakat yang menyalurkan opini melalui media sosial menjadikan platform ini sebagai sumber data yang kaya untuk memahami pandangan publik. Dalam konteks ini, analisis sentimen berperan penting sebagai indikator untuk melihat kepuasan masyarakat terhadap kinerja Shin Tae Yong yang dikategorikan ke dalam sentimen positif, negatif, dan netral.

Penggunaan metode BERT dalam analisis sentimen menghasilkan akurasi sebesar 99% pada penelitian (Mas *dkk.*, 2021) dan akurasi sebesar 98% pada penelitian (Septian, Ivana Lucia Kharisma dan Kamdan, 2023). Hasil ini menunjukkan bahwa BERT memiliki kemampuan unggul dalam memahami konteks bahasa secara *bidirectional*, memungkinkan analisis yang lebih akurat pada teks dengan struktur kompleks. Kemampuan ini telah dibuktikan dalam penelitian (Devlin *dkk.*, 2019), bahwa BERT secara efektif menangkap hubungan antar kata dalam berbagai konteks, menjadikannya pilihan yang kuat untuk tugas-tugas seperti analisis sentimen. Dengan penelitian ini, analisis sentimen terhadap kinerja Shin Tae Yong di media sosial dapat dilakukan secara lebih akurat dan mendalam.

1.2 Perumusan Masalah

Bagaimana analisis sentimen diterapkan sebagai indikator untuk melihat kepuasan masyarakat terhadap kinerja pelatih tim nasional Indonesia Shin Tae Yong berdasarkan ulasan pengguna sosial media?

1.3 Pembatasan Masalah

1. Data yang digunakan didapatkan dari aplikasi youtube, x, dan instagram serta berbahasa Indonesia.
2. Data yang dikumpulkan mencakup periode dari tahun 2022 hingga 2024.
3. Pelabelan sentimen diklasifikasikan menjadi sentimen positif, negatif, dan netral.

1.4 Tujuan

Penelitian ini bertujuan untuk merancang sebuah sistem dan melakukan validasi untuk membuat analisis sentimen terhadap kinerja pelatih tim nasional Indonesia Shin Tae Yong berdasarkan ulasan pada sosial media menggunakan metode BERT.

1.5 Manfaat

Penelitian ini memiliki manfaat sebagai sistem untuk memprediksi kepuasan masyarakat dan memahami sentimen publik terhadap Shin Tae Yong yang bisa menjadi bahan evaluasi bagi PSSI dan pelatih berikutnya.

1.6 Sistematika Penulisan

Penyusunan laporan tugas akhir menerapkan tata cara penulisan yaitu:

BAB 1 : PENDAHULUAN

Bab ini memuat penjelasan mengenai latar belakang, perumusan masalah, pembatasan masalah, tujuan penelitian, dan metodologi penelitian, serta sistematika penulisan.

BAB 2 : TINJAUAN PUSTAKA DAN DASAR TEORI

Bab ini membahas penelitian-penelitian sebelumnya yang berkaitan serta dasar teori yang mendukung penelitian.

BAB 3 : METODE PENELITIAN

Bab ini menjelaskan tahapan penelitian secara sistematis, mulai dari pengumpulan data hingga tahap klasifikasi data.

BAB 4 : HASIL DAN ANALISIS PENELITIAN

Bab ini menampilkan temuan penelitian, khususnya analisis sentimen yang mengidentifikasi kategori positif, negatif atau netral dari teks ulasan menggunakan model BERT.

BAB 5 : KESIMPULAN DAN SARAN

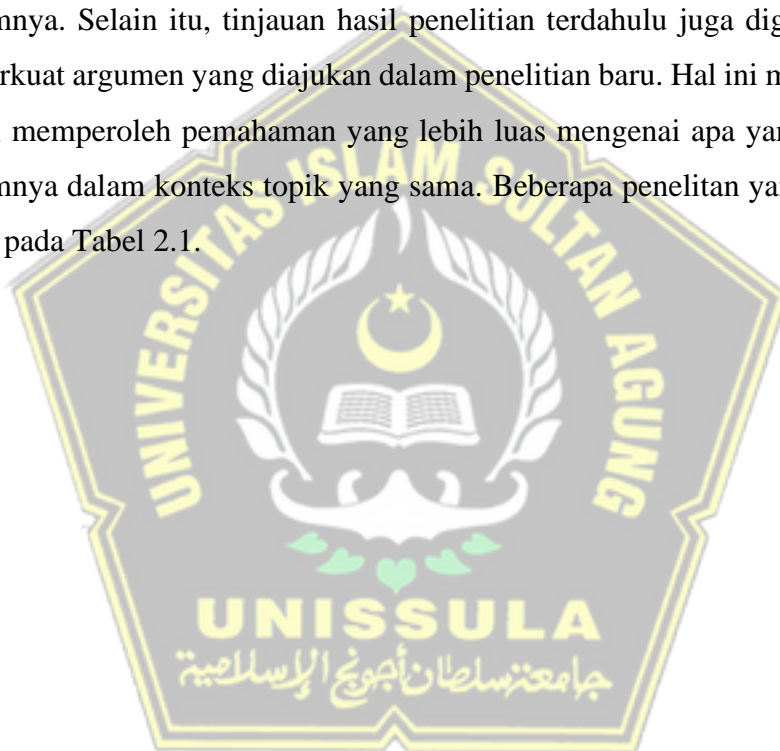
Bab terakhir memuat ringkasan keseluruhan penelitian dari tahap awal hingga akhir, serta memberikan saran bersarkan temuan yang diperoleh.

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Tinjauan hasil penelitian sebelumnya menjadi penopang penting dalam penelitian yang akan dilakukan. Di satu sisi, peninjauan tersebut menyediakan bahan perbandingan yang berharga terhadap penelitian yang telah ada sebelumnya, baik itu dalam mengidentifikasi kelebihan maupun kekurangan dari penelitian sebelumnya. Selain itu, tinjauan hasil penelitian terdahulu juga digunakan untuk memperkuat argumen yang diajukan dalam penelitian baru. Hal ini memungkinkan peneliti memperoleh pemahaman yang lebih luas mengenai apa yang telah dikaji sebelumnya dalam konteks topik yang sama. Beberapa penelitian yang serupa bisa terlihat pada Tabel 2.1.



Tabel 2. 1 Tinjauan Pustaka

No.	Judul	Penulis	Metode	Dataset	Hasil	Kekurangan
1.	Analisis Sentimen Customer Review Aplikasi Ruang Guru dengan Metode BERT (<i>Bidirectional Encoder Representations from Transformers</i>)	(Mas dkk., 2021)	BERT	Data berasal dari <i>Google Play Store</i> terdiri dari 10.418 data komentar ulasan.	Hasil akurasi mencapai 99%, dengan presisi mencapai 64,13%, sementara <i>recall</i> sebesar 60,51%.	Memiliki keterbatasan dalam menjaga konsistensi data yang diambil, belum ada eksplorasi penerapan algoritma ini pada platform lain atau dataset yang berbeda.
2.	Analisis Sentimen Terhadap Game Genshin Impact Menggunakan Bert	(Kusnadi dkk., 2021)	BERT	Data berasal dari <i>Google Play Store</i> sejumlah 12000.	Hasil dari penelitian nilai presisi tertinggi sebesar 0,86%, dengan <i>recall</i> mencapai 0,78% dan <i>f1-score</i> sebesar 0,82%.	Data yang digunakan kurang terstandarisasi, sehingga dapat memengaruhi keakuratan dan validitas analisis sentimen.

No.	Judul	Penulis	Metode	Dataset	Hasil	Kekurangan
3.	Analisis Sentimen Review Film Berbahasa Inggris dengan Pendekatan <i>Bidirectional Encoder Representations from Transformers</i>	(Putri, Adiwijaya dan Faraby, 2020)	BERT	Data diunduh dari situs website <i>cornelledu</i> sebanyak 2000 <i>review</i> .	Hasil penelitian ini memiliki akurasi yang relatif tinggi yaitu mencapai 73.7%.	Penelitian ini belum sepenuhnya mengenali karakteristik dataset yang digunakan. Hal ini terlihat dari saran yang menekankan pentingnya mengenali opini dan bagian penting dari dataset.
4.	Peningkatan Akurasi Klasifikasi Sentimen Ulasan Makanan Amazon dengan <i>Bidirectional Lstm dan Bert Embedding</i>	(Pasaribu, Kusrini dan Sudarmawan, 2020)	<i>Bidirectional Lstm dan Bert Embedding</i>	Dataset dari Anees et al. (2019) di <i>Kaggle</i> berisi 568.454 ulasan <i>Amazon Fine Food</i> .	Hasil pengujian mendapatkan akurasi 93 %.	Hanya menggunakan ukuran embedding sebesar 300, meskipun hasilnya sudah baik, tidak ada eksplorasi terhadap ukuran <i>embedding</i> yang lebih besar, seperti 500 atau lebih, yang mungkin dapat memberikan akurasi lebih

No.	Judul	Penulis	Metode	Dataset	Hasil	Kekurangan
						btinggi atau pola analisis yang lebih mendalam.
5.	Analisis Sentimen Review Hotel menggunakan Metode <i>Deep Learning BERT</i>	(Chandradev, Suarjaya dan Bayupati, 2023)	<i>SmallBERT</i>	Menggunakan dataset 515 ribu ulasan hotel dari <i>Kaggle</i> .	<i>SmallBERT</i> mencapai akurasi 91,40%, dengan presisi, recall, dan f1-score masing-masing sebesar 90,51%.	Hanya menggunakan <i>SmallBERT</i> dengan konfigurasi tertentu tanpa eksplorasi lebih lanjut terhadap variasi model, konfigurasi <i>fine-tuning</i> , atau kombinasi <i>hyperparameter</i> lainnya.
6.	Analisis Sentimen Ulasan Pengguna Aplikasi Joox pada Android menggunakan Metode <i>Bidirectional Encoder Representations from Transformers</i> (BERT)	(Lazuardi dan Juarna, 2023)	IndoBERT	Data tentang JOOX di <i>Google Play Store</i> yaitu data komentar sebanyak 10.000 data.	Hasil akurasi <i>F1-score</i> adalah 86%, 51%, dan 76% dari 990 data uji, sedangkan akurasi <i>baseline</i> sebesar 83%, 79%, dan 75%.	Penelitian belum melakukan pengujian berbagai <i>learning rate</i> untuk menemukan nilai terbaik. Hal ini dapat memengaruhi performa model.

No.	Judul	Penulis	Metode	Dataset	Hasil	Kekurangan
7.	Analisis Sentimen Pendapat Masyarakat Mengenai Vaksin Covid-19 Pada Media Sosial Twitter dengan <i>Robustly Optimized BERT Pretraining Approach</i>	(Zain dkk., 2021)	<i>Robustly Optimized BERT Pretraining Approach (RoBERTa)</i>	Dataset disaring menggunakan kata kunci yaitu "vaksin covid" dengan jumlah batasan sebanyak 109.202 tweet.	Penelitian ini menunjukkan bahwa model berhasil mencapai akurasi tinggi dengan rata-rata keseluruhan 95%. Selain itu, akurasi prediksi untuk label positif mencapai 84%, netral sebesar 97%, dan negatif sebesar 93%.	Distribusi kinerja yang tidak merata antar kelas akurasi untuk label positif (84%) lebih rendah dibandingkan label netral (97%) dan negatif (93%). Hal ini menunjukkan bahwa model kesulitan mengklasifikasikan data pada kelas positif, yang dapat disebabkan oleh jumlah data positif yang tidak cukup banyak atau fitur yang kurang representatif untuk kelas tersebut.

No.	Judul	Penulis	Metode	Dataset	Hasil	Kekurangan
8.	Analisis Sentimen Terhadap Minat Belajar pada Tayangan Acara CoC by Ruangguru berdasarkan <i>Tweets</i> menggunakan Metode NLP dan Model <i>BERT</i>	(Helmiyah dan Verdian, 2024)	BERT	Data dikumpulkan dengan kata kunci seperti " <i>Clash of Champions</i> ", " <i>CoC by Ruangguru</i> ", dan " <i>Ruangguru</i> " didapat sebanyak 351 <i>tweets</i> .	Hasil penelitian menunjukkan bahwa sentimen positif sebesar 85,27% dan sentimen negatif sebesar 3,99% menggunakan metode NPL. Serta sentimen positif 86,12% dan sentiment negatif 0% menggunakan model BERT.	Pengumpulan data yang terbatas hanya dari platform tertentu yaitu hanya dari X menyebabkan kurangnya representasi spektrum pendapat masyarakat secara menyeluruh. Penelitian belum sepenuhnya memanfaatkan <i>preprocessing</i> yang mendalam, yang bisa mencakup langkah-langkah seperti normalisasi, <i>lemmatization</i> , atau penghapusan <i>noise</i> pada data teks.

No.	Judul	Penulis	Metode	Dataset	Hasil	Kekurangan
9.	Analisis Sentimen Berbasis Aspek pada Ulasan Aplikasi Gojek	(Rahman, Pranatawijaya dan Sari, 2024)	BERT	Dataset diperoleh dari ulasan pengguna Gojek pada <i>Google Play Store</i> , mulai dari bulan Januari 2020 hingga Desember 2023 sebanyak 14.147 ulasan.	Hasil meraih akurasi tertinggi untuk analisis sentimen sebesar 96.67% dan aspek layanan sebesar 98.78%.	Meskipun penelitian menemukan tiga aspek utama (<i>user experience, service, dan payment</i>), tidak ada analisis mendalam mengenai kontribusi masing-masing aspek terhadap keseluruhan sentimen pengguna. Hal ini dapat mengurangi wawasan yang praktis untuk perbaikan aplikasi.
10.	Implementasi Metode <i>Bidirectional Encoder Representations from Transformers (BERT)</i> untuk Analisis	(Septian, Ivana Lucia Kharisma dan Kamdan, 2023)	IndoBERT	Data yang terkumpul sebanyak 1331 yang merupakan data komentar di	Hasil akurasi 98% dari data latih serta validasi akurasi sebesar 93% dengan data	Penelitian hanya berfokus pada dua kelas (positif dan negatif), tanpa mempertimbangkan kelas netral atau ulasan yang

No.	Judul	Penulis	Metode	Dataset	Hasil	Kekurangan
	Sentimen Komentar pengguna Aplikasi Dana di Instagram			Instagram akun resmi Dana.	latih dan data uji 70:30.	lebih kompleks. Hal ini dapat membatasi kemampuan model dalam mengidentifikasi variasi sentimen yang lebih halus.
11.	Analisis Sentimen Terhadap Pembelajaran Secara Daring Pasca Pandemi Covid-19 Menggunakan Metode IndoBERT	(Hidayat dan Pramudita, 2024)	IndoBERT	Data dikumpulkan melalui survei Google Form kepada mahasiswa tentang pengalaman perkuliahan online, dengan 35 opini terkumpul pada 3–10 September 2023.	Sistem klasifikasi opini dengan metode IndoBERT mampu mengelompokkan opini secara otomatis dikelompokkan dalam tiga kategori, yaitu positif, netral, dan negatif dengan	Penelitian selanjutnya dapat dikembangkan dengan menambah fitur dan visualisasi untuk penyajian data yang lebih sederhana, meningkatkan kualitas dataset dengan menambah data dan melakukan <i>pre-processing</i> yang lebih efisien, serta membandingkan model IndoBERT dengan metode klasifikasi lain guna

No.	Judul	Penulis	Metode	Dataset	Hasil	Kekurangan
					tingkat akurasi sebesar 87% (0,87).	menentukan pendekatan terbaik dalam analisis sentimen.
12.	Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan Indobert	(Merdiansah dan Ali Ridha, 2024)	IndoBERT	Penelitian ini menggunakan tiga kata kunci yaitu "motorlistrik", "mobillistrik", dan "kendaraanlistrik". Data yang dikumpulkan adalah 1.000 tweet terbaru.	Hasil penelitian ini mencapai akurasi 98% dengan menggunakan 5 <i>epoch</i> .	Penelitian ini terbatas pada data dari platform X (Twitter) dan jumlah data yang masih sedikit, sehingga hasilnya kurang mewakili opini secara luas. Selain itu, model IndoBERT belum dibandingkan dengan metode lain, dan kesulitan menangkap sentimen ambigu seperti sarkasme tetap menjadi tantangan.

No.	Judul	Penulis	Metode	Dataset	Hasil	Kekurangan
13.	Penerapan Metode Transfer Learning pada Indobert Untuk Analisis Sentimen Teks Bahasa Jawa Ngoko Lugu	(Prabowo dan Indra Sanjaya, 2024)	IndoBERT	Penelitian ini mengumpulkan 3.000 tweet berbahasa Jawa Ngoko Lugu dari platform media sosial Twitter sebagai data penelitian.	Hasil penelitian mengindikasikan bahwa model IndoBERT menunjukkan kinerja unggul dalam mendeteksi sentimen positif dan negatif pada teks Bahasa Jawa Ngoko Lugu, dengan akurasi sebesar 99% pada data pelatihan serta 90% pada data validasi dan pengujian.	Penelitian ini masih memiliki beberapa keterbatasan, seperti data yang hanya berasal dari Twitter, sehingga kurang mencerminkan opini dari platform lain. Selain itu, klasifikasi sentimen hanya terbatas pada dua kategori, tanpa mempertimbangkan sentimen netral. Jumlah data yang digunakan untuk pemodelan juga relatif kecil, dan proses pelabelan yang dilakukan secara manual berpotensi menyebabkan bias subjektif.

Beberapa penelitian yang telah disebutkan pada Tabel 2.1 dapat disimpulkan bahwa penggunaan BERT dan variannya, seperti IndoBERT, *RoBERTa*, dan *SmallBERT*, terbukti efektif dalam analisis sentimen. BERT memungkinkan pemahaman konteks dua arah dalam teks, membuatnya unggul dibandingkan metode konvensional. IndoBERT lebih akurat dalam menangani bahasa Indonesia, *RoBERTa* memiliki optimasi pretraining yang lebih efisien, dan *SmallBERT* menawarkan model lebih ringan dengan performa kompetitif. Namun, tantangan seperti ketidakseimbangan data, kurangnya eksplorasi *hyperparameter*, dan keterbatasan sumber data masih ditemukan. Oleh karena itu, penelitian ini diharapkan dapat memanfaatkan IndoBERT untuk memahami opini publik mengenai STY secara lebih akurat, mengatasi ketidakseimbangan data, serta mengeksplorasi *hyperparameter* optimal guna menghasilkan model analisis sentimen yang lebih representatif dan berkualitas.

2.2 Dasar Teori

2.2.1 *Natural Language Processing* (NLP)

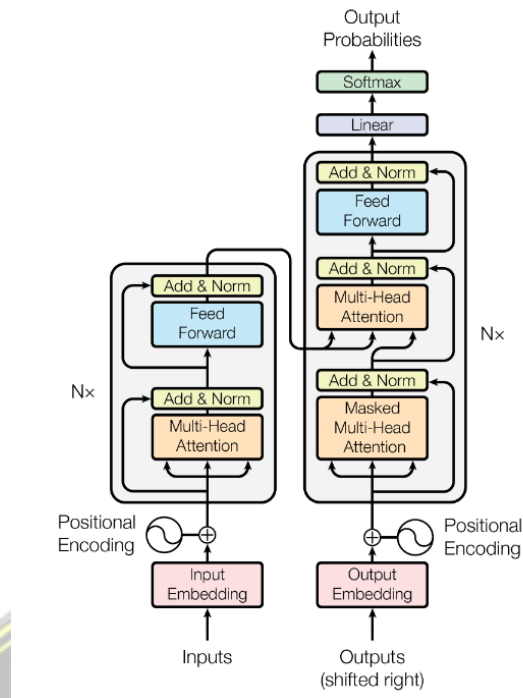
Natural Language Processing (NLP) merupakan bidang dalam ilmu komputer dan kecerdasan buatan yang berfokus pada kemampuan komputer dalam memahami, menganalisis, dan memproses bahasa manusia secara efisien. Dalam konteks bahasa Indonesia, NLP mencakup berbagai teknik dasar seperti *stemming*, *part-of-speech tagging*, dan analisis sentimen. Teknologi ini membantu dalam pengurangan kompleksitas teks dan peningkatan efisiensi pemrosesan data berbasis teks. Arsitektur dalam NLP melibatkan berbagai langkah seperti *preprocessing* data teks untuk membersihkan data, tokenisasi berfungsi untuk membagi teks ke dalam bagian-bagian yang lebih kecil, dan penerapan model untuk klasifikasi atau penggalian informasi. Dengan perkembangan teknik seperti pembelajaran mesin dan jaringan saraf, NLP mampu menangkap konteks kalimat yang lebih baik dibandingkan metode statistik tradisional. Teknik NLP telah digunakan secara luas, termasuk dalam analisis sentimen untuk memahami opini publik di media sosial (Amien, 2023).

2.2.2 Analisis Sentimen

Analisis sentimen merupakan metode untuk mengenali dan mengategorikan opini, emosi, atau sikap dalam teks ke dalam kategori positif, negatif, atau netral. Metode ini memanfaatkan teknik pemrosesan bahasa alami (NLP) untuk menganalisis data teks yang tidak terstruktur (Cindo dan Rini, 2019). Secara umum, analisis sentimen bertujuan untuk memahami bagaimana perasaan atau opini seseorang terhadap suatu entitas, baik itu produk, layanan, kebijakan, maupun isu tertentu. Selain itu, analisis ini juga membantu organisasi dalam mengambil keputusan strategis berdasarkan wawasan yang diperoleh dari data teks yang besar dan tidak terstruktur. Dengan kemajuan teknologi NLP dan pembelajaran mesin modern seperti *deep learning*, analisis sentimen terus berkembang untuk menangani tantangan seperti ambiguitas bahasa dan ironi dalam teks (Thomas, Yuliana dan Noviyanti. P, 2021).

2.2.3 Transformers

Berdasarkan penelitian yang dilakukan oleh (Vaswani *dkk.*, 2017) dalam paper berjudul "*Attention is All You Need*" mengenalkan tentang arsitektur model *transformers*. Model ini dirancang untuk mengatasi keterbatasan model sebelumnya seperti RNN dan LSTM, terutama dalam hal menangani urutan data yang panjang. Salah satu inovasi utama dalam *Transformers* adalah penggunaan mekanisme *self-attention*, yang memungkinkan model untuk mempertimbangkan hubungan antara kata-kata dalam kalimat secara bersamaan, sehingga meningkatkan pemahaman konteks. Berikut pada Gambar 2.1 menggambarkan arsitektur dari *Transformers*.



Gambar 2. 1 Arsitektur *Transformers*

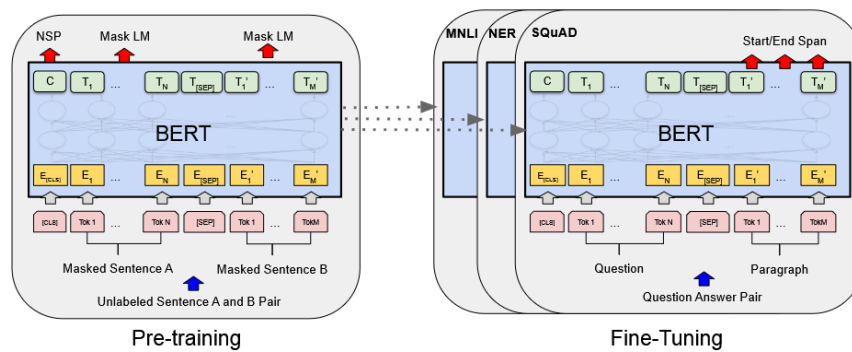
Arsitektur *Transformers* terdiri dari dua elemen utama, yaitu *encoder* dan *decoder*. *Encoder* memiliki peran untuk memproses *input* dan menghasilkan representasi kontekstual dari data tersebut. Setiap lapisan pada *encoder* memiliki dua sub-lapisan yaitu *self-attention mechanism* dan *feed forward neural network*. *Self-attention mechanism* bekerja dengan memperhatikan setiap kata dalam input untuk memahami konteks keseluruhan. Sementara itu, *feed-forward neural network* mengolah hasil dari *self-attention* dengan menerapkan transformasi non-linear agar informasi yang diperoleh lebih kompleks dan bermakna. Di sisi lain, *decoder* bertugas menghasilkan *output* berdasarkan representasi yang diberikan oleh *encoder*. *Decoder* juga memiliki dua sub-lapisan yang serupa dengan yang terdapat pada *encoder*, namun dengan penambahan bagian yang berfungsi untuk memperhatikan dan menghubungkan informasi dari *output encoder*. Mekanisme *self-attention* menjadi inti dari arsitektur ini, di mana setiap kata diubah menjadi tiga vektor: *query*, *key*, dan *value* untuk memahami hubungan antar kata dalam suatu kalimat.

Sistem menentukan seberapa penting setiap kata dengan mengalikan *query* dan *key*, lalu menerapkan fungsi *softmax* untuk menghasilkan bobot perhatian

dalam prosesnya. Bobot ini kemudian digunakan untuk menyesuaikan *value*, sehingga menghasilkan representasi akhir yang lebih bermakna. Keunggulan *Transformers* dibandingkan model sebelumnya mencakup kemampuan untuk melakukan paralelisasi, karena tidak bergantung pada urutan waktu, sehingga proses pelatihan dapat dilakukan lebih cepat. Selain itu, model ini mampu menangani konteks yang jauh satu sama lain dalam teks, serta fleksibel untuk diterapkan pada berbagai tugas NLP seperti terjemahan bahasa dan analisis sentimen. *Transformers* banyak diterapkan dalam berbagai aplikasi NLP serta bidang lainnya. Contoh terkenal dari penerapan *Transformers* adalah *BERT* (*Bidirectional Encoder Representations from Transformers*), yang memungkinkan pemahaman konteks dua arah; *GPT* (*Generative Pre-trained Transformer*), yang berfokus dalam pembuatan teks; serta *RoBERTa* dan *ALBERT*, yang merupakan variasi dari *BERT* yang dioptimalkan untuk performa lebih baik (Tanjung dkk., 2023).

2.2.4 *BERT (Bidirectional Encoder Representations from Transformers)*

BERT (Bidirectional Encoder Representations from Transformers) adalah model pemrosesan bahasa alami yang diperkenalkan oleh *Google AI* pada tahun 2018. Model ini telah merevolusi pemrosesan bahasa alami (NLP) dengan pendekatan baru dalam memahami konteks teks. *BERT* berbeda dari model sebelumnya karena menggunakan arsitektur *Transformer* dan menerapkan pelatihan *bidirectional*, sehingga model dapat memahami keterkaitan kata dalam kalimat dari dua arah, yaitu dari kiri ke kanan dan sebaliknya, dari kanan ke kiri. Struktur *BERT* dibangun di atas komponen *encoder* dari arsitektur *Transformers*. Pendekatan ini membuat *BERT* mampu memahami arti kata dalam pengertian yang lebih komprehensif, sehingga menghasilkan pemahaman yang lebih akurat terhadap teks. Dalam penelitian yang dilakukan oleh (Devlin dkk., 2019), *BERT* dilatih melalui dua tahap utama, yaitu proses *pre-training* dan *fine-tuning* yang bisa dilihat pada Gambar 2.2.



Gambar 2. 2 *Pre-training* dan *Fine Tuning* BERT

Selama *pre-training*, model dilatih dengan dua tugas utama yaitu *Masked Language Model* (MLM) dan *Next Sentence Prediction* (NSP). Pada MLM, beberapa kata dalam teks digantikan dengan token [MASK], lalu model berusaha menebak kata yang hilang berdasarkan konteks di sekitarnya. Pendekatan ini memungkinkan BERT memahami makna kata secara *bidirectional*, berbeda dari model sebelumnya yang hanya mempertimbangkan konteks dari satu arah. Selain itu, dalam tugas NSP, model diberikan dua kalimat dan harus menentukan apakah kalimat kedua merupakan kelanjutan dari kalimat pertama dalam teks aslinya. Hal ini bertujuan untuk mengajarkan model memahami hubungan antarkalimat, yang sangat penting dalam tugas pemahaman teks. Setelah tahap *pre-training* selesai, model dapat dioptimalkan untuk berbagai tugas khusus melalui proses *fine-tuning*. Pada tahap ini, BERT diadaptasi ke dalam berbagai tugas NLP dengan dataset berlabel yang lebih kecil, seperti klasifikasi teks (*Multi-Genre Natural Language Inference/ MNLI*), pengenalan entitas dalam teks (*Named Entity Recognition/ NER*), dan pencarian jawaban dalam teks (*Stanford Question Answering Dataset/ SQuAD*). Keunggulan utama BERT terletak pada kemampuannya memahami konteks secara *bidirectional*, yang membuatnya lebih unggul dalam menangkap makna kata dibandingkan dengan model NLP sebelumnya. Hal ini menjadikan BERT sebagai model yang sangat efisien dalam berbagai tugas pemrosesan bahasa alami (NLP), seperti analisis sentimen, klasifikasi teks, penerjemahan, hingga penjawaban pertanyaan.

2.2.5 IndoBERT

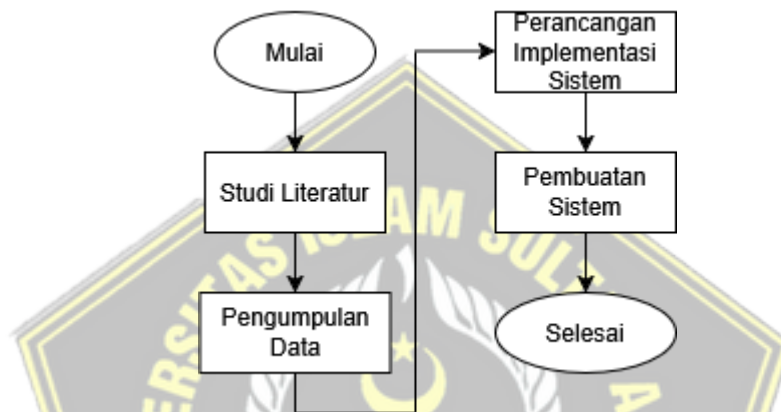
IndoBERT merupakan model BERT yang dikustomisasi untuk memahami bahasa Indonesia. Model ini dikembangkan untuk mengatasi keterbatasan model-model sebelumnya yang lebih banyak berfokus pada bahasa Inggris dan hanya menggunakan model multibahasa yang tidak optimal untuk bahasa Indonesia. IndoBERT dibangun dari awal dengan menggunakan dataset besar yang disebut INDOLEM, yang mencakup berbagai teks dalam bahasa Indonesia. Penelitian ini bertujuan untuk menciptakan model *monolingual* yang dapat memberikan performa lebih baik dalam beragam tugas pemrosesan bahasa alami (NLP), termasuk sentimen analisis dan pembuatan ringkasan teks (Wijaya, Romadhony dan Richasdy, 2023).

Dalam pengembangan IndoBERT, peneliti menerapkan teknik *pre-training* yang sama dengan *BERT*, termasuk *Masked Language Model* (MLM) dan *Next Sentence Prediction* (NSP). Teknik MLM memungkinkan model untuk memprediksi kata-kata yang hilang dalam kalimat dengan memperhatikan konteks dari kedua sisi, sedangkan NSP membantu model memahami hubungan antar kalimat. Hasil dari penelitian menunjukkan bahwa IndoBERT secara konsisten mengungguli model *Multilingual BERT* (*M-BERT*) dalam berbagai aplikasi NLP berbasis bahasa Indonesia. Keunggulan IndoBERT terletak pada kemampuannya untuk memahami nuansa dan konteks bahasa Indonesia secara lebih efektif, sehingga dapat diterapkan dalam beragam aplikasi seperti klasifikasi teks, analisis sentimen, ekstraksi informasi, penerjemahan bahasa, dan generasi teks. Dengan kemampuan ini, IndoBERT diharapkan dapat membantu penulis dalam mengolah data berbahasa Indonesia dengan lebih akurat dan efisien (Hidayat dan Nastiti, 2024).

BAB III METODE PENELITIAN

3.1 Metode Penelitian

Penelitian ini akan membangun sistem analisis sentimen terhadap kinerja Shin Tae Yong berdasarkan masyarakat pengguna sosial media menggunakan model BERT. Alur penelitian tertera pada Gambar 7.1.



Gambar 3. 1 Flowchart Alur Penelitian

3.2 Studi Literatur

Studi literatur dilakukan dengan mengumpulkan dan menelaah artikel serta jurnal ilmiah dari penelitian-penelitian terdahulu yang berhubungan dengan penulisan tugas akhir. Dari daftar pustaka yang dikaji, penelitian ini mengacu pada jurnal akademik dari berbagai institusi dan penerbit seperti *Elsevier*, *IEEE Xplore*, dan *Google Scholar*; serta jurnal seperti *Jurnal Buana Informatika*, *JATISI*, dan *Jurnal Komputer Terapan*, yang membahas analisis sentimen dan implementasi BERT. Selain itu, data pendukung diperoleh dari website seperti *Tempo*, *DataReportal*, dan *Nielsen* untuk memahami konteks publik terkait kinerja Shin Tae-yong. Referensi ini digunakan dalam penulisan tugas akhir sehingga membantu peneliti dalam menyelesaikan penelitian ini.

3.3 Pengumpulan Data

Peneliti melakukan scraping data dengan mengambil data dari media sosial youtube, X, dan instagram tentang Shin Tae Yong. *Scraping* data dilakukan menggunakan bahasa pemrograman *Python*. Jumlah data yang didapat dari masing-masing sosial media sebanyak 500 jadi jumlahnya 1500 data. Data disimpan dalam file csv dan kemudian dilakukan pelabelan secara manual dengan kategori pelabelannya yaitu positif, negatif, dan netral. Sehingga didapat 650 positif, 602 negatif, dan 248 netral.

3.4 Perancangan Implementasi Sistem

Pada tahap ini, metode analisis sentimen dirancang sesuai kebutuhan penelitian yaitu dengan menggunakan pendekatan berbasis bert, untuk klasifikasi sentimen. Perancangan model ini Bertujuan untuk mengklasifikasikan sentimen publik ke dalam kategori positif, negatif, atau netral. Proses perancangan alur sistem dapat dilihat pada Gambar 3.2.



Gambar 3. 2 *Flowchart* Alur Sistem

- Pelabelan dilakukan setelah data terkumpul, setiap data diberikan label sesuai dengan kategori yang diinginkan yaitu berupa sentimen "positif", "negatif", atau "netral". Pelabelan dilakukan secara manual.
- Tahap *preprocessing* yaitu tahap mengolah data untuk mendapatkan data mentah (data kotor) sehingga dapat diolah dalam pengujian data. Proses *preprocessing* melibatkan proses *cleansing*, *casefolding*, dan *tokenization*.

Tahap *preprocessing* data ini dilakukan untuk menghilangkan *noise* dan mempercepat waktu analisis sehingga mendapatkan hasil analisis yang lebih akurat.

- c. Segmentasi Data, data dibagi menjadi tiga bagian: data latih untuk melatih model, data validasi untuk penyesuaian parameter, dan data uji untuk menilai kinerja model.
- d. Klasifikasi metode IndoBERT yaitu IndoBERT bekerja untuk mengklasifikasi sentimen dengan memanfaatkan arsitektur *transformers*, yang dirancang untuk memahami hubungan antar kata dalam sebuah kalimat secara kontekstual. Proses dimulai dengan tokenisasi, di mana kalimat diubah menjadi token menggunakan *tokenizer* berbasis *subword* IndoBERT. Token spesial seperti [CLS] ditambahkan di awal sebagai representasi keseluruhan kalimat, dan [SEP] di akhir untuk menandai batas input. Setelah itu, token diubah menjadi vektor numerik melalui *embedding layer* yang menyertakan informasi tentang posisi kata dalam kalimat dan makna semantiknya. Mekanisme *self-attention* digunakan untuk menentukan hubungan antar kata dalam kalimat. Setiap kata diperbandingkan dengan kata-kata lain untuk memahami pengaruhnya. Sebagai contoh, dalam kalimat "Shin Tae-yong adalah pelatih hebat," kata "hebat" memiliki hubungan yang kuat dengan "pelatih" dan "Shin Tae-yong." Representasi kontekstual yang dihasilkan oleh *encoder* IndoBERT menangkap hubungan semacam ini untuk setiap kata, dan representasi CLS digunakan untuk merangkum informasi dari seluruh kalimat. Selanjutnya, representasi CLS diteruskan ke lapisan linear untuk menghasilkan skor sentimen, yang kemudian diubah menjadi probabilitas melalui fungsi *softmax*. Probabilitas ini menentukan kategori sentimen positif, negatif, atau netral dengan memilih kategori yang memiliki nilai tertinggi. Misalnya, sebuah komentar seperti "Shin Tae-yong membawa perubahan besar untuk Timnas!" akan menghasilkan probabilitas yang menunjukkan sentimen positif. Sebelumnya, model ini dikembangkan dengan sejumlah besar data komentar yang sudah diberi label (positif, negatif, atau netral). Ini

membantu IndoBERT belajar mengenali pola sentimen dalam berbagai kalimat.

- e. Hasil Klasifikasi, sistem memberikan hasil berupa prediksi sentimen untuk setiap data yang diuji.
- f. Evaluasi model dalam analisis sentimen bertujuan untuk mengukur kinerja model menggunakan beberapa metrik utama, seperti akurasi, presisi, *recall*, dan *f1-score*.
 1. Akurasi yaitu mengukur seberapa banyak prediksi yang benar dibandingkan dengan seluruh data.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

2. Presisi yaitu mengukur sejauh mana prediksi positif yang dihasilkan oleh model benar.

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (2)$$

3. *Recall* yaitu mengukur sejauh mana model dapat menangkap semua data positif yang sebenarnya.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

4. *F1-score* yaitu rata-rata harmonik dari presisi dan recall yang digunakan untuk menyeimbangkan keduanya.

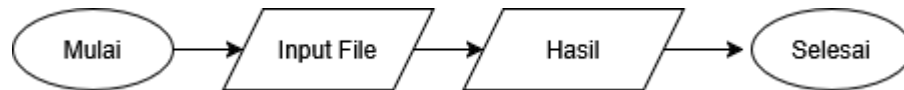
$$\text{F1-score} = 2 \times \frac{\text{presisi} \times \text{recall}}{\text{presisi} + \text{recall}} \quad (4)$$

Keterangan:

- TP (*True Positive*) = Prediksi positif yang sesuai dengan kondisi sebenarnya.
- TN (*True Negative*) = Prediksi negatif yang benar sesuai dengan data asli.
- FP (*False Positive*) = Prediksi positif yang ternyata salah.
- FN (*False Negative*) = Prediksi negatif yang seharusnya positif.

3.5 Pembuatan Sistem

Tahap ini merupakan implementasi dari model analisis sentimen yang telah dirancang sebelumnya. Sistem disambungkan menggunakan framework *Streamlit* dari file *app.py* yang didalamnya terdapat model yang telah disimpan. Alur kerja sistemnya dapat ditemukan pada Gambar 3.3.



Gambar 3. 3 *Flowchart* Alur Kerja Sistem

Berikut merupakan penjabaran *flowchart* berdasarkan tampilan *Streamlit* yang telah dibuat:

1. Mulai

Proses dimulai ketika pengguna mengakses aplikasi berbasis *Streamlit*.

2. Input Data

Pengguna memasukkan teks atau memilih dataset yang akan dianalisis. Kemudian sistem melakukan *preprocessing*, seperti pembersihan teks, tokenisasi, dan normalisasi agar data siap untuk diproses oleh model IndoBERT.

3. Hasil Prediksi

Sentimen yang terdeteksi ditampilkan di tampilan *Streamlit*, jika yang dimasukkan hanya teks maka hasil prediksinya hanya positif, negatif, dan netral. Jika dalam bentuk dataset hasil prediksinya berupa tampilan file yang ada kolom prediksinya dan bisa diunduh lalu ada visualisasi diagram *pie*.

4. Selesai

Proses dapat diulang dari tahap input jika pengguna ingin menganalisis teks lain.

BAB IV

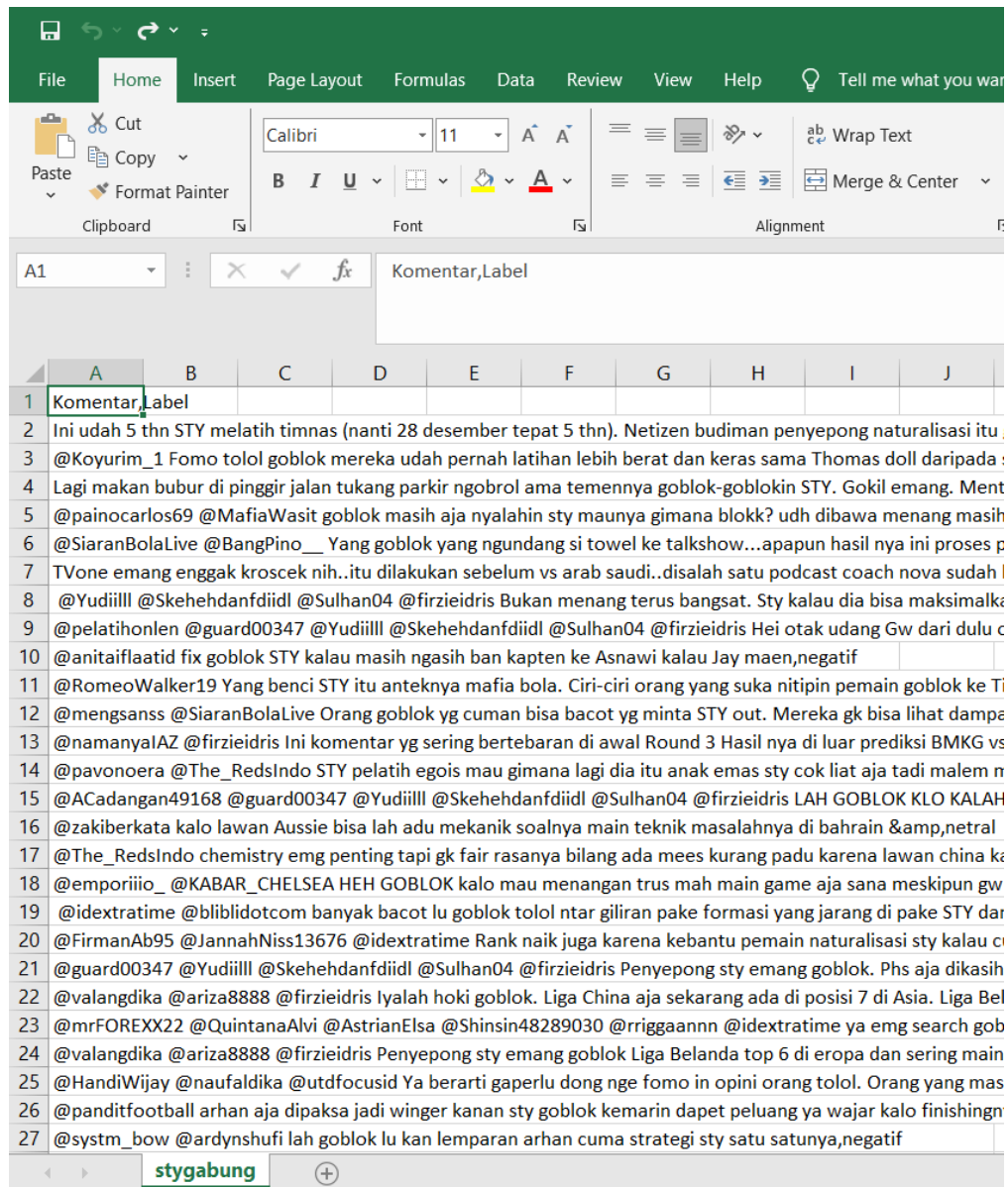
HASIL DAN ANALISIS PENELITIAN

4.1 Persiapan Data

Data dikumpulkan melalui teknik *scraping* dari tiga sosial media yaitu X, youtube, dan instagram. Jumlah data yang didapat dari masing-masing sosial media sebanyak 500 jadi jumlahnya 1500 data. Data disimpan dalam file csv dan kemudian dilakukan pelabelan secara manual dengan kategori pelabelannya yaitu positif, negatif, dan netral. Sehingga didapat 650 positif, 602 negatif, dan 248 netral.

Metode pengambilan data dalam penelitian ini dilakukan melalui teknik web scraping dengan memanfaatkan *tweet harvest*, sebuah alat yang dikembangkan oleh Helmi Satria. Proses ini menggunakan kata kunci "shin tae-yong" atau "sty" dalam Bahasa Indonesia, dengan batas pencarian sebanyak 500 data, serta mengharuskan penggunaan token dari akun X untuk mengakses dan mengunduh data yang dibutuhkan. Data yang diperoleh dari platform X ini berupa *tweet* yang mengandung kata kunci yang telah ditentukan.

Selain dari X, data juga dikumpulkan melalui platform *youtube* dengan menggunakan API *youtube*. Proses ini melibatkan penggunaan API *key* serta ID video yang diambil dari tautan *youtube* yang relevan dengan topik penelitian. Video yang dipilih berasal dari berbagai sumber, terutama *podcast* atau wawancara yang membahas Shin Tae-yong secara langsung, baik dalam kapasitasnya sebagai pelatih tim nasional Indonesia maupun dalam konteks lain yang berkaitan dengan karir dan kinerjanya. Sedangkan yang dari instagram dilakukan secara manual dengan mengambil komentar dari postingan saat terjadi pertandingan sepakbola. Data lalu digabung menjadi satu untuk dilakukan proses selanjutnya dapat terlihat pada Gambar 4.1.



Gambar 4. 1 Hasil gabungan data

4.2 Preprocessing Data

4.2.2 Cleaning

Proses ini dilakukan untuk membersihkan data dari teks yang masih memiliki emoji, *username*, *hashtag*, *link*, spasi berlebih, dan lainnya. Kemudian dilakukan pembersihan agar meningkatkan kualitas data dan siap untuk dilakukan analisis. Berikut pada Tabel 4.1 menampilkan contoh data sebelum dan setelah proses pembersihan.

oversampling karena sederhana dan efisien dilakukan pada dataset kecil. Pada Tabel 4.2 menunjukkan data label sebelum dan sesudah dilakukan *oversampling*.

Tabel 4. 2 Jumlah data sebelum dan sesudah *oversampling*

Label	Sebelum <i>Oversampling</i>	Sesudah <i>Oversampling</i>
Positif	650	650
Negatif	602	650
Netral	248	650

4.3 *Optimasi Model*

4.3.1 *Split Data*

Pembagian data dilakukan dengan rasio 80:20 yang merupakan salah satu standar pembagian umum. Jumlah data yang digunakan sebanyak 1560 sampel untuk training (80%), 195 sampel untuk validasi (10%), dan 195 sampel untuk testing (10%). Pembagian ini bertujuan agar model bisa belajar pola dengan cukup data sehingga nanti bisa memberikan gambaran mengenai kemampuan model dalam menggeneralisasi data baru.

4.3.2 *Parameter*

Untuk memastikan model dapat bekerja secara optimal, diperlukan pengaturan *hyperparameter* yang sesuai dengan spesifikasi perangkat serta karakteristik data yang digunakan. *Hyperparameter* ini memengaruhi kinerja model dalam proses pelatihan, termasuk efisiensi komputasi dan akurasi hasil. Pengaturan *hyperparameter* dapat diamati pada Tabel 4.3.

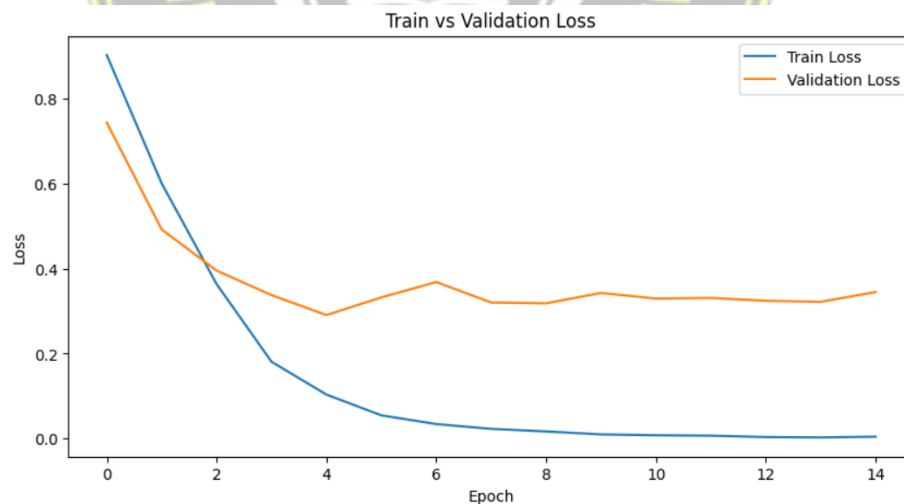
Tabel 4. 3 *Hyperparameter*

<i>Hyperparameter</i>	Nilai
<i>Max Sequence Length</i>	512
<i>Batch Size</i>	8
<i>Num Workers</i>	2
<i>Optimizer</i>	Adam
<i>Learning Rate</i>	3e-6 (0,000003)
<i>Epoch</i>	15

Pemilihan *hyperparameter* dalam penelitian ini disesuaikan dengan kebutuhan model dan keterbatasan perangkat. *Max sequence length* 512 digunakan agar model dapat menangani teks panjang secara optimal. *Batch size* 8 dan *num workers* 2 dipilih karena menyesuaikan dengan kapasitas laptop yang digunakan. *Optimizer* Adam dipilih karena mampu menyesuaikan *learning rate* secara adaptif, sementara *learning rate* $3e-6$ digunakan agar *fine-tuning* berjalan stabil tanpa kehilangan informasi dari model *pre-trained*. *Epoch* 15 ditetapkan untuk memastikan model memiliki cukup waktu belajar tanpa *overfitting*.

4.4 Hasil

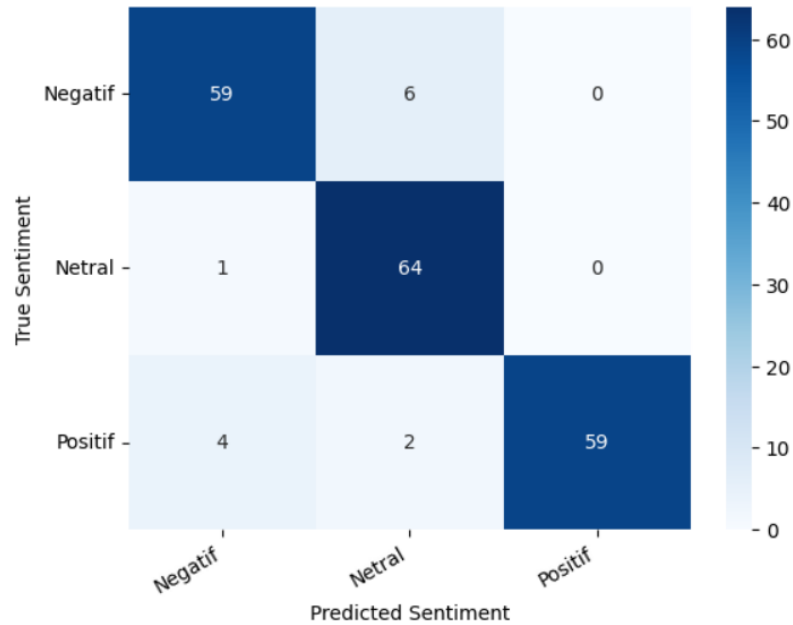
Sebelum dilakukan evaluasi keseluruhan, model diuji terlebih dahulu dengan metode random sampling pada 10 data acak. Dari hasil pengujian ini, model berhasil memberikan prediksi yang benar pada 9 dari 10 data, menunjukkan akurasi awal yang cukup baik sebelum dilakukan analisis lebih lanjut. Hasil pelatihan model menampilkan perubahan nilai *train loss* dan *valid loss* selama beberapa *epoch* terdapat grafik dalam Gambar 4.2.



Gambar 4. 2 Grafik *Train Loss* dan *Validation Loss*

Pada awal pelatihan, *train loss* memiliki nilai yang sangat tinggi, tetapi secara bertahap menurun seiring peningkatan *epoch*, menandakan perbaikan kinerja model dalam menyesuaikan diri dengan data latih. Sementara itu, *validation loss* juga mengalami penurunan pada awalnya, tetapi kemudian cenderung stabil setelah beberapa *epoch*. Berdasarkan grafik tersebut menunjukkan bahwa proses pelatihan

telah berjalan dengan baik. Untuk mengevaluasi lebih lanjut performa model, dilakukan analisis *confusion matrix* yang tersaji dalam Gambar 4.3.



Gambar 4. 3 *Confusion Matrix*

Confusion matrix menunjukkan performa model dalam mengklasifikasikan sentimen berdasarkan data uji. Nilai *true* positif untuk kategori negatif sebanyak 59, netral 64, dan positif 59, menunjukkan bahwa model berhasil mengategorikan mayoritas data dengan tepat. Namun, terdapat beberapa kesalahan klasifikasi, seperti 6 data negatif yang diprediksi sebagai netral dan 4 data positif yang salah diklasifikasikan sebagai negatif. Setelah melihat performa model melalui *confusion matrix*, evaluasi lebih lanjut dilakukan dengan menghitung metrik *precision*, *recall*, dan *f1-score*. Data hasilnya tercantum pada Tabel 4.4.

Tabel 4. 4 Matrik Evaluasi

	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
Negatif	0,92	0,91	0,91	65
Netral	0,89	0,98	0,93	65
Positif	1,00	0,91	0,95	65
<i>accuracy</i>			0,93	195
<i>macro avg</i>	0,94	0,93	0,93	195
<i>weighted avg</i>	0,94	0,93	0,93	195

Model mencapai akurasi keseluruhan sebesar 0.93, yang menunjukkan bahwa 93% dari seluruh data uji berhasil diklasifikasikan dengan benar. Selain itu, nilai *macro average* dan *weighted average* untuk *precision*, *recall*, dan *f1-score* juga berada pada angka 0.93, yang menunjukkan bahwa model bekerja dengan performa yang optimal dalam mengklasifikasikan ketiga kategori sentimen.

4.5 Tampilan *Streamlit*

4.5.1 Tampilan Analisis Dataset


Setelah pemodelan dilakukan menggunakan model IndoBERT dan disimpan, langkah selanjutnya adalah membuat sistem berbasis *framework Streamlit* untuk memudahkan pengguna dalam melakukan analisis sentimen. Tampilan awal bisa dilihat pada Gambar 4.4.

Analisis Sentimen Kinerja Shin Tae Yong berdasarkan Sosmed

 Analisis Dataset  Analisis Teks

Upload Dataset untuk Analisis Sentimen

Upload CSV



Drag and drop file here

Limit 200MB per file • CSV

Browse files

Gambar 4. 4 Tampilan Awal *Streamlit*

Pada bagian atas, termuat judul utama yang menampilkan tujuan utama sistem. Di bawahnya, terdapat dua tab navigasi, yaitu Analisis Dataset dan Analisis Teks, yang memungkinkan pengguna untuk memilih metode analisis yang diinginkan. Pada tab Analisis Dataset berfungsi untuk mengunggah dataset dalam format CSV sebagai input untuk analisis sentimen. Terdapat area unggah file yang memungkinkan pengguna untuk *drag and drop* file atau memilih file secara manual melalui tombol *Browse files*. Batas ukuran file yang dapat diunggah adalah 200MB. File yang diunggah merupakan dataset asli berjumlah 1500 data berisi mengenai komentar-komentar publik tentang kinerja Shin Tae Yong sebagai pelatih sepakbola Indonesia. Setelah dataset diunggah, pengguna mengklik *enter* kemudian menunggu *running* sampai keluar hasil prediksi analisis sentimennya yang ditunjukkan dalam Gambar 4.5.

stycleaning(1).csv 228.5KB

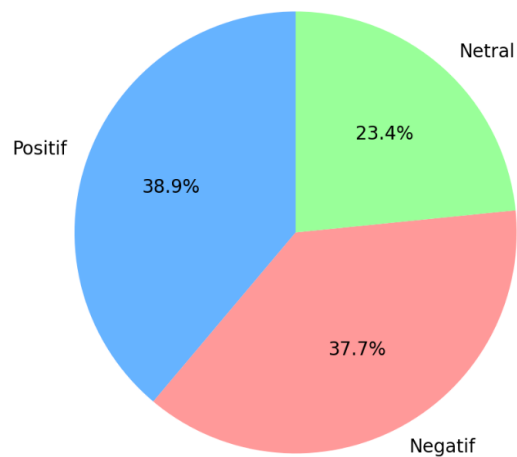
×

Hasil Analisis Sentimen ↻

	komentar	label	pred
0	ini udah 5 thn sty melatih timnas nanti 28 desember tepat 5 thn netizen budiman per	negatif	Negatif
1	fomo tolol goblok mereka udah pernah latihan lebih berat dan keras sama thomas de	negatif	Negatif
2	lagi makan bubur di pinggir jalan tukang parkir ngobrol ama temennya goblokgoblok	netral	Netral
3	goblok masih aja nyalahin sty maunya gimana blokk udh dibawa menang masih ae p	positif	Positif
4	yang goblok yang ngundang si towel ke talkshowapapun hasil nya ini proses perbaik	netral	Netral
5	tvone emang enggak kroscek nihitu dilakukan sebelum vs arab saudidialah satu poc	positif	Positif
6	bukan menang terus bangsat sty kalau dia bisa maksimaln laga bahrain dan gak ek	negatif	Negatif
7	hei otak udang gw dari dulu objektif pada sty kalau dia bagus gw puji dan kalau gak t	netral	Netral
8	fix goblok sty kalau masih ngasih ban kapten ke asnawi kalau jay maen	negatif	Negatif
9	yang benci sty itu anteknya mafia bola ciriciri orang yang suka nitipin pemain goblok	netral	Netral

Gambar 4. 5 Hasil Prediksi Dataset

Tampilan Gambar 4.5 menunjukkan file hasil prediksi analisis sentimen, kemudian ada hasil visualisasi *pie chart* agar lebih mudah bagi para pembaca mengetahui seberapa banyak distribusi sentimennya yang tersaji pada Gambar 4.6.



Distribusi Sentimen

Positif: 583 data

Negatif: 566 data

Netral: 351 data

Gambar 4. 6 Distribusi Sentimen

4.5.2 Tampilan Analisis Teks

Tab Analisis Teks memungkinkan pengguna untuk menganalisis sentimen dari teks yang dimasukkan secara langsung. Terdapat sebuah kotak *input* tempat pengguna dapat mengetik atau menempelkan teks yang ingin dianalisis. Setelah teks dimasukkan, pengguna dapat mengklik tulisan “Analisis” untuk memproses teks dan mendapatkan hasil analisis sentimen, untuk mengkategorikan sentimen dalam teks menjadi positif, negatif, atau netral. Tampilan analisis teks dapat ditemukan pada Gambar 4.7.

Analisis Sentimen Kinerja Shin Tae Yong berdasarkan Sosmed

 Analisis Dataset  Analisis Teks

Analisis Sentimen Teks

Masukkan teks:

Analisis

Gambar 4. 7 Tampilan Analisis Teks

Setelah memasukkan teks tunggu *running* hingga keluar hasil sentimennya sebagaimana ditunjukkan pada Gambar 4.8.

Masukkan teks:

sty itu orangnya disiplin jadi kita beruntung dapet coach kayak dia timnas bisa berkembang sejauh ini

Analisis

Hasil Sentimen: Positif

Gambar 4. 8 Tampilan Hasil Sentimen Positif

Pada Gambar 4.8 dapat disimpulkan bahwa streamlit melakukan analisis sentimen berdasarkan teks yang dimasukkan dan memberikan hasil berupa kategori sentimen positif terhadap Shin Tae-yong.

Masukkan teks:

aduh pulang aja deh coach ke negara lu dasar miskin taktik

Analisis

Hasil Sentimen: Negatif

Gambar 4. 9 Tampilan Hasil Sentimen Negatif

Pada Gambar 4.9 dapat disimpulkan bahwa streamlit melakukan analisis sentimen berdasarkan teks yang dimasukkan dan memberikan hasil berupa kategori sentimen negatif terhadap Shin Tae-yong.

Masukkan teks:

menurut gua malam ini agak mengecewakan sih kenapa eliano ga dimainin malah arhan terus tapi balik lagi pelatih pastinya udah tau strategi yang terbaik buat timnas sebagai penonton cuma bisa dukung aja

Analisis

Hasil Sentimen: Netral

Gambar 4. 10 Tampilan Hasil Sentimen Netral

Pada Gambar 4.10 dapat disimpulkan bahwa streamlit melakukan analisis sentimen berdasarkan teks yang dimasukkan dan memberikan hasil berupa kategori sentimen netral terhadap Shin Tae-yong.

Masukkan teks:

Hujan deras yang mengguyur Boyolali, Jawa Tengah, pada Sabtu (8/3/2025) petang, menyebabkan banjir melanda tiga kecamatan, yakni Sambu, Simo, dan Nogosari. Kepala Badan Penanggulangan Bencana Daerah (BPBD) Boyolali, Suratno, mengungkapkan bahwa beberapa desa terdampak banjir dengan ketinggian air yang bervariasi. Di Kecamatan Sambu, banjir merendam Desa Senting

Analisis

Hasil Sentimen: Netral

Gambar 4. 11 Tampilan Hasil Sentimen Kalimat non Shin Tae-yong

Pada Gambar 4.11 menunjukkan teks yang dimasukkan adalah sebuah cuplikan berita tentang banjir yang melanda Boyolali, Jawa Tengah, pada 8 Maret 2025. Setelah dianalisis, sistem memberikan hasil sentimen netral, yang sesuai karena teks tersebut bersifat informatif tanpa adanya opini yang bernada positif atau negatif. Hal ini menunjukkan bahwa model analisis sentimen tidak hanya terbatas pada teks tentang Shin Tae-yong, tetapi juga dapat menganalisis teks berita umum.

Masukkan teks:

semoga diberikan ketabahan dan kekuatan,, dan untuk segala kerugian Allah Ganti dengan yang lebih baik,

Analisis

Hasil Sentimen: Positif

Gambar 4. 12 Tampilan Hasil Sentimen Kalimat non Shin Tae-yong

Pada Gambar 4.12 menunjukkan teks merupakan komentar *youtube* tentang banjir, yang berisi doa dan harapan agar para korban diberi ketabahan serta mendapat pengganti yang lebih baik. Hasil analisis sentimen menunjukkan positif, yang sesuai karena teks ini mengandung empati dan doa baik. Ini menunjukkan bahwa sistem mampu menganalisis sentimen dari berbagai konteks, tidak hanya yang berkaitan dengan Shin Tae-yong.

4.6 Pembahasan

Model yang dikembangkan untuk analisis sentimen terhadap kinerja Shin Tae-yong menunjukkan performa yang baik dengan akurasi sebesar 93% berdasarkan *confusion matrix* dan metrik evaluasi. Hasil pengujian awal dengan random sampling pada 10 data acak juga menandakan bahwa model dapat membuat prediksi yang tepat pada 9 dari 10 data, menandakan kinerja awal yang cukup baik. Dari grafik *Train vs Validation Loss*, terlihat bahwa model mengalami penurunan *loss* yang signifikan pada awal pelatihan, tetapi *validation loss* mulai stabil setelah beberapa *epoch*, mengindikasikan bahwa *overfitting* pada model tidak terlalu besar. Evaluasi lebih lanjut dengan *confusion matrix* mengindikasikan bahwa model memiliki performa yang cukup baik dalam mengklasifikasikan sentimen negatif, netral, dan positif, meskipun terdapat beberapa kesalahan prediksi. Metrik *precision*, *recall*, dan *f1-score* yang tinggi pada ketiga kelas menandakan bahwa model mampu memberikan prediksi yang konsisten dan seimbang.

Penggunaan metode IndoBERT dalam penelitian ini terbukti berhasil dalam melakukan analisis sentimen terhadap kinerja Shin Tae-yong. Ini terbukti dari tingginya nilai akurasi serta distribusi klasifikasi yang cukup baik pada setiap kategori sentimen. Dibandingkan dengan metode klasik, IndoBERT menunjukkan keunggulan dalam memahami konteks bahasa Indonesia dengan lebih baik, sehingga mampu menghasilkan prediksi yang lebih tepat.

Selain evaluasi model, implementasi berbasis *Streamlit* dibuat untuk memudahkan pengguna dalam melakukan analisis sentimen. Sistem ini memiliki dua fitur utama, yaitu “Analisis Dataset” dan “Analisis Teks”. Fitur “Analisis Dataset” memungkinkan pengguna mengunggah dataset dalam format CSV untuk diproses, dengan hasil yang ditampilkan dalam bentuk tabel dan visualisasi *pie chart* untuk melihat distribusi sentimen. Sementara itu, fitur “Analisis Teks” memungkinkan pengguna memasukkan teks secara langsung untuk dianalisis sentimennya. Implementasi ini memberikan kemudahan dalam memahami dan mengakses hasil analisis.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis sentimen terhadap kinerja Shin Tae Yong yang telah dilakukan dalam penelitian ini berdasarkan dari persepsi masyarakat melalui sosial media dari *x*, *youtube* dan *instagram* dapat disimpulkan bahwa:

1. Hasil prediksi sentimen adalah 583 sentimen positif, 566 negatif, dan 351 netral dari total 1.500 data, dari data awal terdiri dari 650 positif, 602 negatif, dan 248 netral.
2. Opini Publik banyak yang memberikan apresiasi positif terhadap kinerja Shin Tae-yong. Namun, kritik terhadap kinerjanya juga cukup banyak.
3. Model BERT yang digunakan memiliki akurasi 93% yang berarti model mampu mengklasifikasikan sentimen dengan baik.
4. Publik secara umum cenderung puas dengan kinerja Shin Tae-yong sehingga hasil analisis ini memberikan gambaran yang cukup akurat mengenai persepsi masyarakat.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, berikut beberapa rekomendasi untuk penelitian selanjutnya:

1. Meningkatkan variasi dan ukuran dataset yang digunakan untuk meningkatkan generalisasi model, serta berasal dari berbagai platform media sosial.
2. Membandingkan performa BERT dengan model lain.
3. Melibatkan lebih banyak annotator untuk mengurangi bias subjektif dalam penentuan sentimen.

DAFTAR PUSTAKA

- Amien, M. (2023) “Sejarah dan Perkembangan Teknik Natural Language Processing (NLP) Bahasa Indonesia: Tinjauan tentang sejarah, perkembangan teknologi, dan aplikasi NLP dalam bahasa Indonesia,” (2007), hal. 1–7. Tersedia pada: <http://arxiv.org/abs/2304.02746>.
- Bagja, M. H. M. (2024) *Resmi Perpanjangan Kontrak, Ini 9 Prestasi Shin Tae-yong Selama Latih Timnas Indonesia*. Tersedia pada: <https://www.tempo.co/olahraga/resmi-perpanjangan-kontrak-ini-9-prestasi-shin-tae-yong-selama-latih-timnas-indonesia-44512>.
- Chandradev, V., Suarjaya, I. M. A. D. dan Bayupati, I. P. A. (2023) “Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT,” *Jurnal Buana Informatika*, 14(02), hal. 107–116. doi: 10.24002/jbi.v14i02.7244.
- Cindo, M. dan Rini, D. P. (2019) “Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) Literatur Review: Metode Klasifikasi Pada Sentimen Analisis,” *Januari*, hal. 66–70. Tersedia pada: <https://seminar-id.com/semnas-sainteks2019.html>.
- Devlin, J. dkk. (2019) “BERT: Pre-training of deep bidirectional transformers for language understanding,” *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1(Mlm), hal. 4171–4186.
- Helmiyah, S. dan Verdian, A. (2024) “Analisis Sentimen Terhadap Minat Belajar pada Tayangan Acara CoC by Ruangguru Berdasarkan Tweets Menggunakan Metode NLP dan Model BERT: Analisis Sentimen Terhadap Minat Belajar pada Tayangan Acara CoC by Ruangguru Berdasarkan Tweets Menggunakan Metode NLP,” *Jurnal Pendidikan Rosalia*, 7(2), hal. 138–149.
- Hidayat, M. N. dan Pramudita, R. (2024) “Analisis Sentimen Terhadap Pembelajaran Secara Daring Pasca Pandemi Covid-19 Menggunakan

- Metode IndoBERT,” *INFORMATION MANAGEMENT FOR EDUCATORS AND PROFESSIONALS: Journal of Information Management*, 8(2), hal. 161. doi: 10.51211/imbi.v8i2.2719.
- Hidayat, W. A. dan Nastiti, V. R. S. (2024) “PERBANDINGAN KINERJA PRE-TRAINED TOKOPEDIA SELLER CENTER,” 11(2), hal. 13–20. doi: 10.30656/jsii.v11i2.9168.
- Kemp, S. (2024) *Digital 2024: Indonesia*. Tersedia pada: <https://datareportal.com/reports/digital-2024-indonesia>.
- Kusnadi, R. *dkk.* (2021) “Analisis Sentimen Terhadap Game Genshin Impact Menggunakan Bert,” *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, 6(2), hal. 122–129. doi: 10.36341/rabit.v6i2.1765.
- Lazuardi, J. U. S. dan Juarna, A. (2023) “Analisis Sentimen Ulasan Pengguna Aplikasi Joox Pada Android Menggunakan Metode Bidirectional Encoder Representation From Transformer (Bert),” *Jurnal Ilmiah Informatika Komputer*, 28(3), hal. 251–260. doi: 10.35760/ik.2023.v28i3.10090.
- Mas, R. *dkk.* (2021) “Analisis Sentimen Customer Review Aplikasi Ruang Guru dengan Metode BERT (Bidirectional Encoder Representations from Transformers),” *Jeisbi*, 02(3), hal. 55–62.
- Merdiansah, R. dan Ali Ridha, A. (2024) “Sentiment Analysis of Indonesian X Users Regarding Electric Vehicles Using IndoBERT,” *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, 7(1), hal. 221–228.
- Nielsen (2022) “The 2022 world football report.” Tersedia pada: <https://www.nielsen.com/wp-content/uploads/sites/2/2022/07/Nielsen-World-Football-Report-2022.pdf>.
- Nugraha, P. *dkk.* (2024) “Analisis Sentimen Kepuasan Publik Terhadap Masa Kepemimpinan Shin Tae Yong Menggunakan Algoritma Naïve Bayes,” *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, 9(Oktober), hal. 149–158.
- Pasaribu, D. J. M., Kusriani, K. dan Sudarmawan, S. (2020) “Peningkatan Akurasi Klasifikasi Sentimen Ulasan Makanan Amazon dengan Bidirectional

- LSTM dan Bert Embedding,” *Inspiration: Jurnal Teknologi Informasi dan Komunikasi*, 10(1), hal. 9–20. doi: 10.35585/inspir.v10i1.2568.
- Prabowo, A. dan Indra Sanjaya, F. (2024) “Penerapan Metode Transfer Learning Pada Indobert Untuk Analisis Sentimen Teks Bahasa Jawa Ngoko Lugu,” *Simkom*, 9(2), hal. 205–217. doi: 10.51717/simkom.v9i2.478.
- Putri, C. A., Adiwijaya dan Faraby, S. Al (2020) “Analisis Sentimen Review Film Berbahasa Inggris Dengan Pendekatan Bidirectional Encoder Representations from Transformers,” *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 6(2), hal. 181–193. doi: 10.35957/jatisi.v6i2.206.
- Rahman, R. A., Pranatawijaya, V. H. dan Sari, N. N. K. (2024) “Analisis Sentimen Berbasis Aspek pada Ulasan Aplikasi Gojek,” *Jurnal Konstelasi: Konvergensi Teknologi dan Sistem Informasi*, 13(1), hal. 659. doi: 10.35889/jutisi.v13i1.1826.
- Septian, F. I., Ivana Lucia Kharisma, H. dan Kamdan (2023) “Implementasi Metode Bidirectional Encoder Representations from Transformers (BERT) untuk Analisis Sentimen Komentar Pengguna,” 3(1).
- Tanjung, T. *dkk.* (2023) “Transformers in Machine Learning: Literature Review,” *Jurnal Penelitian Pendidikan IPA*, 9(9), hal. 604–610. doi: 10.29303/jppipa.v9i9.5040.
- Thomas, S., Yuliana dan Noviyanti, P (2021) “Study Analisis Metode Analisis Sentimen pada YouTube,” *Journal of Information Technology*, 1(1), hal. 1–7. doi: 10.46229/jifotech.v1i1.201.
- Vaswani, A. *dkk.* (2017) “Attention is all you need,” *Advances in Neural Information Processing Systems*, 2017-Decem(Nips), hal. 5999–6009.
- Wijaya, K. A., Romadhony, A. dan Richasdy, D. (2023) “Implementasi Model IndoBERT pada Dashboard Sentimen Media Sosial (Studi Kasus Universitas XYZ),” *eProceedings of Engineering*, 10(4), hal. 3910–3926. Tersedia pada: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/20853%0Ahttps://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/download/20853/20321>.

Zain, M. mahrus *dkk.* (2021) “Analisis Sentimen Pendapat Masyarakat Mengenai Vaksin Covid-19 Pada Media Sosial Twitter dengan Robustly Optimized BERT Pretraining Approach,” *Jurnal Komputer Terapan*, 7(2), hal. 280–289. doi: 10.35143/jkt.v7i2.4782.

