

# SKRIPSI

## ANALISIS SENTIMEN BERBASIS PENGENALAN UCAPAN PADA VIDEO REVIEW PRODUK MENGGUNAKAN METODE BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS (BERT)

Disusun dan diajukan oleh:

MUH. IKBAL  
D121 17 1317



PROGRAM STUDI SARJANA  
TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS HASANUDDIN  
GOWA  
2024



## LEMBAR PENGESAHAN SKRIPSI

### ANALISIS SENTIMEN BERBASIS PENGENALAN UCAPAN PADA VIDEO REVIEW PRODUK MENGGUNAKAN METODE BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS (BERT)

Disusun dan diajukan oleh

**Muh. Ikbal**  
**D121 17 1323**

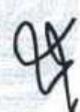
Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka  
Penyelesaian Studi Program Sarjana Program Studi Teknik Informatika  
Fakultas Teknik Universitas Hasanuddin  
Pada tanggal 26 Januari 2024  
dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,

Pembimbing Pendamping,

  
Prof. Dr. Ir. Indrabayu, S.T., M.T.,  
M.Bus.Sys., IPM, ASEAN. Eng.  
NIP 19750716 2002121 004

  
Anugrayani Bustamin, S.T., M.T.  
NIP 19901201 2018074 001

Ketua Program Studi,

  
Prof. Dr. Ir. Indrabayu, S.T., M.T.,  
M.Bus.Sys., IPM, ASEAN. Eng.  
NIP 19750716 2002121 004



## PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini ;

Nama : Muh. Ikbal  
NIM : D121 17 1317  
Program Studi : Teknik Informatika  
Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

Analisis Sentimen Berbasis Pengenalan Ucapan Pada Video Review Produk  
Menggunakan Metode Bidirectional Encoder Representations from Transformers  
(BERT)

Adalah karya tulisan saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain dan bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Semua informasi yang ditulis dalam skripsi yang berasal dari penulis lain telah diberi penghargaan, yakni dengan mengutip sumber dan tahun penerbitannya. Oleh karena itu semua tulisan dalam skripsi ini sepenuhnya menjadi tanggung jawab penulis. Apabila ada pihak manapun yang merasa ada kesamaan judul dan atau hasil temuan dalam skripsi ini, maka penulis siap untuk diklarifikasi dan mempertanggungjawabkan segala resiko.

Segala data dan informasi yang diperoleh selama proses pembuatan skripsi, yang akan dipublikasi oleh Penulis di masa depan harus mendapat persetujuan dari Dosen Pembimbing.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan isi skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Gowa, 03 MARET 2024

Yang Menyatakan

  
Muh. Ikbal



## ABSTRAK

**MUH. IKBAL.** *Analisis Sentimen Berbasis Pengenalan Ucapan Pada Video Review Produk Menggunakan Metode Bidirectional Encoder Representations from Transformers* (dibimbing oleh Anugrayani Bustamin dan Indrabayu)

Dalam era informasi saat ini, YouTube berperan sebagai *platform* media sosial utama dengan pengguna aktif harian yang sangat banyak. Popularitas konten review produk di Youtube mempermudah pengguna untuk mencari review produk yang diinginkan namun seringkali video tersebut memiliki durasi panjang untuk kebutuhan pemasangan iklan, sehingga dibutuhkan pendekatan analisis sentimen untuk dapat mengetahui kesimpulan dari review produk secara efisien.

Penelitian ini bertujuan untuk mengembangkan model yang mampu menganalisis sentimen pada video review produk. Penelitian ini memanfaatkan teknologi pengenalan ucapan dan model *pre-trained* BERT agar dapat menganalisis sentimen berdasarkan ucapan pada video review produk.

Metodologi yang digunakan pada penelitian ini meliputi pengumpulan dataset video review produk, kemudian mengekstrak audio lalu mengkonversi data audio menjadi data teks yang di augmentasi menggunakan teknik *random insertion* dan *random replacement* dan menerapkan *fine tuning* model IndoBERT untuk tugas analisis sentimen. Pengujian dan evaluasi model dilakukan dalam 4 skenario berbeda dengan ukuran *batch* 4, 8, 16, dan 32, kemudian mengukur performa berdasarkan *precision*, *recall*, dan *f1-score*.

Hasil penelitian menunjukkan bahwa model yang dikembangkan mampu melakukan sentimen analisis dengan performa yang baik dengan nilai *f1-score* yang tinggi. Skor tersebut mencapai 97,04% untuk ukuran *batch* 4 dan pada ukuran *batch* 8 mencapai 97,14%. Sedangkan pada ukuran *batch* 16 dan 32 mencapai 97,23% dan 97,41%, menunjukkan konsistensi performa yang baik pada berbagai ukuran *batch*.

Kata Kunci: Analisis Sentimen, Pengenalan Ucapan, BERT, Video Review Produk, NLP



## ABSTRACT

**MUH. IKBAL.** *Sentiment Analysis Based on Speech Recognition in Product Review Videos Using Bidirectional Encoder Representations from Transformers Method* (supervised by Anugrayani Bustamin and Indrabayu)

In the current era of information, YouTube serves as a primary social media platform with a vast daily active user base. The popularity of product review content on YouTube facilitates users in searching for desired product reviews. However, these videos often have a lengthy duration due to advertising needs, creating a demand for sentiment analysis approaches to efficiently determine the conclusions of product reviews.

This research aims to develop a model capable of accurately analyzing sentiments in product review videos. It leverages speech recognition technology and the pre-trained BERT model to analyze sentiments based on speech within product review videos.

The methodology employed in this study involves collecting a dataset of product review videos, extracting audio, converting the audio data into text which augmented using random insertion and random replacement techniques, and applying fine-tuning of the IndoBERT model for the sentiment analysis task. The model's testing and evaluation were conducted in four different scenarios with batch sizes of 4, 8, 16, and 32, measuring performance based on precision, recall, and F1-score.

The results indicate that the developed model performs well in sentiment analysis, achieving high F1-scores. The scores reached 97.04% for a batch size of 4, and 97.14% for a batch size of 8. For batch sizes of 16 and 32, the scores were 97.23% and 97.41%, respectively, demonstrating good performance consistency across various batch sizes.

**Keywords:** Sentiment Analysis, Speech Recognition, BERT, Product Review Videos, NLP



## DAFTAR ISI

PERNYATAAN KEASLIAN .....	i
ABSTRAK .....	ii
ABSTRACT .....	iii
DAFTAR ISI .....	iv
DAFTAR GAMBAR .....	vi
DAFTAR TABEL .....	vii
DAFTAR SINGKATAN DAN ARTI SIMBOL.....	viii
DAFTAR LAMPIRAN.....	ix
KATA PENGANTAR .....	x
BAB 1 PENDAHULUAN .....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian .....	3
1.4 Manfaat Penelitian .....	3
1.5 Ruang Lingkup .....	3
BAB 2 TINJAUAN PUSTAKA.....	5
2.1 Analisis Sentimen .....	5
2.2 <i>Speech Recognition</i> .....	6
2.2.1 <i>Google Automatic Speech Recognition (ASR) API</i> .....	7
2.3 <i>Word Error Rate (WER)</i> .....	8
2.4 Data Augmentation .....	8
2.4.1 Random Replacement.....	9
2.4.1 Random Insertion .....	9
2.4 <i>Transformer</i> .....	9
2.5 Bidirectional Encoder Representations From Transformers (BERT) .....	10
2.5.1 IndoBERT.....	15
2.6 Batch Size.....	15
2.7 Training & Validation Loss.....	15
2.8 Multiclass Confusion Matrix .....	17
BAB 3 METODE PENELITIAN .....	19
3.1 Tahapan Penelitian .....	19
3.2 Waktu dan Lokasi Penelitian .....	20
3.3 Instrumen Penelitian .....	20
3.3.1 Software.....	20
3.3.2 Hardware.....	20
3.3.3 Bahasa Pemrograman .....	21
3.3.4 <i>Library</i> .....	21
3.4 Teknik Pengambilan Data .....	22
3.4.1 Pengambilan Data YouTube .....	22
3.4.2 Pengambilan Data Lazada .....	22
3.5 Perancangan Sistem .....	22
Data Preparation.....	23
ASIL DAN PEMBAHASAN .....	32
.....	32
Hasil Skenario Pertama .....	32
Hasil Skenario Kedua .....	35



4.1.3 Hasil Skenario Ketiga.....	38
4.1.4 Hasil Skenario Keempat .....	41
4.2 Pembahasan .....	44
BAB 5 KESIMPULAN DAN SARAN .....	49
5.1 Kesimpulan.....	49
5.2 Saran.....	49
DAFTAR PUSTAKA .....	50



## DAFTAR GAMBAR

Gambar 1 Analisis Sentimen menggunakan <i>Machine Learning</i> .....	6
Gambar 2 Analisis Sentimen menggunakan <i>Deep Learning</i> .....	6
Gambar 3. Proses Umum Sistem <i>Speech Recognition</i> .....	7
Gambar 4. Arsitektur <i>Transformer</i> (Vaswani, 2017) .....	10
Gambar 5. Perbandingan BERT-Base dan BERT-Large .....	11
Gambar 6. Arsitektur <i>Encoder</i> pada <i>Transformer</i> (Vaswani, 2017) .....	11
Gambar 7. <i>Input Embedding Layer</i> (Kenton et al., 2018).....	12
Gambar 8. Arsitektur <i>Pre-trained</i> BERT (Kenton et al., 2018).....	13
Gambar 9. Arsitektur <i>fine tune</i> BERT (Kenton et al., 2018) .....	14
Gambar 10 <i>Multiclass Confusion Matrix</i> (Marjolein Fokkema, 2022) .....	17
Gambar 11. Diagram Tahapan Penelitian .....	19
Gambar 12. Block Diagram Sistem .....	23
Gambar 13. Sinyal Audio Hasil Segmentasi (1) .....	25
Gambar 14. Sinyal Audio Hasil Segmentasi (2) .....	26
Gambar 15. Sinyal Audio Hasil Segmentasi (3) .....	26
Gambar 16. Grafik <i>Training Loss &amp; Validation Loss</i> Skenario Pertama .....	33
Gambar 17. Perbandingan <i>F1-Score</i> Skenario Pertama.....	34
Gambar 18 <i>Confusion Matrix</i> Model Pelatihan Skenario Pertama pada <i>epoch</i> ke-10 .....	34
Gambar 19. Grafik <i>Training Loss &amp; Validation Loss</i> Skenario Kedua .....	36
Gambar 20. Perbandingan <i>F1-Score</i> Skenario Kedua .....	37
Gambar 21 <i>Confusion Matrix</i> Model Pelatihan Skenario Kedua pada <i>epoch</i> ke- 9 .....	37
Gambar 22. Grafik <i>Training Loss &amp; Validation Loss</i> Skenario Ketiga.....	39
Gambar 23. Perbandingan <i>F1-Score</i> Skenario Ketiga.....	40
Gambar 24 <i>Confusion Matrix</i> Model Pelatihan Skenario Ketiga pada <i>epoch</i> ke- 9 .....	40
Gambar 25. Grafik <i>Training Loss &amp; Validation Loss</i> Skenario Keempat .....	42
Gambar 26. Perbandingan <i>F1-Score</i> Skenario Keempat .....	43
Gambar 27 <i>Confusion Matrix</i> Model Pelatihan Skenario Ketiga pada <i>epoch</i> ke- 10 .....	43
Gambar 28. Grafik Perbandingan <i>Training Loss</i> .....	45
Gambar 29. Grafik Perbandingan <i>Validation Loss</i> .....	45
Gambar 30. Perbandingan Rata-rata <i>Validation Loss</i> .....	46
Gambar 31. Perbandingan <i>F1-Score</i> tiap skenario .....	47



## DAFTAR TABEL

Tabel 1 Percobaan parameter <i>silence threshold</i> dan <i>length</i> .....	25
Tabel 2 Nilai WER <i>Transcript</i> .....	27
Tabel 3 Contoh Data yang Telah Diberi Label .....	27
Tabel 4 Contoh Hasil Augmentasi <i>Random Insertion</i> .....	28
Tabel 5 Contoh Hasil Augmentasi <i>Random Replacement</i> .....	29
Tabel 6 Skenario Pengujian.....	30
Tabel 7 Iterasi Proses Pelatihan Skenario Pertama.....	32
Tabel 8 Metrik Evaluasi Model Pelatihan Skenario Pertama pada <i>epoch</i> ke-10 ..	35
Tabel 9 Iterasi Proses Pelatihan Skenario Kedua .....	35
Tabel 10 Metrik Evaluasi Model Pelatihan Skenario Kedua pada <i>epoch</i> ke-9 .....	38
Tabel 11 Iterasi Proses Pelatihan Skenario Kedua .....	38
Tabel 12 Metrik Evaluasi Model Pelatihan Skenario Ketiga pada <i>epoch</i> ke-9.....	41
Tabel 13 Iterasi Proses Pelatihan Skenario Keempat .....	41
Tabel 14 Metrik Evaluasi Model Pelatihan Skenario Ketiga pada <i>epoch</i> ke-10...	44
Tabel 15. Contoh hasil sentiment analisis .....	47



## DAFTAR SINGKATAN DAN ARTI SIMBOL

Lambang/Singkatan	Arti dan Keterangan
BERT	<i>Bidirectional Encoder Representative from Transformers</i>
ASR	<i>Automatic Speech Recognition</i>
TP	<i>True Positive</i>
FP	<i>False Negative</i>
TN	<i>True Negative</i>
FN	<i>False Negative</i>
MLM	<i>Masked Language Model</i>
NSP	<i>Next Sentence Prediction</i>
WER	<i>Word Error Rate</i>



## DAFTAR LAMPIRAN

Lampiran 1 Perbandingan <i>transcript</i> antara <i>SpeechRecognition</i> and <i>Youtube</i> .....	52
Lampiran 2. Lembar Perbaikan Skripsi .....	58



## KATA PENGANTAR

Assalamualaikum Warahmatullahi Wabarakatuh

Segala puji dan syukur penulis penjatkan kehadirat Allah SWT yang telah melimpahkan karunia, hidayah, kekuatan dan pencerahan kepada penulis sehingga penulis dapat menyelesaikan tugas akhir dengan judul “Analisis Sentimen Berbasis Pengenalan Ucapan Pada Video Review Produk Menggunakan Metode Bidirectional Encoder Representations from Transformers (BERT)”. Sebagai syarat dalam menyelesaikan studi jenjang Strata-1 di Departemen Teknik Informatika Fakultas Teknik, Universitas Hasanuddin.

Penulis sadar benar bahwa dalam penyusunan dan penulisan laporan penelitian ini masih banyak kekurangan dan jauh dari kesempurnaan. Namun berkat dukungan, bantuan, bimbingan, kerja sama dari berbagai pihak dan Allah SWT sehingga kendala-kendala saat proses pengerjaan skripsi ini dapat diatasi dengan usaha dan kemampuan yang penulis miliki.

Selesainya skripsi ini tentunya tidak lepas dari bimbingan serta bantuan dari berbagai pihak. Penulis mengungkapkan terima kasih dan penghargaan setinggi-tingginya kepada semua pihak yang telah membantu, dan dalam kesempatan ini penulis ingin menyampaikan banyak terima kasih yang sebesar-besarnya kepada :

1. Allah S.W.T karena atas semua berkat, karunia serta pertolongan-Nya yang tiada batas, yang telah diberikan kepada Penulis disetiap langkah dalam pembuatan tugas akhir ini hingga penulisan tugas akhir ini;
2. Kedua orang tua Penulis, Bapak Lahmuddin dan Ibu Yayuk Kusmianti yang selalu memberikan dukungan, doa dan semangat;
3. Ibu Anugrayani Bustamin, ST., M.T. selaku pembimbing I dan Bapak Prof. Dr. Ir. Indrabayu, S.T., M.T., M,Bus.Sys., IPM, ASEAN. Eng., selaku pembimbing II, yang senantiasa menyediakan waktu, tenaga, pikiran dan perhatian yang luar biasa untuk mengarahkan Penulis dalam penyusunan gas akhir ini;



4. Bapak Prof. Dr. Ir. Indrabayu, S.T., M.T., M.Bus.Sys., IPM, ASEAN. Eng., selaku Ketua Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin atas bimbingannya selama masa perkuliahan Penulis;
5. Andi Muh. Rizky dan Amiruddin sebagai tim Multimodal Content yang membantu Penulis dalam memberikan dukungan dan pemikiran-pemikiran dalam penyelesaian tugas akhir;
6. Bapak Robert, Bapak Zainuddin dan Ibu Yuanita serta segenap staff Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin yang telah membantu kelancaran penyelesaian tugas akhir Penulis;
7. Keluarga besar RECOGN17ER yang menjadi rekan seperjuangan selama berada di kampus sebagai anak teknik;
8. Teman-teman pengurus Himpunan Mahasiswa Informatika Fakultas Teknik Universitas Hasanuddin Periode 2019/2020 yang telah menemani saya dalam perjalanan organisasi;
9. Seluruh pihak yang tidak sempat disebutkan satu persatu yang telah banyak meluangkan tenaga, waktu dan pikiran selama penyusunan laporan tugas akhir ini.

Akhir kata, penulis berharap semoga Allah SWT. Berkenan membalas segala kebaikan dari semua pihak yang telah banyak membantu. Semoga Tugas Akhir ini dapat memberikan manfaat bagi pengembangan ilmu, Aamiin.

*Wassalam*

Makassar, November 2023

Penulis



# BAB 1 PENDAHULUAN

## 1.1 Latar Belakang

Di era informasi seperti saat ini memungkinkan kita untuk mengakses berbagai macam informasi dengan cepat dan mudah, salah satunya dengan menggunakan sarana media sosial. Youtube telah menjadi salah satu media sosial terpopuler saat ini dengan pengguna aktif tiap harinya mencapai 122 Juta pengguna dan lebih dari 500 jam video diunggah setiap menitnya pada tahun 2021 (Anonim, 2022). Berdasarkan halaman Situs dataindonesia.id pengguna Youtube di Indonesia merupakan terbanyak ketiga di dunia mencapai 127 juta pengguna pada Januari 2022 (Mahmudan, 2022). Youtube menjadi media sosial terpopuler di Indonesia menduduki peringkat pertama dengan jumlah pengguna terbanyak pada awal 2022 mencapai 139 juta pengguna mengalahkan Facebook dengan jumlah 129,9 juta pengguna (Jemadu & Prastya, 2022).

Youtube merupakan sebuah media sosial untuk berbagi video dimana pengguna dapat membagikan video untuk dilihat oleh orang lain secara gratis. Youtube menawarkan berbagai jenis konten baik dari segi informatif maupun hiburan. Salah satu konten yang sangat populer di Youtube adalah konten review produk. Pentingnya mengecek ulasan yang diberikan oleh orang lain sebelum membeli sebuah produk menjadi salah satu faktor mengapa konten ini sangat populer. Dengan kepopuleran konten review di Youtube, penonton dapat dengan mudah mencari review produk yang diinginkan namun konten review produk biasanya berdurasi panjang dikarenakan penjelasan produk yang secara detail ataupun kebutuhan untuk pemasangan iklan. Hal ini mengakibatkan pengguna harus menonton video yang relatif panjang agar dapat mengetahui kesimpulan dari review produk. Dengan menggunakan analisis sentimen kita dapat mengetahui dengan mudah dan cepat hasil atau kesimpulan dari sebuah review. Telah banyak penelitian terkait analisis sentimen yang telah dilakukan terhadap teks salah satunya



yang dilakukan oleh Nugroho Kunchahyo Setyo dkk, dengan judul BERTing for Sentiment Analysis on Indonesian Mobile Apps Review yang menganalisis sentimen terhadap komentar review aplikasi di playstore

(Nugroho et al., 2021). Namun untuk review suatu produk, dengan hanya menganalisa teks komentar dirasa tidaklah cukup dikarenakan komentar teks cenderung sederhana dan tidak mendetail sedangkan komentar/evaluasi yang diberikan oleh seorang konten kreator lebih jelas dan detail, oleh karena itu dengan melakukan sentimen analisis terhadap kalimat yang diucapkan oleh konten kreator pada video review produk kita dapat mengetahui hasil atau kesimpulan review produk dengan jelas.

Berdasarkan uraian diatas, untuk mengetahui dengan mudah hasil atau kesimpulan dari review suatu produk akan dilakukan pendekatan menggunakan analisis sentimen untuk menganalisis kalimat yang diucapkan oleh konten kreator seperti David Brendi pada kanal YouTube GadgetIn pada video review gadget, GadgetIn merupakan salah satu Youtuber review terkenal dari Indonesia yang melakukan review produk gadget dengan jumlah *subscriber* sebanyak 9.32 Juta dan kanal YouTube GadgetIn berhasil menempati ranking ke-26 untuk niche teknologi pada 22 Juni 2022 (Wijayanti, 2022).

Pada penelitian ini, kalimat yang diucapkan pada video akan di ubah menjadi teks menggunakan Google Speech Recognition yang selanjutnya akan dilakukan analisis sentimen menggunakan IndoBERT. IndoBERT merupakan model versi Indonesia dari *Bidirectional Encoder Representations from Transformers* (BERT), BERT merupakan teknik machine learning yang berbasis transformer untuk *natural language processing* (NLP) yang dikembangkan oleh Google. BERT dirancang agar dapat melakukan *pre-train* pada teks tanpa label secara dua arah dengan menggunakan teks disekitar kata yang memungkinkan BERT untuk mengetahui konteks dari sebuah kalimat. (Kenton et al., 2018).

Oleh karena itu, penulis mengangkat penelitian dengan judul “Analisis Sentimen Berbasis Pengenalan Ucapan Pada Video Review Produk Menggunakan Metode Bidirectional Encoder Representations from Transformers (BERT)”. Penelitian ini merupakan bagian awal dari sistem Multimodal Sentiment yang tidak hanya melibatkan satu media data tapi melibatkan berbagai jenis data seperti teks,

n visual dari sebuah video.



## 1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah dari penelitian ini adalah sebagai berikut:

- a. Bagaimana mengekstrak sinyal wicara menjadi teks dan menganalisis sentimennya berdasarkan data video ?
- b. Bagaimana unjuk kerja analisis sentiment berbasis pengenalan ucapan pada video review produk menggunakan metode BERT ?

## 1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut :

- a. Dapat mengekstrak sinyal wicara menjadi teks dan menganalisis sentimennya berdasarkan data video.
- b. Mengetahui hasil unjuk kerja analisis sentiment berbasis pengenalan ucapan pada video review produk menggunakan metode BERT.

## 1.4 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut :

- a. Mempermudah dalam pengambilan kesimpulan atau hasil sentimen dalam konten review produk di *Youtube* yang dapat digunakan oleh perusahaan atau pelanggan ketika akan mengevaluasi produk.
- b. Menjadi referensi bagi peneliti mendatang apabila akan melakukan penelitian dengan tema terkait.

## 1.5 Ruang Lingkup

Ruang lingkup dalam penelitian ini adalah sebagai berikut :

- a. Kumpulan data yang bersumber dari media sosial YouTube.
- b. Sumber video yang akan dianalisis merupakan video dari channel YouTube GadgetIn.

urasi video yang akan dianalisis berupa video durasi 10-20 menit.



- d. Video yang dianalisis merupakan video review yang secara khusus membahas satu produk tertentu, tanpa melibatkan perbandingan atau ulasan terhadap produk lain
- e. .Kumpulan data sekunder diambil dari YouTube kanal GadgetIn dan data review produk dari *platform* lazada
- f. Metrik Evaluasi pengukuran performa menggunakan *Confusion Matrix* yang terdiri dari :
  - *Precision*
  - *Recall*
  - *F1-Score*



## BAB 2 TINJAUAN PUSTAKA

### 2.1 Analisis Sentimen

Analisis sentimen merupakan sebuah proses mengekstrak informasi terkait sebuah entitas dan secara otomatis mengidentifikasi subjektivitas entitas tersebut. Tujuannya untuk menentukan apakah teks yang berasal dari pengguna menyampaikan pendapat positif, negatif, atau netral (Dang et al., 2020).

Analisis sentimen dapat mengatasi masalah mengenai bisnis, program, produk, aplikasi, dan lain-lain yang dapat dikomentari publik. Keunggulan analisis sentimen ini adalah menghemat waktu dan tenaga dalam melakukan penelitian dengan jumlah data yang besar (Herlinawati et al., 2020).

Analisis sentimen merupakan salah satu contoh dari bidang *Natural Language Processing* (NLP) yang paling populer. *Natural Language Processing* (NLP) adalah bidang ilmiah yang membahas tentang bagaimana caranya agar komputer bisa bekerja dan berpikir seperti manusia. *Natural Language Processing* (NLP) merupakan bagian dari *Artificial Intelligence* atau kecerdasan buatan (Irwansyah Saputra, 2022).

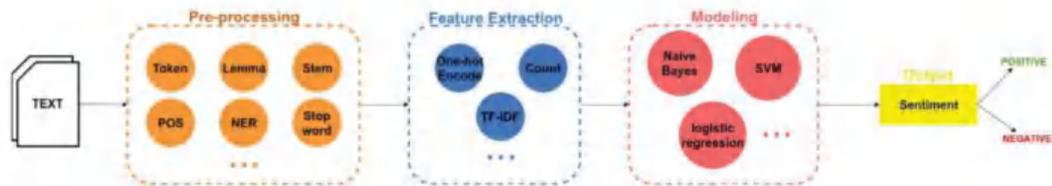
Metode *supervised machine learning* seperti *Naïve Bayes* (NB), *Support Vector Machine* (SVM), atau *Logistic Regression* (LR) telah digunakan sebagai upaya untuk menyelesaikan masalah pengkategorian sentimen teks. Namun, algoritma *machine learning* klasik ini menunjukkan kinerja yang kurang memuaskan saat berhadapan dengan *cross-lingual* atau *cross-domain* data. Berbeda dengan pendekatan berbasis *deep learning* yang menunjukkan kinerja yang lebih unggul dibandingkan dengan algoritma klasik tersebut.

Dengan menggabungkan struktur *multi-layer* ke dalam *hidden layers* pada *neural network's*, *deep learning* dapat mencapai hasil yang lebih kompleks. Metode konvensional *machine-learning* memerlukan fitur yang ditentukan dan diambil secara manual atau melalui teknik seleksi fitur. Sebaliknya, model *deep learning*

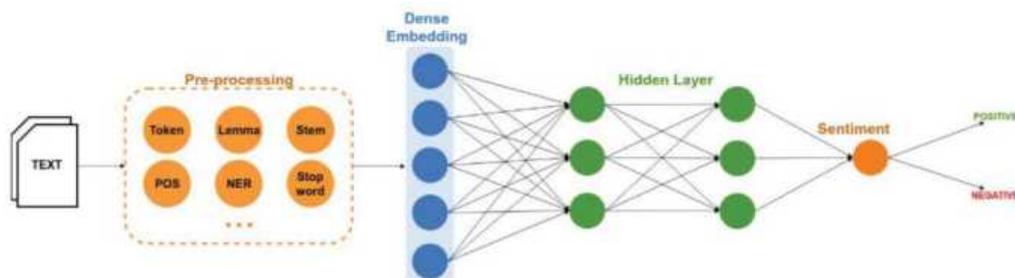


otomatis mempelajari dan mengekstrak informasi, meningkatkan akurasi dan perbandingan antara standar *machine learning* seperti SVM, *Naïve Bayes*, *decision tree* dengan *deep learning* untuk analisis sentimen dapat dilihat pada

Gambar 1 dan Gambar 2. Untuk menyelesaikan masalah yang sulit seperti di bidang gambar, pengenalan suara, dan NLP, solusi terbaik adalah menggunakan *deep learning* (Kalluri, 2023).



Gambar 1 Analisis Sentimen menggunakan *Machine Learning*



Gambar 2 Analisis Sentimen menggunakan *Deep Learning*

## 2.2 Speech Recognition

*Speech Recognition* yang juga dikenal dengan *Automatic Speech Recognition* (ASR) merupakan teknologi yang diterapkan pada perangkat lunak untuk menerima input berupa kata yang diucapkan. Teknologi ini memungkinkan suatu perangkat untuk mengenali dan memahami kata-kata yang diucapkan dengan cara digitalisasi kata dan mencocokkan sinyal digital tersebut dengan suatu pola tertentu yang tersimpan dalam suatu perangkat. Kata-kata yang diucapkan diubah menjadi sekumpulan angka yang kemudian disesuaikan dengan kode-kode tertentu untuk mengidentifikasi kata-kata tersebut (Wahyudi et al., 2015).

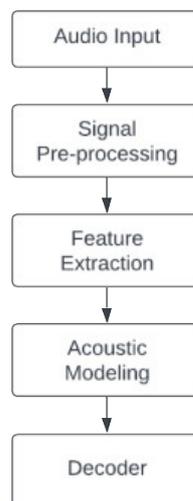
*Speech Recognition* menggabungkan ilmu komputer dan linguistik untuk mengidentifikasi kata – kata yang diucapkan dan mengubahnya menjadi teks. Hal ini memungkinkan komputer untuk memahami bahasa manusia. (Adnan et al.,



### 2.2.1 Google Automatic Speech Recognition (ASR) API

*Google Automatic Speech Recognition (ASR) API* bagian dari layanan Google Cloud Speech-to-Text, adalah alat yang memungkinkan pengembang mengonversi audio menjadi teks dengan menerapkan model jaringan saraf dalam API yang mudah digunakan. API ini mengenali lebih dari 120 bahasa dan varian, melayani basis pengguna global.

Pada penelitian ini, proses *speech-to-text* dilakukan dengan memanfaatkan Google Automatic Speech Recognition API, Dimana secara umum proses *speech recognition* beberapa tahap yang dapat dilihat seperti pada gambar 3.



Gambar 3. Proses Umum Sistem *Speech Recognition*

*Input* audio dapat berupa bahasa lisan yang ditangkap melalui mikrofon atau perangkat perekam audio lainnya, sinyal ucapan yang ditangkap oleh mikrofon, telepon, dll. Bersifat analog sehingga perlu didigitalkan, umumnya frekuensi sampling untuk sinyal ucapan adalah 8 KHz dan 20 KHz, untuk sinyal ucapan lewat telepon disarankan untuk memiliki sampling rate 8 KHz sedangkan 16 KHz umumnya digunakan untuk mikrofon normal.

Input audio umumnya melewati *pre-processing* untuk meningkatkan kualitas audio, ini biasanya melibatkan proses *noise reduction*, *filtering*, dan teknik lain untuk meningkatkan kualitas sinyal. Selanjutnya ekstraksi fitur digunakan untuk mendapatkan sekumpulan property yang stabil dan berkorelasi akustik satu sama lain yang umumnya digunakan adalah *Mel-frequency cepstral coefficients* atau representasi *spectrogram*.



*Accoustic Modeling* melibatkan pembuatan model statistic yang mewakili hubungan antara fitur yang diekstraksi dan fonem atau unit bunyi dasar sesuai dari suatu bahasa. Teknik *machine learning*, seperti *Hidden Markov Model* (HMMs) atau jaringan saraf, sering digunakan dalam langkah ini. *Language Modeling* menggabungkan pengetahuan tentang struktur dan aturan suatu bahasa. Hal ini membantu sistem membuat prediksi yang lebih akurat tentang kemungkinan urutan kata. Model *N-gram* dan model bahasa berbasis jaringan saraf biasanya digunakan untuk tujuan ini.

Tahap *Decoder* dilakukan untuk menemukan urutan kata yang paling mungkin untuk urutan observasi yang diberikan. Umumnya algoritma pemrograman dinamis digunakan untuk memecahkan masalah ini. Tujuan dari algoritma ini adalah untuk mencari jalur tunggal melalui jaringan untuk mendapatkan kecocokan terbaik untuk urutan yang diberikan, algoritma *Viterbi* banyak digunakan untuk tujuan ini (Kumar & Mittal, 2019).

### 2.3 Word Error Rate (WER)

*Word Error Rate* (WER) merupakan pendekatan standar yang digunakan untuk mengevaluasi kinerja dari sistem *Speech Recognition*. Urutan kata yang di prediksi oleh sistem ASR dibandingkan dengan transkrip referensi, dan jumlah kesalahan dihitung berdasarkan jumlah dari *substitutions* (S), *insertions* (I), dan *deletions* (D) (Ali & Renals, 2018).

Dimana *substitution* terjadi ketika ASR salah mengidentifikasi sebuah kata, *insertion* terjadi ketika ASR menambahkan kata yang tidak diucapkan, dan *deletion* terjadi ketika ASR gagal mencatat kata yang diucapkan. Jika ada N total kata dalam transkrip referensi, maka WER dihitung seperti pada formula (1):

$$WER = \frac{I + D + S}{N} \times 100 \quad (1)$$

### 2.4 Data Augmentation



*Data Augmentation* dalam konteks *Machine Learning*, dan lebih khusus lagi teks adalah strategi yang digunakan untuk meningkatkan performa dengan memperkenalkan variabilitas dalam data pelatihan. Hal ini dicapai

dengan menerapkan transformasi atau manipulasi pada data yang mempertahankan informasi inti namun menyajikannya dalam bentuk yang sedikit berbeda

Tujuannya adalah untuk membuat model yang dihasilkan lebih kuat, mengurangi risiko *overfitting*, dan meningkatkan kemampuannya untuk menggeneralisasi data baru yang belum terlihat (Shorten et al., 2021).

#### 2.4.1 Random Replacement

*Random Replacement* merupakan metode untuk membuat data baru dengan cara memilih kata secara acak pada kalimat kemudian menggantinya menggunakan sinonim dari kata tersebut (Fadilah & Priyanta, 2022).

#### 2.4.1 Random Insertion

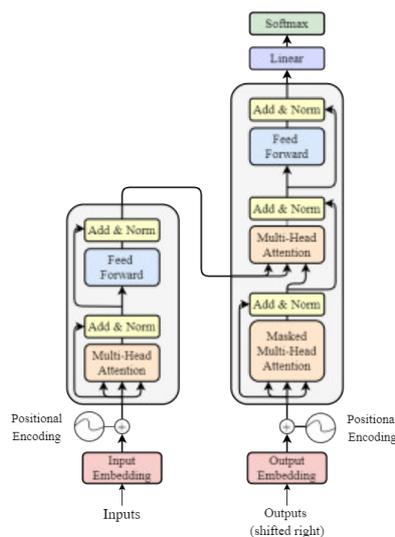
Proses *Random Insertion* hampir sama dengan *Random Replacement*, hanya saja pada proses ini kata secara acak di masukkan kedalam kalimat (Fadilah & Priyanta, 2022).

### 2.4 Transformer

*Transformer* merupakan arsitektur *neural network* yang sepenuhnya berlandaskan pada *attention mechanism*. Berbeda dengan model tradisional yang memproses data secara berurutan, Transformer dapat memahami seluruh teks sekaligus, memungkinkan untuk memahami konteks dengan lebih efisien. Hal ini membuatnya sangat efektif untuk tugas-tugas seperti penerjemahan mesin dan analisis sentimen.

Dalam penelitian ini, model sentiment analisis dibuat menggunakan model yang dibangun berdasarkan arsitektur *Transformer*. Model *Transformer* dapat dilihat pada gambar 4 terdiri dari dua bagian yaitu *encoder* yang berada disebelah kiri dan *decoder* yang berada di sebelah kanan (Vaswani, 2017).



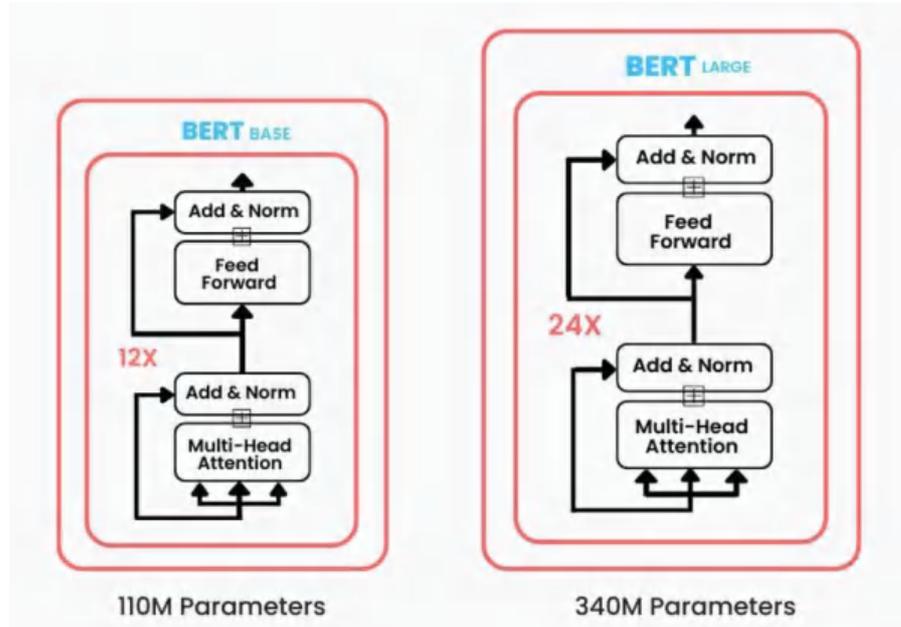


Gambar 4. Arsitektur *Transformer* (Vaswani, 2017)

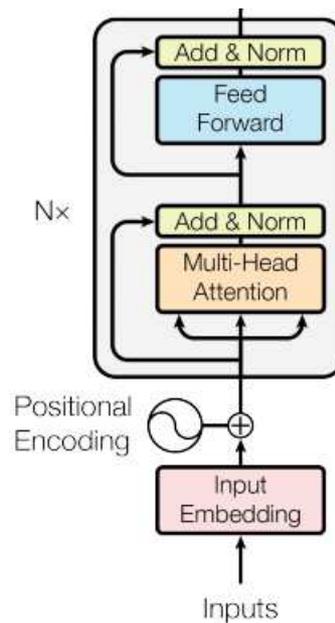
## 2.5 Bidirectional Encoder Representations From Transformers (BERT)

BERT dikembangkan pada tahun 2018 untuk membantu tugas komputer memahami bahasa alami dalam memahami konteks frasa (Kenton et al., 2019). BERT sesuai dengan namanya *Bidirectional Encoder Representations from Transformers* merupakan representasi *encoder* dari *Transformer*, tujuan dari BERT adalah untuk membuat sebuah *Language Model* oleh karena itu BERT hanya membutuhkan bagian *encoder* dari *Transformer* dan tidak memerlukan bagian *decoder*. Model BERT memiliki dua versi yaitu BERT-Base dan BERT-Large yang memiliki perbedaan pada ukuran model, dapat dilihat pada gambar 5 BERT-Base memiliki 12 *layer* atau *encoder block* yang disusun dan memiliki 110 juta *parameters* dan BERT-Large memiliki 24 *layer* atau *encoder block* yang disusun dan memiliki 340 juta *parameters*. Pada penelitian ini versi model BERT yang digunakan adalah versi BERT-Base dengan 12 *layers*.





Gambar 5. Perbandingan BERT-Base dan BERT-Large

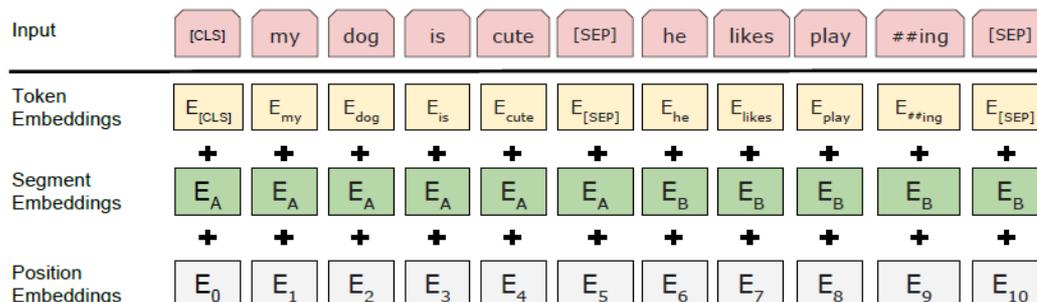


Gambar 6. Arsitektur Encoder pada Transformer (Vaswani, 2017)

Setiap *layers* atau *encoder* pada model BERT terdiri dari 2 bagian utama yaitu *multi-head attention*, dan *feed forward*. Dan sebelum data *input* masuk dan diproses di setiap *layer* data *input* akan melewati proses *input embedding* yang dapat dilihat



gambar 6 *Input embedding* pada BERT merupakan kombinasi dari tiga jenis *embedding* yang berbeda yaitu *token embeddings*, *segment embeddings*, dan *position embeddings*.



Gambar 7. *Input Embedding Layer* (Kenton et al., 2018)

Dapat dilihat pada gambar 7 Pada *token embeddings* input token diubah menjadi vektor yang menyimpan data semantik, *token embedding* pada *BERT* menggunakan *pre-trained embeddings* yaitu *WordPieces* yang memiliki 30K *vocabulary tokens*, kemudian pada *segment embedding* input token diberi kode untuk membedakan kalimat satu dengan kalimat lainnya berdasarkan token <sep> dan dapat dilihat pada gambar 10 token dengan kode A dan token dengan kode B dipisahkan oleh token <sep>, dan yang terakhir pada *positional embeddings* token diberi kode untuk mengetahui posisi token pada input data. Dan dengan menggabungkan tiga jenis *embeddings* akan menghasilkan *input embedding* yang akan masuk ke *layer* atau *encoder* pada model *BERT*.

Pada *layer encoder*, *input embedding* pertama melalui mekanisme *multi-head attention* dimana setiap *token* dibuatkan representasi *attention vectors* yang memiliki informasi relasi kontekstual setiap kata pada kata yang lain, setelah melewati *multi-head attention* dilakukan proses *Add & Norm* dimana proses ini merujuk pada dua langkah proses yang dikenal sebagai *residual connection* dan *layer normalization*, pada *residual connection* output dari *sub-layer* dalam hal ini adalah output dari *multi-head attention* ditambahkan kembali dengan *input* asli dengan tujuan untuk menghindari masalah yang disebut *vanishing gradient*, setelah penambahan tersebut normalisasi dilakukan dengan menyesuaikan skala dan translasi dari output agar memiliki *mean* 0 dan *varians* 1, hal ini untuk membantu memastikan bahwa dari setiap *sub-layer* memiliki rentang nilai yang konsisten, yang membantu model belajar dengan lebih stabil dan efektif.

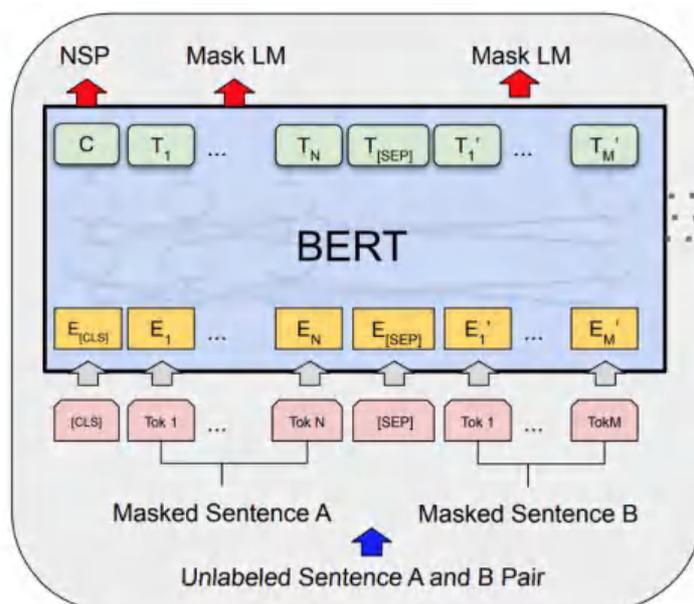


elah proses *multi-head attention* output yang berupa *attention vectors* *ub-layer feed forward* dimana pada proses ini *attention vectors* diubah bentuk yang lebih mudah untuk di proses pada *layer* selanjutnya. Pada

model BERT-Base yang memiliki 12 *layer* proses ini akan dilakukan sebanyak 12 kali sesuai dengan ukuran dari model.

Pada penelitian ini BERT digunakan untuk menyelesaikan tugas analisis sentimen, dan untuk menyelesaikan tugas tersebut terdapat dua tahap yaitu tahap pertama BERT perlu dilatih agar dapat memahami bahasa dengan baik kemudian tahap kedua dilakukan *fine tune* untuk mengajarkan BERT tugas analisis sentimen.

Pada tahap pertama yaitu tahap *pre-trained* BERT kita memerlukan agar model BERT dapat memahami bahasa dengan baik, untuk mencapai tujuan tersebut model BERT dilatih dengan menggunakan *unlabeled datasets* yang besar untuk menyelesaikan dua tugas yaitu *Masked Language Model* (MLM) dan *Next Sentence Prediction* (NSP). Dimana pada tugas MLM *input* token secara acak diberi *mask* untuk menyembunyikan nilai asli dari *token* tersebut dan kemudian model dilatih untuk memprediksi kata yang benar pada token yang diberi *mask*, tugas ini dilakukan agar model dapat memahami bahasa dalam level kata dengan lebih baik. Kemudian pada tugas kedua yaitu NSP *input embeddings* terdiri dari dua kalimat yaitu kalimat A dan kalimat B dan model akan dilatih untuk memprediksi apakah kalimat B merupakan lanjutan dari kalimat A, tugas ini dilakukan agar model dapat memahami bahasa pada level kalimat dengan lebih baik.

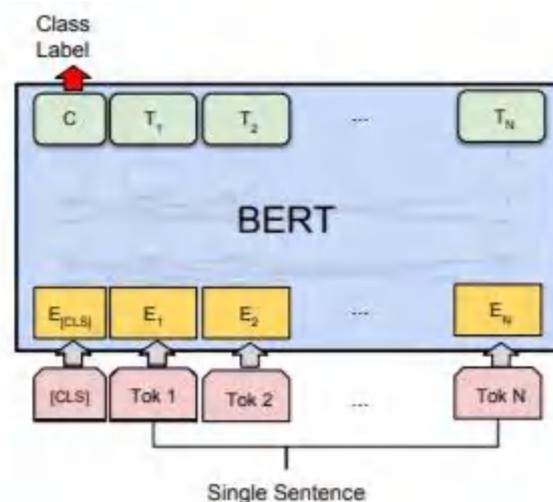


Gambar 8. Arsitektur *Pre-trained* BERT (Kenton et al., 2018)



Arsitektur *pre-trained* BERT dapat dilihat pada gambar 8 dimana *input token* terdiri dari dua kalimat yang dipisahkan oleh *token* [SEP] dan salah satu *token* diberi *mask* yang ditandai dengan nama TokM, token kemudian di *embeddings* dan diumpun ke dalam BERT *neural network* yang terdiri dari *layer encoder transformer*. Pada *output layer* untuk token yang disamarkan BERT akan memprediksi nilai asli dari token tersebut, ini adalah hasil dari tugas MLM. Dan untuk tugas NSP, output dari *token* [CLS] digunakan untuk menjadi representasi vektor yang menentukan apakah kalimat B merupakan lanjutan dari kalimat A. Dengan menjalankan dua tugas tersebut secara bersamaan BERT dapat mempelajari dan memahami bahasa dengan konteks lebih baik.

Pada tahap kedua, dilakukan *fine tune* pada *pre-trained* model dengan menambahkan *output layer* berupa *classification layer* untuk mengklasifikasi kelas sentiment dan dengan melatih model menggunakan dataset yang telah dilabel sebelumnya dengan kelas positif, negatif, dan netral.



Gambar 9. Arsitektur *fine tune* BERT (Kenton et al., 2018)

Dapat dilihat pada gambar 9 merupakan arsitektur *fine tune* BERT untuk tugas sentiment analysis, pada kasus sentiment analysis data input yang dimasukkan dalam bentuk satu kalimat dan dipecah setiap kata menjadi *token* kemudian token di *embeddings* dan diumpun ke BERT *Neural Network* dan pada *output layer* BERT merupakan representasi vektor dari konteks data yang di *input* yang akan dimasukkan ke dalam *classification layer* untuk menentukan kelas tersebut (Kenton et al., 2018).



### 2.5.1 IndoBERT

IndoBERT adalah model berbasis *transformer* bergaya BERT, tetapi dilatih murni sebagai *masked language model* (MLM) yang dilatih menggunakan *framework* Huggingface, mengikuti konfigurasi standar untuk BERT-Base yang memiliki 12 *hidden layer*, 12 *attention head* dan *feed-forward hidden layer*. (Koto & Baldwin, 2020). Model IndoBERT yang digunakan pada penelitian ini adalah model yang dikembangkan oleh Sarah Lintang pada tahun 2020 menggunakan arsitektur BERT-Base pada 16 GB data teks mentah yang memiliki lebih dari 2 milyar kata, model ini terbukti dapat mengalahkan performa model multilingual BERT pada tugas *downstream* (Sariwening, 2020).

### 2.6 Batch Size

*Batch size* merupakan *hyper-parameter* pada pelatihan model *machine learning* yang menentukan ukuran sebenarnya dari data acak yang dilatih pada setiap *steps* pelatihan (Liu et al., 2019).

*Batch size* adalah salah satu *hyper-parameter* terpenting yang harus disesuaikan, di setiap epoch, ini adalah jumlah total gambar yang digunakan dalam pelatihan jaringan. Model mungkin memerlukan waktu terlalu lama untuk mencapai konvergensi jika *hyperparameter* disetel terlalu tinggi. Namun, jika disetel terlalu rendah, model akan terpentol tanpa mencapai performa yang diinginkan (Aldin & Aldin, 2022).

### 2.7 Training & Validation Loss

*Training & Validation loss* merupakan dua metrik penting dalam *machine learning* dan *deep learning* yang digunakan untuk menilai performa suatu model. *Training loss* adalah metrik yang mengukur seberapa cocok model dengan data pelatihan. Ini menilai kesalahan model pada set pelatihan, yang merupakan bagian dari set data yang digunakan untuk melatih model. *Training loss* dihitung dengan mengambil jumlah kesalahan untuk setiap contoh dalam set pelatihan dan biasanya setelah setiap *batch*. Hal ini membantu dalam memahami seberapa baik mempelajari dan menyesuaikan dengan data pelatihan.



Sebaliknya, *validation loss* digunakan untuk menilai performa model pada set validasi, yang merupakan bagian terpisah dari kumpulan data yang disisihkan untuk memvalidasi performa model. Seperti *training loss*, ini dihitung dari jumlah kesalahan untuk setiap contoh dalam set validasi. *Validation loss* diukur setiap *epoch*, memberikan wawasan apakah model memerlukan penyetelan atau penyesuaian lebih lanjut (Baeldung, 2023).

*Training & validation loss* sering divisualisasikan bersama dalam grafik untuk mendiagnosis performa model dan mengidentifikasi aspek mana yang perlu disesuaikan. Misalnya, jika *training & validation loss* tinggi, hal ini mungkin mengindikasikan *underfitting*, yang berarti model tidak belajar secara memadai dari data pelatihan. Jika *training loss* menurun namun *validation loss* meningkat, hal ini dapat mengindikasikan *overfitting*, yaitu model mempelajari data pelatihan dengan terlalu baik namun gagal melakukan generalisasi ke data baru.

Dalam *machine learning*, tujuannya biasanya adalah untuk meminimalkan *training & validation loss*, memastikan bahwa model tidak hanya belajar dengan baik dari data pelatihan tetapi juga menggeneralisasi dengan baik data baru yang belum terlihat (Baeldung, 2023).

Pada pelatihan model BERT untuk tugas analisis sentiment *multi-class*, *Cross-Entropy Loss* biasa digunakan untuk menghitung *training & validation loss* dengan membandingkan distribusi probabilitas model yang diprediksi di seluruh kelas dengan distribusi sebenarnya (label sebenarnya). Adapun cara untuk menghitung *loss* menggunakan *Cross-Entropy Loss* dapat dilihat pada formula (2)

$$\text{Cross - Entropy} = - \sum_x p(x) \cdot \log q(x) \quad (2)$$

Formula diatas menjumlahkan produk dari distribusi probabilitas sebenarnya  $p(x)$  untuk sebuah kelas  $x$  dan logaritma probabilitas yang diprediksi  $q(x)$  untuk kelas yang sama, disemua kelas. Fungsi ini menghitung perbedaan antara probabilitas yang diprediksi dan label kelas sebenarnya. (Vlastimil Martinek, 2020).



## 2.8 Multiclass Confusion Matrix

*Confusion matrix* adalah matriks yang digunakan untuk melakukan evaluasi proses model klasifikasi berupa jumlah data uji yang benar dan salah. Dengan adanya matriks ini dapat mengetahui kualitas kinerja model klasifikasi (Normawati & Prayogi, 2021).

Matriks ini berisi data target prediksi yang dibandingkan dengan data target aktual. Data prediksi merupakan nilai yang didapatkan dari hasil pemodelan *machine learning*, sedangkan data aktual adalah nilai sebenarnya yang dimiliki. Adanya *confusion matrix* untuk mengetahui sejauh mana *machine learning* bekerja sesuai dengan yang diinginkan. *Confusion matrix* berisi berbagai performa yang dapat diukur seperti akurasi, presisi, *recall* dan F1 Score untuk mengetahui seberapa baik kinerja dari pemodelan yang telah dilakukan sebelumnya (Irwansyah Saputra, 2022).

		PREDICTED classification			
		Classes	a	b	c
ACTUAL classification	a	TN	FP	TN	TN
	b	FN	TP	FN	FN
	c	TN	FP	TN	TN
	d	TN	FP	TN	TN

Gambar 10 *Multiclass Confusion Matrix* (Marjolein Fokkema, 2022)

*Multiclass Confusion Matrix* memungkinkan matriks memvisualisasi kinerja algoritma dalam skenario yang melibatkan lebih dari dua kelas. Setiap baris dari matriks mewakili kelas data aktual, sementara setiap kolom mewakili kelas data prediksi atau sebaliknya (Haghighi et al., 2018).

*Precision* adalah hasil bagi antara jumlah elemen *True Positive* dan total nit yang diprediksi sebagai positif (jumlah kolom dari prediksi positif).  
 a. Misalnya, *True Positive* adalah elemen yang diberi label positif oleh model yang positif, sedangkan *False Positive* adalah elemen yang diberi label



positif oleh model tapi sebenarnya negatif. *Precision* memberitahu kita seberapa banyak kita dapat mempercayai model ketika memprediksi unit sebagai positif seperti yang ditunjukkan pada formula (3).

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

*Recall* adalah hasil bagi antara jumlah elemen *True Positive* dan total jumlah unit yang diklasifikasi positif (jumlah kolom dari nilai yang benar-benar positif). Lebih detailnya, *False Negative* adalah elemen yang diberi label negatif oleh model tapi nilai sebenarnya adalah positif. *Recall* mengukur akurasi prediksi model untuk kelas positif, ini mengukur kemampuan model untuk menemukan semua unit positif didalam dataset seperti pada formula (4).

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

*F1-Score* mengevaluasi performa model klasifikasi berdasarkan matriks konfusi. Formula *F1-Score* dapat diinterpretasikan sebagai rata-rata antara *Precision* dan *Recall* yang ditunjukkan oleh formula (5), dimana nilai *F1-Score* mencapai puncaknya pada 1 dan mencapai terendahnya pada 0. Kontribusi relative dari presisi dan *recall* adalah setara dalam *F1-Score*, dan rata-rata harmonik digunakan untuk mencari keseimbangan terbaik antara dua ukuran (Grandini et al., 2020).

$$F1-Score = 2 \times \left( \frac{Precision \times Recall}{Precision + Recall} \right) \quad (5)$$

