



Klasifikasi Emosi Data *Text* Bahasa Indonesia Menggunakan Algoritma BERT, RoBERTa, dan DistilBERT

Faishal Basbeth
20917015

*Tesis diajukan sebagai syarat untuk meraih gelar Magister Komputer
Konsentrasi Data Sains
Program Studi Informatika Program Magister
Fakultas Teknologi Industri
Universitas Islam Indonesia
2024*

Lembar Pengesahan Pembimbing

Klasifikasi Emosi Data *Text* Bahasa Indonesia Menggunakan Algoritma BERT, RoBERTa, dan DistilBERT

Faishal Basbeth

20917015



Yogyakarta, Juli 2024



الجامعة الإسلامية
الاندونيسية

Pembimbing



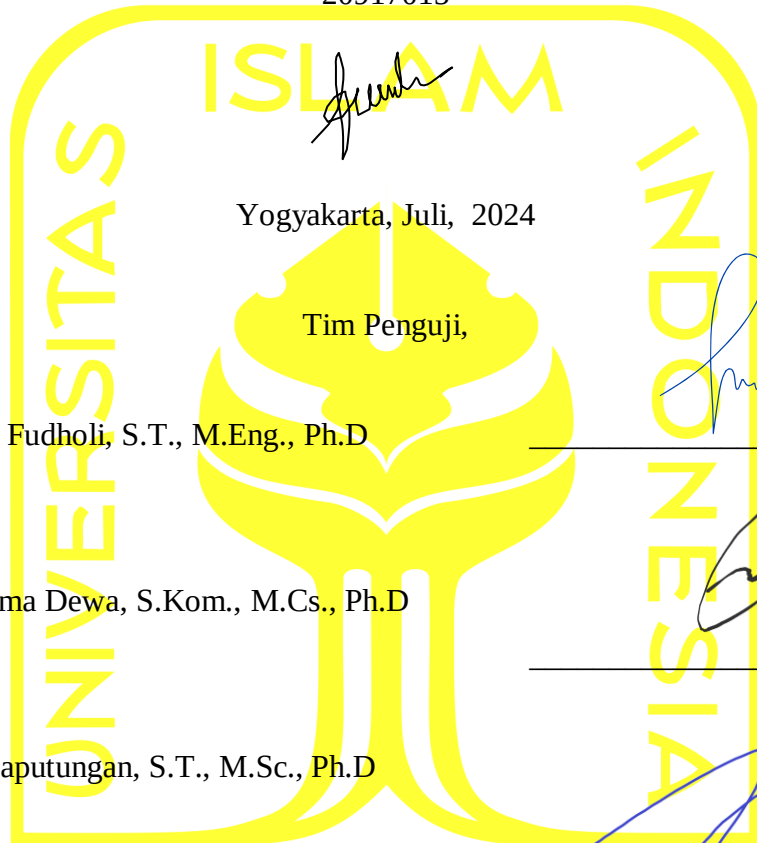
DThomas Hatta Fudholi, S.T., M.Eng., Ph.D

Lembar Pengesahan Penguji

Klasifikasi Emosi Data *Text* Bahasa Indonesia Menggunakan Algoritma BERT, RoBERTa, dan DistilBERT

Faishal Basbeth

20917015



Yogyakarta, Juli, 2024

Tim Penguji,

Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D

Ketua Penguji

Chandra Kusuma Dewa, S.Kom., M.Cs., Ph.D

Penguji I

Irving Vitra Paputungan, S.T., M.Sc., Ph.D

Penguji II

22/07/2024

Mengetahui,

Ketua Program Studi Informatika Program Magister

Universitas Islam Indonesia



Irving Vitra Paputungan, S.T., M.Sc., Ph.D

Abstrak

Klasifikasi Emosi Data *Text* Bahasa Indonesia Menggunakan Algoritma BERT, RoBERTa, dan DistilBERT

Analisis sentimen memiliki beberapa peran penting untuk berbagai macam tujuan. 2 hasil yang didapat dalam analisis sentimen seperti positif dan negatif masih memiliki keterbatasan dalam memahami makna emosional kalimat dalam bentuk *text* digital, berbeda dengan klasifikasi emosi data *text*. Klasifikasi emosi data *text* memiliki kelebihan dalam memahami makna emosional kalimat dalam bentuk *text* digital dengan berbagai macam makna emosional. Data *text* yang digunakan dalam klasifikasi emosi data *text* kebanyakan menggunakan data *text* Bahasa Inggris dan pengembangan penelitian tentang klasifikasi emosi data *text* Bahasa Indonesia masih berjalan lambat. Berjalan lambatnya penelitian klasifikasi emosi data *text* Bahasa Indonesia mengharapakan peneliti untuk membangun klasifikasi emosi data *text* Bahasa Indonesia menggunakan 3 model yang mumpuni : BERT, RoBERTa, DistilBERT untuk mengetahui kualitas masing-masing model dalam nilai ukuran, waktu dalam melakukan *training*, akurasi, dan *f1-score* dengan menggunakan proses *fine-tuning* : penambahan *output layer* dan *loss function* dan *hyperparameter-tuning*. Data *text* yang digunakan adalah indo4B dan mengolah dengan cara *undersample*. Penelitian ini menghasilkan model yang dapat menerapkan klasifikasi emosi dengan hasil yang baik. Skenario yang telah diterapkan pada penelitian ini menghasilkan model BERT mendapatkan nilai parameter = 109486085, waktu *training* = 24 min 00 s, akurasi = 89,60, dan nilai *f1-score* = 90. Model RoBERTa mendapatkan nilai parameter = 124649477, waktu *training* = 25 min 39 s, akurasi = 90,83, dan nilai *f1-score* = 91. Model DistilBERT mendapatkan nilai parameter = 66366725, waktu *training* = 13 min 34 s, akurasi = 88,99, dan nilai *f1-score* = 89.

Kata kunci

analisis sentimen, klasifikasi emosi, fine-tuning, bahasa indonesia

Abstract

Emotion Classification of Indonesian Text Data Using the BERT, RoBERTa, and DistilBERT Algorithms

Sentiment analysis has several important roles for various purposes. The 2 results obtained in sentiment analysis, such as positive and negative, still have limitations in understanding the emotional meaning of sentences in digital text form, in contrast to the emotional classification of text data. Emotional classification of text data has the advantage of understanding the emotional meaning of sentences in the form of digital text with various kinds of emotional meaning. The text data used in emotion classification of text data mostly uses English text data and the development of research on emotion classification of Indonesian text data is still progressing slowly. The slow pace of research on emotional classification of Indonesian text data expects researchers to build an emotional classification of Indonesian text data using 3 capable models: BERT, RoBERTa, DistilBERT to determine the quality of each model in terms of measurement value, time in training, accuracy, and f1-score using a fine-tuning process: adding output layers and loss functions and hyperparameter-tuning. The text data used is Indo4B and processed using undersample. This research produces a model that can apply emotion classification with good results. The scenario that has been applied in this research results in the BERT model getting a parameter value = 109486085, training time = 24 min 00 s, accuracy = 89.60, and f1-score value = 90. The RoBERTa model gets a parameter value = 124649477, training time = 25 min 39 s, accuracy = 90.83, and f1-score = 91. The DistilBERT model gets parameter value = 66366725, training time = 13 min 34 s, accuracy = 88.99, and f1-score = 89.

Keywords

sentiment analysis, emotion classification, fine-tuning, indonesian language

Pernyataan Keaslian Tulisan

Dengan ini saya menyatakan bahwa tesis ini merupakan tulisan asli dari penulis, dan tidak berisi material yang telah diterbitkan sebelumnya atau tulisan dari penulis lain terkecuali referensi atas material tersebut telah disebutkan dalam tesis. Apabila ada kontribusi dari penulis lain dalam tesis ini, maka penulis lain tersebut secara eksplisit telah disebutkan dalam tesis ini.

Dengan ini saya juga menyatakan bahwa segala kontribusi dari pihak lain terhadap tesis ini, termasuk bantuan analisis statistik, desain survei, analisis data, prosedur teknis yang bersifat signifikan, dan segala bentuk aktivitas penelitian yang dipergunakan atau dilaporkan dalam tesis ini telah secara eksplisit disebutkan dalam tesis ini.

Segala bentuk hak cipta yang terdapat dalam material dokumen tesis ini berada dalam kepemilikan pemilik hak cipta masing-masing. Apabila dibutuhkan, penulis juga telah mendapatkan izin dari pemilik hak cipta untuk menggunakan ulang materialnya dalam tesis ini.

Yogyakarta, Juli, 2024



Faishal Basbeth, S.Kom

Daftar Publikasi

Publikasi yang menjadi bagian dari tesis

Basbeth, F., & Fudholi, D. H. (2024). Klasifikasi Emosi Pada Data *Text* Bahasa Indonesia Menggunakan Algoritma BERT, RoBERTa, dan Distil-BERT. 2 (8) : 1160-1170
<http://dx.doi.org/10.30865/mib.v8i2.7472>

Sitasi publikasi 1

Kontributor	Jenis Kontribusi
Faishal Basbeth	Mendesain eksperimen (60%) Menulis <i>paper</i> (60%)
Dhomas Hatta Fudholi	Mendesain eksperimen (40%) Menulis dan mengedit <i>paper</i> (40%)

Halaman Kontribusi

“Tidak ada kontribusi dari pihak lain”

Halaman Persembahan

Alhamdulillah segala puji bagi Allah yang telah memberikan kita semua nikmat dan dengan rahmat dan karunia dari Allah penulis bisa menyelesaikan Pendidikan program Studi Informatika Program Magister di Universitas Islam Indonesia Yogyakarta. Penulis persembahkan penelitian ini kepada :

1. Kedua orang tua, kakak, adik, dan keluarga besar yang telah memberikan dukungan kepada penulis yang tiada hentinya.
2. Seluruh dosen pembimbing, dosen penguji, dan seluruh dosen yang mengajar pada perkuliahan program Studi Informatika.
3. Seluruh sahabat dan rekan seperjuangan yang selalu bersama-sama dalam menyelesaikan studi jurusan Informatika Program Magister ini.

Kata Pengantar

Puji syukur kehadiran Allah SWT yang telah melimpahkan nikmat dan karunianya sehingga penulis dapat menyelesaikan laporan tesis dengan judul “Klasifikasi Emosi Pada Data *Text* Menggunakan Algoritma DistilBERT”. Laporan ini disusun sebagai syarat kelulusan pada Program Studi Informatika Program Magister (S2), Fakultas Teknologi Industri, Universitas Islam Indonesia.

Pada kesempatan ini penulis ingin menyampaikan rasa hormat dan terima kasih kepada :

1. Bapak Prof. Fathul Wahid, S.T., M.Sc., Ph.D., selaku Rektor Universitas Islam Indonesia.
2. Bapak Irving Vitra Paputangan, S.T., M.Sc., Ph.D., selaku Ketua Program Studi Informatika Program Magister, Fakultas Teknologi Industri, Universitas Islam Indonesia.
3. Bapak Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D., selaku dosen Pembimbing 1, Terima kasih atas segala arahan dan bimbingan ilmu pengetahuan yang diberikan sehingga penulis tesis ini dapat berjalan dengan baik.
4. Bapak/Ibu Dosen dan seluruh pegawai Program Studi Informatika Program Magister, Fakultas Teknologi Industri, Universitas Islam Indonesia.
5. Teman-teman Magister Teknik Informatika Universitas Islam Indonesia Yogyakarta.
6. Teman-teman dari mahasiswa perkuliahan lain.

Akhir kata, semoga tesis ini dapat memberikan manfaat bagi pihak-pihak terkait.

Yogyakarta, Juli 2024



(Faishal Basbeth)

Daftar Isi

Lembar Pengesahan Pembimbing.....	i
Lembar Pengesahan Penguji.....	ii
Abstrak	iii
Abstract	iv
Pernyataan Keaslian Tulisan	v
Daftar Publikasi	vi
Halaman Kontribusi	vii
Halaman Persembahan	viii
Kata Pengantar.....	ix
Daftar Isi.....	x
Daftar Tabel.....	xii
Daftar Gambar	xiii
BAB 1 Pendahuluan	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian	5
1.4 Batasan Penelitian.....	5
1.5 Sistematika Penulisan	5
BAB 2 Tinjauan Pustaka	7
2.1 Landasan Teori	7
2.1.1 Klasifikasi Emosi.....	7
2.1.2 BERT	8
2.1.3 RoBERTa	11
2.1.4 <i>Distillation</i> Pada BERT Menjadi DistilBERT	11
2.1.4.1 DistilBERT	13

2.1.5	Matriks Evaluasi	13
2.2	Kajian Pustaka	14
BAB 3 Metodologi.....		22
3.1	Langkah Penelitian	22
3.2	Uraian Metodologi.....	22
3.2.1	Pengumpulan Data.....	22
3.2.2	Preprocessing Data	24
3.2.3	Pemodelan	25
3.2.4	Matriks Evaluasi	27
BAB 4 Hasil dan Pembahasan.....		28
4.1	Hasil Pengumpulan Data.....	28
4.2	Hasil preprocessing Data.....	29
4.3	Hasil Hyperparameter-Tuning dan Fine-tuning per Model	31
4.4	Hasil Total Parameter Setiap <i>Hyperparameter-tuning</i> dan <i>Fine-tuning</i> per Model.	32
4.5	Hasil Waktu <i>Training</i> Setiap <i>Hyperparameter-tuning</i> dan <i>Fine-tuning</i> per Model.	33
4.6	Hasil Akurasi Setiap <i>Hyperparameter-tuning</i> dan <i>Fine-tuning</i> per Model.....	34
4.7	Hasil Analisa Peningkatan Performa <i>Training</i> Setiap <i>Hyperparameter-tuning</i> dan <i>Fine-tuning</i> per Model.	35
4.8	Hasil Prediksi Klasifikasi Emosi dan <i>F1-score</i> Setiap <i>Hyperparameter-tuning</i> dan <i>Fine-tuning</i> per Model.	39
4.9	Hasil Dari Semua Aspek Setiap <i>Hyperparameter-tuning</i> dan <i>Fine-tuning</i> per Model.	43
BAB 5 Kesimpulan dan Saran.....		46
5.1	Kesimpulan.....	46
5.2	Saran	46
Daftar Pustaka.....		48

Daftar Tabel

Tabel 2.1 4 Kombinasi Berbeda Dari Nilai Prediksi dan Nilai Aktual.....	14
Tabel 2.2 Penelitian Sebelumnya Klasifikasi Emosi Data <i>Text</i> Bahasa Inggris	17
Tabel 2.3 Penelitian Sebelumnya Klasifikasi Emosi Data <i>Text</i> Bahasa Indonesia	20
Tabel 3.1 Contoh Data <i>Text</i> Indo4b	24
Tabel 4.1 Contoh Dan Pembagian Data per Label	29
Tabel 4.2 Hasil <i>Preprocessing</i> Data	30
Tabel 4.3 Hasil <i>One Hot Encoding</i> Pada Label.....	31
Tabel 4.4 Arsitektur <i>Hyperparameter-Tuning</i> Pada Model Bert, Roberta, Distilbert	31
Tabel 4.5 Arsitektur <i>Fine-tuning</i> Pada Model Bert, Roberta, Distilbert.....	32
Tabel 4.6 Hasil Total Parameter Dari Arsitektur <i>Hyperparameter-tuning</i> dan <i>Fine-tuning</i> Pada Model Bert, Roberta, Distilbert.....	33
Tabel 4.7 Hasil Waktu <i>Training</i> Dari Arsitektur <i>Hyperparameter-tuning</i> dan <i>Fine-tuning</i> Pada Model Bert, Roberta, Distilbert.....	34
Tabel 4.8 Hasil Akurasi Dari Arsitektur <i>Hyperparameter-tuning</i> dan <i>Fine-tuning</i> Pada Model Bert, Roberta, Distilbert	34
Tabel 4.9 Hasil Parameter, Waktu <i>Training</i> , Akurasi, dan <i>F1-score</i> Setiap <i>Hyperparameter-</i> <i>tuning</i> dan <i>Fine-tuning</i> per Model	44
Tabel 4.10 Perbandingan Hasil <i>F1-score</i> Dengan Penelitian Sebelumnya Pada Model Bert, Roberta, Distilbert.....	45

Daftar Gambar

Gambar 2.1 Token model bert.....	8
Gambar 2.2 Struktur <i>pre-training</i> model bert.	10
Gambar 3.1 Diagram alir klasifikasi emosi data text bahasa indonesia dengan model bert, roberta, distilbert.....	22
Gambar 3.2 Urutan tahap <i>preprocessing</i> data.	25
Gambar 3.3 Gambaran klasifikasi emosi data <i>text</i> menggunakan <i>fine-tuning</i> pada model bert, roberta, distilbert.....	26
Gambar 4.1 Keseimbangan data per label.....	28
Gambar 4.2 Perubahan <i>loss</i> pada setiap <i>epoch</i> model bert	36
Gambar 4.3 Perubahan akurasi pada setiap <i>epoch</i> model bert	36
Gambar 4.4 Perubahan <i>loss</i> pada setiap <i>epoch</i> model roberta	37
Gambar 4.5 Perubahan akurasi pada setiap <i>epoch</i> model roberta	37
Gambar 4.6 Perubahan <i>loss</i> pada setiap <i>epoch</i> model distilbert.....	38
Gambar 4.7 Perubahan akurasi pada setiap <i>epoch</i> model distilbert	38
Gambar 4.8 Hasil prediksi klasifikasi pada model bert	40
Gambar 4.9 Hasil <i>f1-score</i> model bert.....	40
Gambar 4.10 Hasil prediksi klasifikasi pada model roberta	41
Gambar 4.11 Hasil <i>f1-score</i> model roberta	41
Gambar 4.12 Hasil prediksi klasifikasi pada model distilbert.....	43
Gambar 4.13 Hasil <i>f1-score</i> model distilbert	43

BAB 1

Pendahuluan

1.1 Latar Belakang

Analisis sentimen sudah sangat populer di bidang *artificial intelligence* yang membantu dalam menganalisa *text* digital dengan menggabungkan beberapa teknik dalam *artificial intelligence* seperti : *Natural Language Processing* (NLP) dan *deep learning* (Huan et al., 2022). Analisis sentimen memiliki beberapa peran penting untuk berbagai macam tujuan serta dapat diaplikasikan dalam berbagai skenario seperti : evaluasi produk dan layanan suatu perusahaan yang telah diberi penilaian oleh kustomer dalam bentuk *text* digital (Lin & Nuha, 2023) dan evaluasi program pemerintahan dari suatu negara yang telah diberi penilaian oleh rakyat dalam bentuk *text* digital. Analisis sentimen dapat memahami makna emosional kalimat dalam bentuk *text* digital yang mengandung kalimat bermakna positif dan negatif (Medhat et al., 2014). Kalimat positif dapat diberi persamaan terhadap makna emosional bahagia, cinta, dsb. Kalimat negatif dapat diberi persamaan terhadap makna emosional marah, sedih, dsb. 2 hasil yang didapat dalam analisis sentimen seperti positif dan negatif masih memiliki keterbatasan dalam memahami makna emosional kalimat dalam bentuk *text* digital (Aman & Szpakowicz, 2007), berbeda dengan klasifikasi emosi data *text*. Klasifikasi emosi data *text* memiliki tujuan yang sama seperti analisis sentimen, tetapi klasifikasi emosi data *text* memiliki kelebihan dalam memahami makna emosional kalimat dalam bentuk *text* digital dengan berbagai macam makna emosional (Acheampong et al., 2021).

Pembangunan klasifikasi emosi data *text* membutuhkan sebuah kumpulan data *text* dari berbagai sumber data dan berbagai macam model NLP dan *deep learning* pada penelitian sebelumnya. Sumber data pada data *text* berisikan data kalimat dan data label. Data kalimat berisikan tentang berbagai macam kalimat yang ada pada ungkapan manusia yang di dapat dalam bentuk data *text* dan data label memberi kesimpulan terhadap kalimat tersebut memiliki makna emosional apa yang telah diungkapkan dalam bentuk data *text*. Data *text* yang digunakan dalam klasifikasi emosi data *text* kebanyakan menggunakan data *text* Bahasa Inggris, penelitian yang dilakukan oleh (Wilie et al., 2020) dan (Yanuar & Shiramatsu, 2020) memaparkan informasi bahwa pengembangan penelitian tentang klasifikasi emosi data *text* Bahasa Indonesia masih berjalan lambat. Berjalan lambatnya

penelitian klasifikasi emosi data *text* Bahasa Indonesia mengharapkan peneliti untuk membangun penelitian tersebut.

Penelitian sebelumnya dalam kasus klasifikasi emosi data *text* Bahasa Indonesia menggunakan beberapa model yang mumpuni dalam kasus tersebut, seperti model BERT penelitian yang dilakukan oleh (Wilie et al., 2020), model RoBERTa penelitian yang dilakukan oleh (Wiciaputra et al., 2021), dan model DistilBERT penelitian yang dilakukan oleh (Nissa & Yulianti, 2023). Model BERT adalah model berbasis *deep learning* yang menggunakan *neural network* untuk melakukan sistem *pre-training* model NLP. Model BERT dapat melakukan *training* data *text* berdasarkan seluruh rangkaian kata dalam kalimat dengan melakukan *training* dari 2 arah secara langsung, arah yang dituju model BERT adalah dari urutan kata dari kiri ke kanan dan urutan kata dari kanan ke kiri, berbeda dengan model tradisional seperti model RNN yang hanya melakukan *training* data *text* berdasarkan seluruh rangkaian kata dalam kalimat dengan melakukan *training* dari 1 arah, yaitu urutan kata dari kiri ke kanan (González-Carvajal & Garrido-Merchán, 2020). Model BERT memungkinkan untuk mempelajari konteks kata berdasarkan kata – kata disekitarnya, bukan hanya kata yang mendahului atau mengikutinya (Sun et al., 2019). Model BERT melakukan *training* data *text* berdasarkan seluruh rangkaian kata dalam kalimat dengan melakukan *training* dari 2 arah secara langsung dan menggunakan model *Masked Language Model* (MLM) dan *Next Sentence Prediction* (NSP) (Devlin et al., 2018). Penelitian yang dilakukan oleh (Wilie et al., 2020) menerapkan model BERT dalam kasus klasifikasi emosi data *text* Bahasa Indonesia dengan *fine-tuning* model BERT menggunakan model MLM yang melakukan prediksi token tertutup dengan pembatasan prediksi sebanyak 20 token tertutup, model tersebut diberi nama model IndoBERT BASE, penelitian ini juga membangun data *text* yang diberi nama indo4B. Indo4B memiliki 5 label, data *training* = 3.521, *validation* = 440, dan *test* = 442. Model klasifikasi yang digunakan, menggunakan model klasifikasi *single* label. Hasil akurasi yang didapat pada model IndoBERT-lite BASE = 73,88 dan model IndoBERT BASE = 75,48.

Model RoBERTa adalah model berbasis *deep learning* yang menggunakan *neural network* untuk melakukan sistem *pre-training* model NLP, model RoBERTa melakukan beberapa perkembangan pada model BERT dengan proses *pre-training* yang dilakukan dengan data *text* yang berukuran besar = 160 GB. Proses *pre-training* dilakukan dengan memodifikasi *hyperparameter* dan *word embedding*. *Input* yang digunakan sama dengan *input* model BERT tetapi model RoBERTa tidak menggunakan *segment embedding* (*token_type_ids*) dan tidak menggunakan model NSP. Model RoBERTa juga menggunakan

MLM dengan model *masking dynamic* (Liu et al., 2019). Penelitian yang dilakukan oleh (Wiciaputra et al., 2021) melakukan klasifikasi data *text* Bahasa Indonesia dengan model XLM-RoBERTa tanpa melakukan uji coba *hyperparameter-tuning*, tetapi menggunakan 5 data *text* yang berbeda dan digabungkan, seperti *english news dataset* = 37.886 Bahasa Inggris, *mixed news dataset* = 75.772 dari 37.886 Bahasa Inggris dan 37.886 Bahasa Indonesia, *mixed news dataset* = 108.190 dari Bahasa Inggris 37.886 dan 70.304 Bahasa Indonesia, *english news dataset* = 2.000, dan *mixed news dataset* = 4.000 dari 2.000 Bahasa Inggris dan 2.000 Bahasa Indonesia. Penelitian ini menghasilkan *f1-score* terbaik pada data *mixed news dataset* = 75.772 dengan *f1-score* = 93,30.

Model DistilBERT adalah model yang menyaring beberapa perkembangan pada model BERT, model BERT disebut model guru dan model DistilBERT disebut model siswa. Model siswa menghapus sementara *token-type-embedding* dan *pooler*. Model DistilBERT memiliki kelebihan : pengurangan sebesar 40% di ukuran model, kenaikan kecepatan waktu dalam melakukan *training* model 60%, dan pencapaian kemampuan mempelajari konteks kata hingga 97% dalam kemampuan pemahaman bahasa (Sanh et al., 2020). Penelitian yang dilakukan oleh (Nissa & Yulianti, 2023) menerapkan model m-BERT-base dan distil-BERT-base dengan model multilingual untuk melakukan klasifikasi data *text* Bahasa Indonesia, penelitian ini menggunakan data *airy rooms hotel reviews* yang berisi tentang komentar pada bagian/ruangan mana yang telah diberi penilaian oleh kustomer di *hotel airy rooms*. Data *airy rooms hotel reviews* memiliki 10 label, dan dibagi pada data *train* = 2.283, *validation* = 286, dan *test* = 285. Model klasifikasi yang digunakan, menggunakan model klasifikasi multi label. Hasil akurasi yang didapat pada model m-BERT-base = 90.9 dan distil-BERT-base = 87,8.

3 model yang telah dijelaskan diatas memiliki kelebihan masing-masing setiap model, seperti model BERT yang mampu memungkinkan untuk mempelajari konteks kata berdasarkan kata – kata disekitarnya, bukan hanya kata yang mendahului atau mengikutinya, model RoBERTa yang memiliki kelebihan sama seperti model BERT tetapi memiliki ukuran dan data yang lebih besar = 144 GB dan dapat melebihi kualitas dari model BERT, dan model DistilBERT yang memiliki ukuran model yang lebih kecil tapi memiliki kualitas 97% dari model BERT. Penelitian ini bertujuan dalam membangun model BERT, RoBERTa, dan DistilBERT untuk mengetahui kualitas masing-masing model pada kasus klasifikasi emosi data *text* Bahasa Indonesia dengan menggunakan proses *fine-tuning* dan *hyperparameter-tuning*.

Fine-tuning adalah teknik menggunakan model yang sudah terlatih pada tugas tertentu dan menyatukan beberapa *layer* terakhir dari model tersebut untuk memecahkan tugas baru dan *fine-tuning* dapat secara signifikan meningkatkan kinerja model (Sun et al., 2019). Proses *fine-tuning* dalam penelitian ini melakukan penambahan *output layer* pada model agar mengetahui hasil prediksi klasifikasi, penambahan *loss function* agar mengetahui nilai *loss* dari prediksi untuk label sebenarnya, dan melakukan *hyperparameter-tuning*. Data *text* yang digunakan dalam melakukan *fine-tuning* pada model menggunakan data *text* indo4B dan mengolah data tersebut dengan cara menstabilkan data *text* dengan cara mengikuti data terendah yang dimiliki oleh masing-masing label atau disebut dengan cara *undersample*.

Penelitian ini mengharapkan untuk mengetahui kualitasnya masing-masing model yang digunakan pada kasus klasifikasi emosi data *text* Bahasa Indonesia yang masih berjalan lambat dengan mengukur nilai ukuran, waktu dalam melakukan *training*, akurasi, dan *f1-score* pada model. Penelitian ini berkontribusi untuk membantu dalam melakukan klasifikasi emosi pada kalimat Bahasa Indonesia yang telah diungkapkan dalam bentuk *text* digital dengan berbagai tujuan seperti : melakukan evaluasi produk dan layanan suatu perusahaan yang telah diberi penilaian oleh kustomer dalam bentuk *text* digital dan evaluasi program pemerintahan dari suatu negara yang telah diberi penilaian oleh rakyat dalam bentuk *text* digital.

1.2 Rumusan Masalah

Penelitian dalam pembangunan model klasifikasi emosi data *text* Bahasa Indonesia mendapatkan beberapa rumusan masalah di antaranya:

1. Bagaimana menerapkan model BERT, RoBERTa, dan DistilBERT dalam membangun klasifikasi emosi data *text* Bahasa Indonesia dengan *fine-tuning* yang ditetapkan?
2. Bagaimana hasil ukuran, waktu *training*, akurasi, dan *f1-score* pada model BERT, RoBERTa, dan DistilBERT dalam membangun klasifikasi emosi data *text* Bahasa Indonesia dengan *fine-tuning* yang ditetapkan?

1.3 Tujuan Penelitian

Rumusan masalah yang telah dirumuskan dalam pembangunan model klasifikasi emosi data *text* Bahasa Indonesia, dapat dituliskan tujuan penelitian sebagai berikut:

1. Mengetahui penerapan model BERT, RoBERTa, dan DistilBERT dalam membangun klasifikasi emosi data *text* Bahasa Indonesia dengan *fine-tuning* yang ditetapkan.
2. Mengetahui hasil ukuran, waktu *training*, akurasi, dan *f1-score* pada model BERT, RoBERTa, dan DistilBERT dalam membangun klasifikasi emosi data *text* Bahasa Indonesia dengan *fine-tuning* yang ditetapkan.

1.4 Batasan Penelitian

Dalam melakukan penelitian klasifikasi emosi data *text* Bahasa Indonesia, digunakan beberapa batasan penelitian, yaitu:

1. Hanya menggunakan data *text* indo4B dari *twitter* yang menggunakan Bahasa Indonesia dengan memiliki 5 label yaitu *Fear*, *Anger*, *Sadness*, *Love*, *Happy* dan diolah dengan teknik *undersample*.
2. Hanya menggunakan model klasifikasi *single* label.
3. Hanya menggunakan 3 model : BERT, RoBERTa, DistilBERT, dan melakukan *fine-tuning* yang ditetapkan.

1.5 Sistematika Penulisan

Laporan penelitian pada penelitian ini menggunakan sistematika penulisan sebagai berikut.

BAB 1 PENDAHULUAN

Bagian pendahuluan memuat materi terkait latar belakang penelitian, rumusan masalah, tujuan penelitian, dan batasan penelitian.

BAB 2 TINJAUAN PUSTAKA

Bagian tinjauan pustaka menyuguhkan landasan teori yang berupa teori pendukung penelitian klasifikasi emosi data *text* dan berisi kajian pustaka terhadap penelitian-penelitian sebelumnya yang berkaitan dengan klasifikasi emosi data *text* Bahasa Indonesia.

BAB 3 METODOLOGI PENELITIAN

Bagian metodologi penelitian menyajikan metode atau langkah-langkah yang digunakan penulis untuk melakukan penelitian. Bagian ini memuat penjelasan singkat mengenai langkah penelitian, uraian metodologi, pengumpulan data, *preprocessing* data, pemodelan, dan matriks evaluasi.

BAB 4 HASIL DAN PEMBAHASAN

Bagian hasil dan pembahasan menyajikan hasil serta pembahasan dari implementasi klasifikasi emosi data *text* Bahasa Indonesia. Bagian ini penulis menjelaskan hasil pengumpulan data, hasil *preprocessing* data, hasil *hyperparameter-tuning* dan *fine-tuning*, hasil total parameter, hasil waktu *training*, hasil akurasi, hasil analisa peningkatan performa model, hasil prediksi klasifikasi emosi dan *f1-score*, dan hasil dari semua aspek.

BAB 5 PENUTUP

Bagian penutup memuat butir-butir kesimpulan penting yang dirangkum penulis dari penelitian yang telah dilakukan serta saran penelitian yang dapat dilakukan ke depannya di bidang klasifikasi emosi data *text* Bahasa Indonesia.

BAB 2

Tinjauan Pustaka

2.1 Landasan Teori

2.1.1 Klasifikasi Emosi

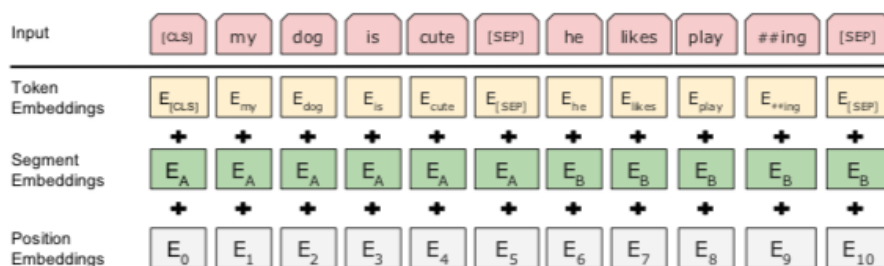
Analisis sentimen sudah sangat populer di bidang *artificial intelligence* yang membantu dalam menganalisa *text* digital. Analisis sentimen menggabungkan beberapa teknik dalam *artificial intelligence* seperti : NLP dan *deep learning* (Huan et al., 2022). Analisis sentimen memiliki beberapa peran penting untuk berbagai macam tujuan serta dapat diaplikasikan dalam berbagai skenario seperti : evaluasi produk dan layanan suatu perusahaan yang telah diberi penilaian oleh kustomer dalam bentuk *text* digital (Lin & Nuha, 2023), dan evaluasi program pemerintahan dari suatu negara yang telah diberi penilaian oleh rakyat dalam bentuk *text* digital. Analisis sentimen dapat memahami makna emosional kalimat dalam bentuk *text* digital yang mengandung kalimat bermakna positif dan negatif (Medhat et al., 2014). Kalimat positif dapat diberi persamaan terhadap makna emosional bahagia, cinta, dsb. Kalimat negatif dapat diberi persamaan terhadap makna emosional marah, sedih, dsb. 2 hasil yang didapat dalam analisis sentimen seperti positif dan negatif, analisis sentimen masih memiliki keterbatasan dalam memahami makna emosional kalimat dalam bentuk *text* digital (Aman & Szpakowicz, 2007), berbeda dengan klasifikasi emosi data *text*. Klasifikasi emosi data *text* memiliki tujuan yang sama seperti analisis sentimen, tetapi klasifikasi emosi data *text* memiliki kelebihan dalam memahami makna emosional kalimat dalam bentuk *text* digital dengan berbagai macam makna emosional (Acheampong et al., 2021).

Penelitian yang dilakukan oleh (Li & Xu, 2014) menjelaskan tentang kelebihan yang dimiliki klasifikasi emosi data *text* bahwa klasifikasi emosi data *text* dapat memberikan manfaat yang baik dalam pengambilan keputusan yang tepat dari informasi yang ada pada *text* digital yang di dapat dari berbagai macam media sosial seperti : *Instagram*, *facebook*, *twitter*, dsb. Perusahaan dapat mengelola data *text* tersebut dan melihat apa tanggapan emosional kustomer pada sebuah produk yang di jual oleh perusahaan dan tanggapan emosional masyarakat terhadap program departemen yang dilakukan.

2.1.2 BERT

Penelitian yang dilakukan oleh (Devlin et al., 2018) memperkenalkan model yang bernama BERT, arsitektur model BERT menggunakan *encoder* yang dimiliki transformer (Vaswani et al., 2017), model BERT memiliki (12 *encoder*, 768 *hidden layer*, 12 *self-attention*). Model BERT memiliki ukuran yang sama dengan model Open AI GPT tetapi BERT menggunakan *bidirectional self-attention* yang melakukan *self-attention* dengan 2 arah dari kata kiri ke kanan dan kanan ke kiri.

Input/output model BERT dapat dikondisikan dengan tugas yang akan dibangun dan mampu mewakili satu kalimat dan sepasang kalimat dalam satu urutan token. Urutan kalimat mengacu pada urutan token *input* model BERT yang dapat berupa satu kalimat atau dua kalimat yang digabungkan menjadi satu, model BERT menggunakan *wordpiece embeddings* dengan 30.000 kosa kata token, perbedaan antara berbagai kalimat dipisahkan dengan token khusus yaitu [SEP], token pertama sudah diatur menggunakan token [CLS] atau token klasifikasi khusus, pasangan keseluruhan kalimat dikemas menjadi satu rangkaian. *Input* token dibuat dengan 3 macam : token *embeddings* (token yang dimiliki masing-masing kata), *segment embeddings* (token yang menginformasikan kata yang dimiliki setiap kalimat), dan *position embeddings* (token yang menginformasikan urutan keseluruhan kata dalam kalimat), Gambar 2.1 menampilkan contoh token model BERT.



Gambar 2.1 Token model bert

Model BERT melakukan *pre-training* menggunakan 2 tugas : MLM dan NSP. MLM melakukan *pre-training* pada model BERT dengan fungsi untuk menutupi token *input* dari seluruh token *wordpiece embeddings* secara acak dengan ukuran 15%, agar model BERT dapat melatih dengan prediksi token yang tertutup tersebut, jika token ke i dipilih untuk melakukan MLM, token tersebut akan diubah menjadi token [MASK] sebanyak 80%, token acak sebanyak 10%, token yang tidak dirubah 10%. Hasil prediksi MLM ditampilkan pada token [MASK], token [MASK] di lanjut dengan fungsi aktivasi *softmax* dan menghitung *loss*

dengan *loss cross entropy*. Fungsi aktivasi *softmax* dapat dilihat di persamaan 2.1, *loss cross entropy* dapat dilihat di persamaan 2.2.

(2.1)

$$\text{softmax}(z_i) = \frac{\exp(z_i)}{\sum_{j=1}^V \exp(z_j)}$$

$\exp()$ = Fungsi eksponensial

z_i = Nilai skor atau logit untuk token ke i dalam kosa kata yang mungkin mengisi posisi yang di [MASK]

z_j = Nilai yang menunjukkan tingkat kecocokan bahwa token ke j adalah token yang benar pada saat [MASK]

V = Jumlah total token dalam kosa kata yang relevan untuk model

(2.2)

$$LCE = -\frac{1}{N} \sum_i^N \sum_j y_{ij} \log(\hat{y}_{ij})$$

y_i = Token aktual yang di [MASK] pada posisi i

$\hat{y}_i$ = Distribusi probabilitas prediksi token dari model untuk posisi i

N = Jumlah token yang di [MASK]

y_{ij} = *One-hot encoding* dari token aktual j pada posisi i , yaitu sebuah vector dengan nilai 1 pada indeks token aktual dan nilai 0 di indeks lainnya

$\hat{y}_{ij}$ = Distribusi probabilitas prediksi token j pada posisi i , yang diperoleh dari model

NSP melakukan *pre-training* pada model BERT dengan fungsi untuk memahami hubungan antar kalimat, seperti kalimat A dan B. 50% dari kalimat B diberi label *IsNext* (kalimat aktual berikutnya setelah kalimat A) dan 50% dari kalimat B diberi label *NotNext* (kalimat acak dari *corpus*). Hasil prediksi NSP ditampilkan pada token [CLS], token [CLS] di lanjut dengan fungsi aktivasi *softmax* dan menghitung *loss* dengan *loss cross entropy*. Fungsi aktivasi *softmax* dapat dilihat di persamaan 2.3, *loss cross entropy* dapat dilihat di persamaan 2.4, dan Gambar 2.2 menggambarkan struktur *pre-training* model BERT.

(2.3)

$$\text{softmax}(z) = \frac{\exp(z_{isnext})}{\exp(z_{isnext}) + \exp(z_{notnext})}$$

$\exp()$ = Fungsi eksponensial

z_{isnext} = Skor atau nilai logit untuk pasangan kalimat yang seharusnya berurutan

$z_{notnext}$ = Skor atau nilai logit untuk pasangan kalimat yang seharusnya tidak berurutan

(2.4)

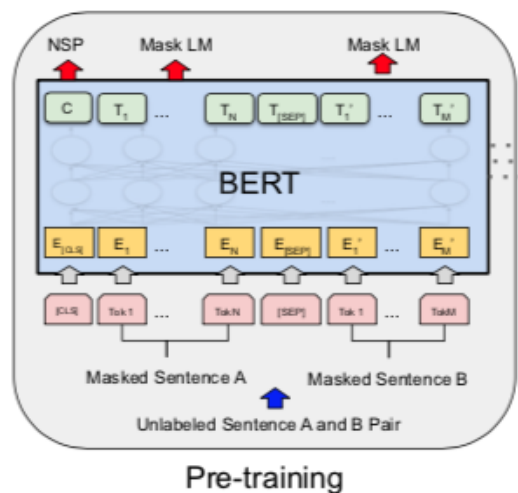
$$LCE = -(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y}))$$

y = Label yang sesungguhnya untuk pasangan 2 kalimat

$\hat{y}$ = Prediksi dari model untuk label tersebut

Jika $y = 1$, maka model diharapkan untuk memprediksi nilai $\hat{y}$ mendekati label 1 (menunjukkan bahwa 2 kalimat berikutnya dalam *text* sesuai dalam urutan aslinya)

Jika $y = 0$, maka model diharapkan untuk memprediksi nilai $\hat{y}$ mendekati label 0 (menunjukkan bahwa 2 kalimat tidak sesuai dalam urutan aslinya)



Gambar 2.2 Struktur *pre-training* model bert

Data *pre-training* yang digunakan model BERT adalah *BooksCorpus* (800 juta kata) dan *wikipedia* Bahasa Inggris (2.500 juta kata), *wikipedia* hanya mengekstrak bagian *text* dan mengabaikan daftar, tabel, dan *header*.

2.1.3 RoBERTa

Penelitian yang dilakukan oleh (Liu et al., 2019) mengusulkan modifikasi pada prosedur *pre-training* model BERT yang meningkatkan kinerja model dan diberi nama model RoBERTa. Penelitian ini juga mengatakan ada 2 faktor yang kurang ditekan pada model BERT : data yang digunakan dan *training* yang melewati data tersebut di proses lebih lama. RoBERTa melakukan *pre-training* mengikuti arsitektur BERTLARGE = (24 *encoder*, 1024 *hidden layer*, 16 *self-attention*), data yang digunakan sebanding dengan data model BERT, dan akurasi model RoBERTa mengalami peningkatan sebesar 3 dari model BERT, peningkatan tersebut disebabkan dengan model RoBERTa yang dilatih dengan *masking dynamic*, *full-sentences without NSP loss*, *large mini-batch* dan *large byte-level BPE*. *Masking dynamic* melakukan *masked* selama *training* pada model RoBERTa sebanyak 4 kali, model BERT melakukan *masked* 1 kali selama *training* (*masking static*), *full-sentences without NSP loss* melakukan *training* tanpa menggunakan NSP dan *input* data dengan panjang token 512, *large mini-batch* melakukan *training* dengan *mini-batch* yang besar = 8k, *large byte-level BPE* melakukan penambahan 20.000 kosa kata dari model BERT, sehingga model RoBERTa memiliki 50.000 kosa kata. Model RoBERTa juga melatih dengan data tambahan : *BooksCorpus* dan *wikipedia* 16 GB, *CC-news* 76 GB, *OpenWebText* 38 GB, *stories* 31 GB, total keseluruhan data adalah 160 GB dan melakukan 100K *steps*, 300k *steps*, 500k *steps* dalam *training*.

2.1.4 Distillation Pada BERT dan Menjadi DistilBERT

Distillation adalah teknik dalam *training* model, dimana model yang lebih kecil (siswa) dilatih untuk mereproduksi kualitas model yang lebih besar (guru). *Distillation* menghasilkan model yang lebih ringan dan memerlukan pelatihan komputasi yang lebih kecil. Model guru dan siswa dilatih dengan data yang sama, model guru dilatih terlebih dahulu dengan label asli untuk mendapatkan sebuah hasil prediksi dari model guru yang disebut *soft* label. Model siswa dilatih setelah model guru dan model siswa dilatih dengan *soft* label. Tujuan model siswa dilatih dengan *soft label* agar model siswa menghasilkan *training* yang baik meskipun model siswa model yang lebih kecil dengan memanfaatkan model guru. Fungsi aktivasi yang digunakan adalah *softmax temperature*, *temperature* berguna untuk mengontrol kelancaran *output* sesuai kebutuhan, *temperature* = 1 akan menghasilkan hasil seperti *softmax* normal, *temperature* > 1 akan menghasilkan hasil lebih lancar / distribusi probabilitasnya semakin merata, dan *temperature* < 1 akan menghasilkan hasil kurang lancar / distribusi probabilitasnya semakin tajam (Hinton et al., 2015),

temperature dari model guru dan siswa diberi nilai yang sama pada saat *training* tetapi pada saat inferensi, *temperature* diatur ke 1, fungsi aktivasi *softmax temperature* dapat dilihat di persamaan 2.5.

(2.5)

$$\text{softmax}(z_i) = \frac{\exp(z_i/T)}{\sum_{j=1}^n \exp(z_j/T)}$$

$\exp()$ = Fungsi eksponensial

T = Parameter *temperature* yang mengontrol seberapa tajam distribusi probabilitas yang dihasilkan, semakin tinggi T : distribusi probabilitasnya semakin merata dan semakin rendah T : distribusi probabilitasnya semakin tajam

z_i = Nilai skor atau logit (pada tugas yang digunakan)

z_j = Nilai yang menunjukkan tingkat kecocokan bahwa token ke j adalah token yang benar (pada tugas yang digunakan)

n = Jumlah total yang digunakan

Tugas dalam *distillation* pada BERT dan menjadi DistilBERT menggunakan 3 tugas : MLM menggunakan label asli, MLM menggunakan label guru, dan menambahkan *cosine embedding* yang menyelaraskan vektor dari model siswa dan guru. MLM menggunakan label asli menggunakan *masking dynamic* seperti model RoBERTa dan menghitung *loss* dengan *loss cross entropy*. MLM menggunakan label guru menggunakan *masking dynamic* seperti model RoBERTa dan menghitung *loss* dengan *loss cross entropy* atau disebut dengan *loss distillation*. *Cosine embedding* menggunakan 2 vektor : model guru dan model siswa yang akan dihitung keselarasan antara 2 vektor tersebut, *loss cosine embedding* dapat dilihat di persamaan 2.6. Tujuan akhir *training* adalah melakukan kombinasi linier dari *loss distillation*, *loss cross entropy*, dan *loss cosine cosine embedding*.

(2.6)

$$\text{Loss cosine embedding}(x_1, x_2, y) = \frac{1}{2} (1 - y \cdot \text{cosine_similarity}(x_1, x_2))^2$$

x_1 dan x_2 = 2 vector *embedding* dari model guru dan siswa

y = Label yang menunjukkan apakah dua teks tersebut *similarity* (y = 1) atau tidak mirip (y = -1)

cosine_similarity(x_1, x_2) = *Cosine similarity* antara x_1 dan x_2

2.1.4.1 DistilBERT

Penelitian yang dilakukan oleh (Sanh et al., 2020) menunjukkan bahwa kinerja serupa dapat dicapai menggunakan model bahasa yang jauh lebih kecil dengan melakukan *training* dengan *distillation*. Arsitektur dari model siswa (DistilBERT) memiliki arsitektur umum yang sama seperti model BERT, tetapi *token-type-embedding* dan *pooler* dihapus sementara selama jumlah *layer* dikurangi. Variasi pada dimensi *hidden size* memiliki dampak yang lebih kecil terhadap pengurangan parameter, dibandingkan variasi pada jumlah *layer*. Elemen penting dalam prosedur *training* DistilBERT adalah menemukan inisialisasi yang tepat agar sub-jaringan dapat menyatu, memanfaatkan dimensi umum antara jaringan guru dan siswa dengan menginisialisasi model siswa mengambil satu dari dua lapisan model guru. DistilBERT menggunakan *batch* yang sangat besar dengan memanfaatkan akumulasi gradien hingga 4k per *batch*, menggunakan *masking dynamic* dan tidak menggunakan NSP. Data dan daya komputasi melatih model DistilBERT menggunakan *corpus* model BERT. Model DistilBERT memiliki kelebihan : pengurangan sebesar 40% di ukuran model, penaikan kecepatan waktu dalam melakukan *training* model 60%, dan pencapaian kemampuan mempelajari konteks kata hingga 97% dalam kemampuan pemahaman bahasa.

2.1.5 Matriks Evaluasi

Matriks evaluasi model penelitian ini menggunakan model *confusion matrix*, model *confusion matrix* adalah salah satu model pengukuran keputusan dalam kasus *supervised learning* dan model evaluasi visual yang digunakan dalam *deep learning*. model *confusion matrix* memvisualisasikan tingkat kebingungan model *deep learning* dalam kelas yang berbeda. Model *confusion matrix* berisikan hasil kelas prediksi dan hasil kelas sebenarnya yang menghitung semua klasifikasi setiap label (Xu et al., 2020). Model *confusion matrix* yang digunakan pada penelitian ini adalah model *multiclass confusion matrix*. *Multiclass confusion matrix* cocok untuk mengatasi masalah dalam penelitian ini, karena kelas dalam penelitian ini memiliki lebih dari 2 kelas. *Multiclass confusion matrix* memiliki 4 kombinasi berbeda dari nilai prediksi dan nilai aktual yaitu *true positif*, *true negatif*, *false positif*, dan *false negatif* (Rahmad et al., 2020). *true positif*, *true negatif*, *false positif*, dan *false negative* dapat menghasilkan akurasi, *precision*, *recall*, dan *f1-score* (Markoulidakis et al., 2021). Tabel 2.1 menjelaskan tentang 4 kombinasi berbeda dari nilai prediksi dan nilai aktual. Kolom nama berisikan nama dari 4 kombinasi *Multiclass confusion matrix*, kolom singkatan berisikan singkatan dari kalimat yang berada pada kolom nama, dan kolom deskripsi berisikan penjelasan dari fungsi yang dimiliki pada kolom nama.

Tabel 2.1 4 Kombinasi Berbeda Dari Nilai Prediksi dan Nilai Aktual

Nama	Singkatan	Deskripsi
<i>True Positive</i>	TP	Model memprediksi positif dan itu benar
<i>True Negative</i>	TN	Model memprediksi negatif dan itu benar
<i>False Positive</i>	FP	Model memprediksi positif dan itu salah
<i>False Negative</i>	FN	Model memprediksi negatif dan itu salah

2.2 Kajian Pustaka

Jurnal klasifikasi emosi data *text* pada beberapa tahun terakhir telah banyak yang dipublikasikan. Penelitian-penelitian yang diterbitkan kebanyakan menggunakan model *deep learning*. Perkembangan model *deep learning* memiliki perkembangan yang pesat pada tugas klasifikasi emosi data *text* untuk meningkatkan performa dalam mendapatkan arsitektur-arsitektur pada tugas tertentu, data *text* yang digunakan dalam klasifikasi emosi data *text* kebanyakan menggunakan data *text* Bahasa Inggris, seperti penelitian yang dilakukan oleh (Acheampong et al., 2020) melakukan perbandingan kualitas antara model : BERT, RoBERTa, DistilBERT, dan XLNet pada klasifikasi emosi data *text*. Penelitian tersebut menggunakan data ISEAR yang berisi 7.666 *text*, 7 label (*anger, disgust, sadness, shame, fear, joy, dan guilt*) dan menggunakan Bahasa Inggris. 7 label ini diisi data *text* dengan ukuran yang sepadan = 1.096, data *training* diberi 80% dan data *testing* 20%. *Preprocessing* yang digunakan : *stopwords* dan *one-hot-encoding* pada label. Model BERT mendapatkan *f1-score* 70,2, RoBERTa 74,2, DistilBERT 67,2, dan XLNet 73,1.

Penelitian yang dilakukan oleh (Cortiz, 2020) melakukan perbandingan antara kualitas model : BERT, RoBERTa, DistilBERT, XLNet, dan Electra pada klasifikasi emosi data *text*. Penelitian tersebut menggunakan data GoEmotion yang berisi 58.000 *text*, 27 label (*admiration, amusement, anger, annoyance, approval, caring, confusion, curiosity, desire, disappointment, disapproval, disgust, embarrassment, excitement, fear, gratitude, grief, joy, love, nervousness, optimism, pride, realization, relief, remorse, sadness, surprise, neutral*) dan menggunakan Bahasa Inggris, *hyperparameter* yang digunakan *learning rate* = 5e-5, *batch size* = 16, *epochs* = 4, dan *threshold* = 0.30. Model BERT mendapatkan *f1-score* 46, RoBERTa 49, DistilBERT 48, XLNet 48, dan Electra 33. Waktu *training* model BERT mendapatkan 02:40:00, RoBERTa 01:08:56, DistilBERT 00:33:58, XLNet 01:33:49, dan Electra 00:13:04.

Penelitian yang dilakukan oleh (Rezapour, 2021) melakukan perbandingan antara kualitas model : *twitter-roberta-base-emoticon* dan *distilbert-base-uncased* pada klasifikasi

emosi data *text*. Penelitian tersebut menggunakan data GoEmotion yang berisi 58.000 *text*, 27 label (*admiration, amusement, anger, annoyance, approval, caring, confusion, curiosity, desire, disappointment, disapproval, disgust, embarrassment, excitement, fear, gratitude, grief, joy, love, nervousness, optimism, pride, realization, relief, remorse, sadness, surprise, neutral*) dan menggunakan Bahasa Inggris, *preprocessing* data *text* melakukan penghapusan, tanda baca, angka, dan tautan. *twitter-roberta-base-emoticon* menghasilkan *f1-score* 91 dan *distilbert-base-uncased* menghasilkan 88.

Penelitian yang dilakukan oleh (Avvaru & Vobilisetty, 2020) melakukan perbandingan antara kualitas model : BERT, DistilBERT, dan XLNet pada analisis sentimen. Penelitian tersebut menggunakan data Meme yang berisi 6601 dan dibagi menjadi 3 label : positif 3.864, netral 576, negatif 2.070, dan menggunakan Bahasa Inggris. *Preprocessing* penelitian ini memiliki *preprocessing* yang berbeda yaitu tidak menghilangkan tanda baca : !, karena bisa dapat untuk mengungkapkan kejutan dll, menghilangkan kata *we, wont, we've*, dll, dan menghilangkan huruf yang berlebih pada setiap kata seperti *soooo* : *so*. Model BERT menghasilkan *f1-score* 33,2, DistilBERT 30,2, dan XLNet 25,9.

Penelitian yang dilakukan oleh (Saputra et al., 2024) melakukan perbandingan antara kualitas model : BERT dan RoBERTa pada klasifikasi emosi data *text*. Penelitian tersebut menggunakan data *text twitter* yang berisi 119.536 dan dibagi menjadi 2 label : *stress* 59.768, *relax* 59.768, dan menggunakan Bahasa Inggris. *Hyperparameter* yang digunakan (*learning rate* 4e-5 dan *batch size* 32, menghasilkan *f1-score* BERT 84,7 dan RoBERTa 80,4), (*learning rate* 4e-5 dan *batch size* 64, menghasilkan *f1-score* BERT 84,0 dan RoBERTa 84,0), dan (*learning rate* 2e-5 dan *batch size* 32, menghasilkan *f1-score* BERT 84,7 dan RoBERTa 83,4).

Penelitian yang dilakukan oleh (Nimmi et al., 2022) melakukan perbandingan antara kualitas model : BERT, RoBERTa, dan DistilBERT pada klasifikasi emosi data *text*. Penelitian tersebut menggunakan data *text ERSS-tweet* yang berisi 5 label : *neutral* = 19.581, *anticipation* = 19.935, *anger* = 23.138, *sadness* = 19.962, *fear* = 21.099 dan menggunakan Bahasa Inggris. *Preprocessing* yang dilakukan pada penelitian ini : menghilangkan simbol pada kalimat, menghilangkan *emot* pada kalimat, dan menghilangkan *hashtags* (#) pada kalimat. *Hyperparameter* yang digunakan memiliki perbandingan yang sepadan, *learning rate* (1e-5, 5e-5, 3e-5, 4e-5, 2e-5), fungsi aktivasi (*softmax*), *batch size* (8, 16, 32), *epochs* (10), *maximum sequence length* (125), *dropout* (0,2), *hidden size* (768), dan *Optimizer* (adam). Model BERT mendapatkan *f1-score* terbaik = 83,55 dengan *hyperparameter* (*batch*

size 16, learning rate 1e-5), model RoBERTa mendapatkan *f1-score* terbaik = 84,47 dengan *hyperparameter (batch size 8, learning rate 2e-5)*, dan model DistilBERT mendapatkan *f1-score* terbaik = 83,52 dengan *hyperparameter (batch size 16, learning rate 2e-5)*.

Penelitian yang dilakukan oleh (Putri, 2023) melakukan perbandingan antara hasil akurasi dari *hyperparameter* yang digunakan : *learning rate* dan *adam optimizer*. Penelitian tersebut menggunakan data *text* ISEAR yang berisi 7666, memiliki 7 label : *joy, fear, anger, sadness, disgust, shame, guilt*. Data *text* ISEAR menggunakan Bahasa Inggris dan data *training* berisi 6132, data *testing* berisi 1534. Penelitian ini hanya menggunakan model DistilBERT. Akurasi terendah pada model DistilBERT saat menggunakan *hyperparameter learning rate* = 1e-03 dan *adam optimizer* = [0, 0,99], akurasi tertinggi pada model DistilBERT saat menggunakan *hyperparameter learning rate* = 1e-05 dan *adam optimizer* = [0, 0,99].

Penelitian yang dilakukan oleh (Abas et al., 2022) melakukan klasifikasi emosi data *text* dengan model BERT yang ditambahkan dengan model CNN dan menjadi model BERT-CNN, data *text* yang digunakan : ISEAR dan Semeval, data ISEAR berisi 7.666 *text*, 7 label (*anger, disgust, sadness, shame, fear, joy, dan guilt*) dengan Bahasa Inggris dan data Semeval berisi 38.424 *text*, 4 label (*happy, sad, angry, dan others*) dengan Bahasa Inggris. *Hyperparameter* yang digunakan *learning rate* = 4e-5, *loss function* : *categorical cross entropy*, *optimizer* : *adam*, *batch size* = 16, *dropout* = 0.5, *convolution size* = 3x3, *kernel size* = [3-5], dan *epochs* = 10. *F1-score* BERT-CNN pada data *semeval* = 94 dan data ISEAR = 76. Tabel 2.2 memberi informasi tentang penelitian sebelumnya dalam kasus klasifikasi emosi data *text* menggunakan data *text* Bahasa Inggris.

Tabel 2.2 Penelitian Sebelumnya Klasifikasi Emosi Data *Text* Bahasa Inggris

Judul	Penulis	Metode	Data
Comparative Analyses of BERT, RoBERTa, DistilBERT, and XLNet for Text-based Emotion Recognition	(Acheampong et al., 2020)	BERT, RoBERTa, DistilBERT, XLNet	ISEAR
Exploring Transformers in Emotion Recognition: a comparison of BERT, DistilBERT, RoBERTa, XLNet, dan ELECTRA	(Cortiz, 2020)	BERT, DistilBERT, RoBERTa, XLNet, ELECTRA	GoEmotion
Emotion Detection with Transformers: A Comparative Study	(Rezapour, 2021)	twitter-roberta-base-emoticon, distilbert-base-uncased	GoEmotion
BERT at SemEval-2020 Task 8: Using BERT to analyse meme emotions	(Avvaru & Vobilisetty, 2020)	BERT, DistilBERT, XLNet	Meme
Psychological Stress Detection Using a Transformer-Based Model	(Saputra et al., 2024)	BERT, RoBERTa	<i>Twitter</i>
Pre-trained ensemble model for identification of emotion during COVID-19 based on emergency response support system dataset	(Nimmi et al., 2022)	BERT, RoBERTa, DistilBERT	ERSS- <i>tweet</i>
DistilBERT with Adam Optimizer Tuning for Text-based Emotion Detection	(Putri, 2023)	DistilBERT	ISEAR
BERT-CNN: A deep learning model for detecting emotions from text	(Abas et al., 2022)	BERT-CNN	ISEAR

Penelitian klasifikasi emosi data *text* Bahasa Inggris telah memberi informasi penuh dalam perkembangannya dan perkembangan klasifikasi emosi data *text* Bahasa Indonesia menurut penelitian yang dilakukan oleh (Wilie et al., 2020) dan (Yanuar & Shiramatsu,

2020) memaparkan informasi bahwa pengembangan penelitian tentang klasifikasi emosi data *text* Bahasa Indonesia masih berjalan lambat, beberapa penelitian klasifikasi emosi data *text* Bahasa Indonesia yang telah dilakukan adalah penelitian yang dilakukan oleh (Budiman et al., 2024) melakukan perbandingan antara kualitas model : BERT dan IndoBERT pada analisis sentimen. Penelitian tersebut menggunakan data *text twitter* yang berisi 1.000 dan dibagi menjadi 2 label : positif 500, negatif 500, dan menggunakan Bahasa Indonesia. *Hyperparameter* yang digunakan : *batch size* 16, *learning rate* 2e-5, dan *epoch* 10. Pembagian data yang digunakan 80% untuk *training* dan 20% *testing*. Model BERT dan IndoBERT dibandingkan dengan data yang dilakukan *preprocessing* dan tidak dilakukan *preprocessing*. Model BERT *preprocessing* menghasilkan akurasi 82 dan tidak dilakukan *preprocessing* 81,5, model IndoBERT *preprocessing* menghasilkan akurasi 94 dan tidak dilakukan *preprocessing* 89,5.

Penelitian yang dilakukan oleh (Christian et al., 2021) melakukan perbandingan antara kualitas model : BERT, RoBERTa, XLNet pada klasifikasi emosi data *text*. Penelitian tersebut menggunakan *facebook* data set, dibagi menjadi 2 label dan menggunakan Bahasa Indonesia, *facebook* data set dibagi menjadi 5 : data *openness*, *conscientiousness*, *extraversion*, *agreeableness*, dan *neuroticism*. Data *training* diberi 80%, *validation* 10%, *test* 10%. *Preprocessing* menggunakan penghapusan simbol, URL, penerjemahan pada bahasa lain, *stopwords*, *stemming*, dan *lower case*. Rata-rata dari keseluruhan nilai *f1-score* setiap data pada model BERT tanpa *preprocessing* = 68,8, RoBERTa tanpa *preprocessing* = 67,7, dan XLNet tanpa *preprocessing* = 68,3. Model BERT menggunakan *preprocessing* = 71,8, RoBERTa menggunakan *preprocessing* = 72,0, dan XLNet menggunakan *preprocessing* = 72,3.

Penelitian yang dilakukan oleh (Fathan Hidayatullah et al., n.d.) melakukan perbandingan antara kualitas model : m-BERT, XLM-RoBERTa pada klasifikasi emosi data *text*. Penelitian tersebut menggunakan *twitter* data set dan dibagi menjadi 2 label : *adult*, *non adult* dan menggunakan Bahasa Indonesia, *adult* memiliki 27.146 dan *non adult* memiliki 27.192. *Preprocessing* menggunakan penghapusan simbol, URL, menghilangkan *hashtags* (#), *user*, *retweet*, angka, tanda baca, dan emoji. *Hyperparameter* yang digunakan mBERT : *learning rate* 3e-05, *batch size* 16, *decay* 0,01 dan XLM-RoBERTa : *learning rate* 3e-05, *batch size* 8, *decay* 0,001. M-BERT menghasilkan *f1-score* 94,48 dan XLM-RoBERTa 96,61.

Penelitian yang dilakukan oleh (Wiciaputra et al., 2021) melakukan klasifikasi data *text* Bahasa Indonesia dengan model XLM-RoBERTa tanpa melakukan uji coba

hyperparameter-tuning, tetapi menggunakan 5 data *text* yang berbeda dan digabungkan, seperti *english news dataset* = 37.886 Bahasa Inggris, *mixed news dataset* = 75.772 dari 37.886 Bahasa Inggris dan 37.886 Bahasa Indonesia, *mixed news dataset* = 108.190 dari Bahasa Inggris 37.886 dan 70.304 Bahasa Indonesia, *english news dataset* = 2.000, dan *mixed news dataset* = 4.000 dari 2.000 Bahasa Inggris dan 2.000 Bahasa Indonesia. Penelitian ini menghasilkan *f1-score* terbaik pada data *mixed news dataset* = 75.772 dengan *f1-score* = 93.30.

Penelitian yang dilakukan oleh (Nissa & Yulianti, 2023) menerapkan model m-BERT-base dan distil-BERT-base dengan model multilingual untuk melakukan klasifikasi data *text* Bahasa Indonesia, penelitian ini menggunakan data *airy rooms hotel reviews* yang berisi tentang komentar pada bagian/ruangan mana yang telah diberi penilaian oleh kustomer di *hotel airy rooms*. Data *airy rooms hotel reviews* memiliki 10 label, dan dibagi pada data *train* = 2.283, *validation* = 286, *test* = 285, dan menggunakan Bahasa Indonesia. Model label klasifikasi yang digunakan, menggunakan model label klasifikasi multi label. Hasil akurasi yang didapat pada model m-BERT-base = 90.9 dan distil-BERT-base = 87,8. Tabel 2.3 memberi informasi tentang penelitian sebelumnya dalam kasus klasifikasi emosi data *text* menggunakan data *text* Bahasa Indonesia.

Tabel 2.3 Penelitian Sebelumnya Klasifikasi Emosi Data *Text* Bahasa Indonesia

Judul	Penulis	Metode	Data
Classification Performance Comparison of BERT and IndoBERT on Self-Report of COVID-19 Status On Social Media	(Budiman et al., 2024)	BERT, IndoBERT	<i>Twitter</i>
Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging	(Christian et al., 2021)	BERT, RoBERTa, XLNet	<i>Facebook</i>
Adult Content Detection on Indonesian Tweets by Fine-tuning Transformer-based Models	(Fathan Hidayatullah et al., n.d.)	m-BERT, XLM-RoBERTa	<i>Twitter</i>
Bilingual text classification in english and Indonesian via transfer learning using XLM-RoBERTa	(Wiciaputra et al., 2021)	XLM-RoBERTa	<i>English news dataset, mixed news dataset</i>
Multi-label text classification of Indonesian customer reviews using bidirectional encoder representations from transformers language model	(Nissa & Yulianti, 2023)	m-BERT-base, distil-BERT-base	<i>Airy rooms hotel reviews</i>

Penelitian klasifikasi emosi data *text* Bahasa Indonesia yang ditampilkan pada Tabel 2.3 menggunakan perbandingan antara model (BERT, IndoBERT), (BERT, RoBERTa, XLNet), (m-BERT, XLM-RoBERTa), XLM-RoBERTa, dan (m-BERT-base, distil-BERT-base) masih memiliki keterbatasan dalam perbandingan model, penelitian ini melihat kelebihan dari model BERT yang mampu memungkinkan untuk mempelajari konteks kata berdasarkan kata – kata disekitarnya, bukan hanya kata yang mendahului atau mengikutinya, model RoBERTa yang memiliki kelebihan sama seperti model BERT tetapi memiliki ukuran dan data yang lebih besar = 144 GB dan dapat melebihi kualitas dari model BERT, dan model DistilBERT yang memiliki ukuran model yang lebih kecil tapi memiliki kualitas 97% dari model BERT. Penelitian ini mengharapkan untuk mengetahui kualitasnya masing-masing model yang digunakan pada kasus klasifikasi emosi data *text* Bahasa Indonesia yang

masih berjalan lambat dengan mengukur nilai ukuran, waktu dalam melakukan *training*, akurasi, dan *f1-score* pada model.

Penelitian ini memiliki perbandingan dengan penelitian sebelumnya seperti : menggunakan model BERT, RoBERTa, dan DistilBERT untuk mengetahui kualitas masing-masing model pada kasus klasifikasi emosi data *text* Bahasa Indonesia dengan menggunakan proses *fine-tuning* yang melakukan penambahan *output layer* pada model agar mengetahui hasil prediksi klasifikasi dan penambahan *loss function* agar mengetahui nilai *loss* dari prediksi untuk label sebenarnya, *hyperparameter-tuning* dan menggunakan data *text* indo4B Berbahasa Indonesia yang diolah dengan teknik menstabilkan data *text* tersebut dengan cara mengikuti data terendah yang dimiliki oleh masing – masing label atau disebut dengan cara *undersample*.

BAB 3

Metodologi

3.1 Langkah Penelitian

Bab ini akan diuraikan metodologi yang digunakan dalam melakukan klasifikasi emosi data *text* Bahasa Indonesia menggunakan model BERT, RoBERTa, dan DistilBERT. Terdapat 4 tahap yang dilakukan : pengumpulan data, *preprocessing* data, pemodelan, dan matriks evaluasi. Diagram alir klasifikasi emosi data *text* Bahasa Indonesia dengan model BERT, RoBERTa, dan DistilBERT dapat dilihat pada Gambar 3.1.



Gambar 3.1 Diagram alir klasifikasi emosi data *text* bahasa indonesia dengan model bert, roberta, dan distilbert

3.2 Uraian Metodologi

Bagian ini menguraikan langkah penelitian klasifikasi emosi data *text* Bahasa Indonesia dengan model BERT, RoBERTa, dan DistilBERT sesuai dari diagram alir pada Gambar 3.1. Diagram alir berisikan : pengumpulan data, *preprocessing* data, pemodelan, dan matriks evaluasi. Tahap pertama melakukan pengumpulan data yang terdapat dari indo4B, tahap kedua melakukan beberapa model *preprocessing* data sebelum data *text* dimasukkan ke dalam model BERT, RoBERTa, dan distilBERT, tahap ketiga memberikan *set* latih kedalam model BERT, RoBERTa, dan DistilBERT, dan tahap terakhir melakukan evaluasi dengan menggunakan matriks evaluasi untuk mengevaluasi model dengan akurasi, *precision*, *recall*, dan *f1-score*.

3.2.1 Pengumpulan Data

Penelitian ini mendapatkan 1 sumber data *text*, yaitu data indo4B yang dilakukan pada penelitian (Wilie et al., 2020), data indo4B diambil dari situs *twitter*. *Twitter* adalah situs yang sangat populer dalam mencari informasi sosial yang tepat waktu seperti berita terkini, postingan selebriti, topik yang sedang tren (Lee et al., 2011), kehidupan sehari-hari, komentar film (Li & Xu, 2014), dan acara sosial (Li & Xu, 2014). Pengguna dapat posting

data *text* singkat dengan ukuran panjang 140 karakter (Lee et al., 2011). Layanan *twitter* bertindak sebagai pusat berita organisasi seperti perusahaan dan departemen pemerintahan. *Twitter* menjadi platform dimana informasi, ide, dan opini bertemu. Pengguna *twitter* banyak yang membuat keputusan dibawah pengaruh *twitter* yang mereka ikuti. Informasi atau postingan dari *twitter* dianggap sebagai sumber data *text* yang memiliki makna opini yang luas (Li & Xu, 2014). Kelebihan dalam sumber data *text* yang kaya dan lengkap, penelitian ini menggunakan data *text twitter* dalam membangun model klasifikasi emosi data *text* Bahasa Indonesia.

Data indo4B memiliki 5 label, *Fear*, *Anger*, *Sadnes*, *Love*, dan *Happy*. Pemilihan label emosi tersebut menggunakan emosi dasar Shaver (Shaver, 1987), sebagian besar klasifikasi emosi data *text* menggunakan leksikon, leksikon adalah keseluruhan kata terkecil yang terdapat pada suatu bahasa, dan kebanyakan klasifikasi emosi pada Bahasa Indonesia bukan berbasis leksikon. Leksikon untuk Bahasa Inggris ada berbagai macam, namun hanya ada satu leksikon Bahasa Indonesia yang dikembangkan pada penelitian yang dilakukan oleh (Shaver, 1987). Penelitian (Dong, 2018) dan penelitian (Wilie et al., 2020) juga menggunakan label emosi yang mengacu pada penelitian yang dilakukan oleh (Shaver, 1987).

Pelabelan data indo4B dilakukan oleh 2 ahli bahasa (Dong, 2018), data indo4B memiliki 3521 data *training*, 440 data *validation*, dan 442 data *test*, data tersebut memiliki 5 label. Berbeda dari penelitian klasifikasi emosi data *text* Bahasa Indonesia yang telah dilakukan oleh (Wilie et al., 2020), data *text* yang digunakan pada penelitian ini menggunakan data *text* indo4B dan menstabilkan data *text* tersebut dengan cara mengikuti data terendah yang dimiliki oleh masing-masing label atau disebut dengan cara *undersample*, *undersample* digunakan karena keseimbangan sebuah data berpengaruh terhadap kualitas sebuah model, seperti penelitian yang dilakukan oleh (Kaur et al., 2019) mengatakan, model klasifikasi emosi data *text* dengan data yang tidak seimbang adalah masalah dimana kelas mayoritas akan mengarah pada akurasi yang lebih baik, dan kelas minoritas akan mengarah akurasi yang kurang baik. Pembagian dataset penelitian ini dibagi menjadi 3 data yaitu : data *training* 80%, data *validation* 10%, dan data *test* 10% (Wilie et al., 2020). Tabel 3.1 memberikan informasi contoh data *text* indo4B. Kolom *tweet* berisikan contoh kalimat yang ada pada data *text* indo4B, dan kolom label berisikan tentang label yang dimiliki setiap kalimat.

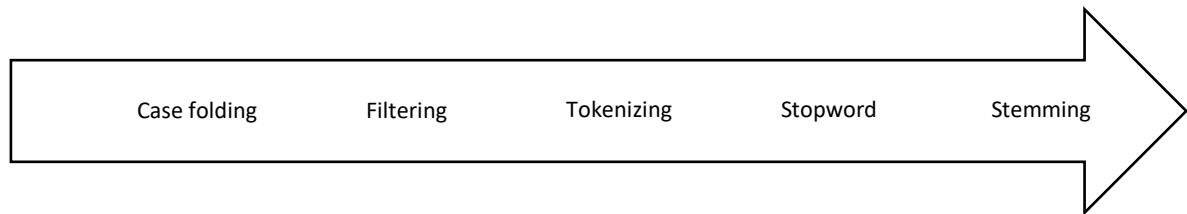
Tabel 3.1 Contoh Data *Text* Indo4b

<i>Text</i>	<i>Label</i>
"[USERNAME] [USERNAME] Bcs doi w anak pelayaran, tp emg beneran sih ditinggal 5 bulan gt, ketemu cm 3 hari dri 5 bulan itu Parahnya atasannya yg udh beristri main cewek gtu, kan ngeri :("	<i>Fear</i>
"[USERNAME] [USERNAME] [USERNAME] [USERNAME] [USERNAME] [USERNAME] [USERNAME] [USERNAME] Sampaikan juga informasi kegiatan dana desa, jangan hanya kirim hastag dan mantion saja"	<i>Anger</i>
"Kapan bandara [USERNAME] 1 di Ambon ada perbaikan dan mungkin renofasi. Melihat ada banyak perkembangan bandara di daerah lain rasanya sudah amat ketinggalan sebagai bandara internasional . [URL]"	<i>Sadness</i>
"[USERNAME] Apaa eh sengajanya tu , sayang dah baca b ku sayang sudah take note already . Baby love you unconditionally. Baby cinta b right and treat my king well arasso? . Kiss sini MUAHHHH!"	<i>Love</i>
[USERNAME] Wuish [USERNAME] saya tungguin nih film nya... Bikin film superhero trus kirim ke negara tetangga lagi. Biar film indonesia jadi trend.	<i>Happy</i>

3.2.2 *Preprocessing Data*

Preprocessing data adalah suatu proses pengubahan bentuk suatu data *text* menjadi lebih terstruktur dan sesuai dengan kebutuhan dalam proses data *mining* (Huang et al., 2015). Gambar 3.2 menunjukkan urutan tahap *preprocessing* data. Tahap pertama melakukan cara *case folding*, *case folding* akan mengubah huruf besar menjadi huruf kecil, *casefolding* memberi kontribusi signifikan terhadap model, karena penelitian ini menggunakan model *uncased*, model *uncased* akan mengelola data yang menggunakan huruf kecil semua, tahap kedua melakukan cara *filtering*, *filtering* akan menghapus berbagai kata yang telah dipilih pada penelitian ini, pilihan kata tersebut adalah @username, *retweet*, #*hashtags*, *emoticon*, dan *URL*, karena kata tersebut termasuk kata yang memiliki makna sederhana untuk kasus klasifikasi emosi data *text*, tahap ketiga melakukan cara *tokenizing*, *tokenizing* akan melakukan pemotongan *text* menjadi setiap bagian kata yang disebut token, *tokenizing* memberi kontribusi signifikan dalam pemrosesan *preprocessing* data lainnya, tahap keempat melakukan cara *stopword*, *stopword* akan menghapus kata tambahan depan agar menjadi kata dasar, *stopword* memberi kontribusi signifikan terhadap makna kata suatu data *text*, dan tahap kelima melakukan cara *stemming*, *stemming* akan mengubah kata yang memiliki kata imbuhan menjadi kata dasar, *stemming* memberi kontribusi signifikan terhadap

mempertahankan akar kata yang sama (Hidayatullah & Ma'Arif, 2017). *Preprocessing* data akhir akan mengubah label menjadi model *one hot encoding*.



Gambar 3.2 Urutan tahap *preprocessing* data

3.2.3 Pemodelan

Penelitian ini melakukan pembangunan model BERT, RoBERTa, dan DistilBERT menggunakan proses *fine-tuning* dan *hyperparameter-tuning*. *Fine-tuning* adalah teknik menggunakan model yang sudah terlatih pada tugas tertentu dan menyatukan beberapa *layer* terakhir dari model tersebut untuk memecahkan tugas baru dan *fine-tuning* dapat secara signifikan meningkatkan kinerja model (Sun et al., 2019). Pembangunan klasifikasi emosi data *text* Bahasa Indonesia pada Model BERT, RoBERTa, dan DistilBERT pada penelitian ini mengambil model yang sudah jadi dan diolah dengan cara *fine-tuning*:

1. Model BERT : bert-base-uncased model yang menggunakan 12 model *hidden-layer*, 12 model *multi-head-attention*, dan 768 model *hidden size*. Panjang token berisi 512 dan vocab_size = 30522 (Devlin et al., 2018).
2. Model RoBERTa : roberta-base model yang menggunakan 12 model *hidden-layer*, 12 model *multi-head-attention*, dan 768 model *hidden size*. Panjang token berisi 512 dan vocab_size = 50265 (Liu et al., 2019).
3. Model DistilBERT : distilbert-base-uncased model yang menggunakan 6 model *hidden-layer*, 12 model *multi-head-attention*, dan 768 model *hidden size*. Panjang token berisi 512 dan vocab_size = 30522 (Sanh et al., 2020).

3 model diatas, akan dilakukan *fine-tuning* seperti : Penambahan *output layer* dan *loss function* : *cross entropy loss* pada model BERT, RoBERTa, dan DistilBERT. *Output layer* berfungsi untuk mengetahui hasil prediksi klasifikasi, fungsi aktivasi yang digunakan pada *output layer* adalah fungsi aktivasi *softmax* dan penambahan *loss cross entropy* yang digunakan untuk menghitung perbedaan antara keluaran yang dibuat oleh model dan keluaran yang diharapkan, Fungsi aktivasi *softmax* dapat dilihat di persamaan 3.1, *loss cross entropy* dapat dilihat di persamaan 3.2. Gambaran klasifikasi emosi data *text* menggunakan *fine-tuning* pada model BERT, RoBERTa, dan DistilBERT dapat dilihat pada Gambar 3.3.

(3.1)

$$\text{softmax}(z_j) = \frac{\exp(z_j)}{\sum_{k=1}^K \exp(z_k)}$$

$\exp()$ = Fungsi eksponensial

K = Jumlah kelas yang mungkin dalam konteks klasifikasi

$\exp(z_j)$ = Eksponensial dari nilai *input* z untuk kelas j

$\sum_{k=1}^K \exp(z_k)$ = Jumlah dari eksponensial dari semua nilai *input* z untuk semua kelas

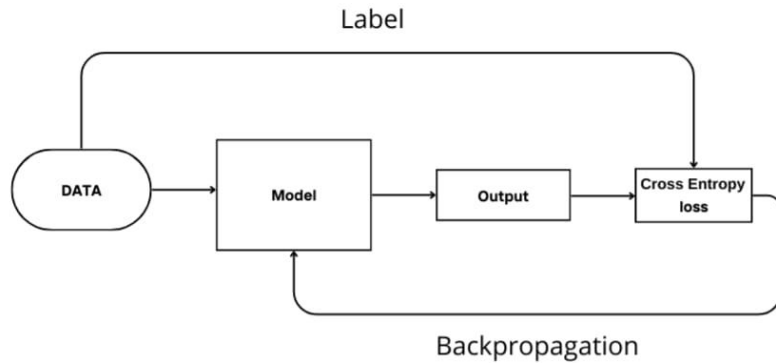
(3.2)

$$LCE = - \sum_{i=1}^K y_i \log(\hat{y}_i)$$

K = Jumlah kelas yang mungkin dalam konteks klasifikasi

y_i = Elemen ke i dari vektor label sebenarnya, nilainya adalah 1 jika contoh tersebut adalah kelas ke i , 0 untuk kelas lainnya (*one-hot-encoding*)

$\hat{y}_i$ = Elemen ke i dari vector *output* prediksi *softmax* dari model, ini adalah probabilitas prediksi bahwa contoh tersebut masuk dalam kelas i



Gambar 3.3 Gambaran klasifikasi emosi data *text* menggunakan *fine-tuning* pada model bert, roberta, dan distilbert

Hyperparameter-tuning terbaik model yang digunakan ditentukan sesuai dengan fluktuasi *gradient*, *batch size* $\{16, \dots, 320\}$, *learning rate* $\{0.01, \dots, 0.20\}$, *epoch* $\{10, \dots, 25\}$ (Hastomo et al., 2021), dan *optimizer* menggunakan *adam optimizer* dengan *beta* 1 ($\beta_1 =$

0.9) dan β_2 ($\beta_2 = 0.999$) (Sun et al., 2019) akan dilakukan penyesuaian (He et al., 2019).

3.2.4 Matriks Evaluasi

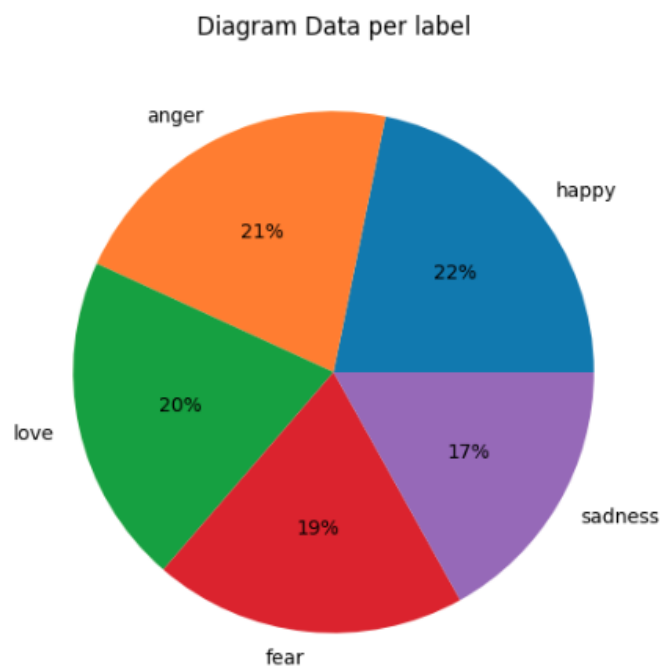
Matriks evaluasi model penelitian ini menggunakan model *confusion matrix*, model *confusion matrix* adalah salah satu model pengukuran keputusan dalam kasus *supervised learning* dan model evaluasi visual yang digunakan dalam *deep learning*. Model *confusion matrix* memvisualisasikan tingkat kebingungan model *deep learning* dalam kelas yang berbeda. Model *confusion matrix* berisikan hasil kelas prediksi dan hasil kelas sebenarnya yang menghitung semua klasifikasi setiap label (Xu et al., 2020). Model *confusion matrix* yang digunakan pada penelitian ini adalah model *multiclass confusion matrix*. *Multiclass confusion matrix* cocok untuk mengatasi masalah dalam penelitian ini, karena kelas dalam penelitian ini memiliki lebih dari 2 kelas. Matriks evaluasi model dapat menghasilkan akurasi, *precision*, *recall*, dan *f1-score* (Markoulidakis et al., 2021).

BAB 4

Hasil dan Pembahasan

4.1 Hasil Pengumpulan Data

Pembangunan model pada penelitian ini terdapat 3401 data *text* yang terbagi pada 5 label dan membagi menjadi 3 bagian yaitu, data *training* 80%, data *validation* 10%, dan data *test* 10%. 3401 data *text* tersebut sudah dilakukan keseimbangan data menggunakan cara *undersample* agar data setiap label seimbang. Keseimbangan data dapat dilihat pada Gambar 4.1. Contoh dan pembagian data per label dapat dilihat pada Tabel 4.1. Tabel 4.1 memberikan informasi contoh dan pembagian data per label. Kolom *text* berisikan contoh kalimat pada data *text*, kolom label berisikan tentang label yang dimiliki setiap kalimat, kolom *training & validation* berisikan total data *text* yang dimiliki data *training* digabung dengan data *validation*, kolom *test* berisikan total data *text* yang dimiliki data *test*, dan kolom total berisikan jumlah data *text* dari data *training & validation* dan data *test*.



Gambar 4.1 Keseimbangan data per label

Tabel 4.1 Contoh dan Pembagian Data per Label

<i>Text</i>	<i>Label</i>	<i>Training & Validation</i>	<i>Test</i>	<i>Total</i>
TAMSIL ISLAM NUSANTARA. Karena kau cuma tiang yang dipancang tergesa-gesa kemarin sore dan pangkalmu cuma dangkal-dangkal saja ditanam, maka engkau jadi takut setengah mati pada angin. Bahkan semilir yang segar pun kau caci dan kau kutuki.	<i>Fear</i>	597	69	666
temen2 masa kecil yang turned out being asshole itu kalo cewe jadi ukhti2 hijrah kalo cowo jadi abdi negara, rasanya semua statement ku di sosmed pasti yang menyangkal orang2 kek gitu, parahnya lagi kalo mereka pacaran bangsat	<i>Anger</i>	657	64	721
berita Via Vallen dapet DM dari pemain sepak bola yang terkesan ""menggoda"" atau sexual harassment, liat komennya pada bilang lebay dan gak harus di publish karena hal kek gitu ""biasa aja"" dan parahnya yang nyinyir perempuan juga :) miris	<i>Sadness</i>	520	69	589
Semoga sehat selalu, dan yg terpenting selalu sayang dan cinta ama [USERNAME], walaupun gw itu followers baru tp kita bisa saling kenal krn kita keluarga we are family I love you	<i>Love</i>	630	63	693
Bahagia beli paket segiga habis limagiga lancar internetan mayanlah tertawa alhamdulillah	<i>Happy</i>	670	62	732

4.2 Hasil *Preprocessing* Data

Hasil dari *preprocessing* data dapat dilihat pada Tabel 4.2. Kolom *text* berisikan contoh data *text* sebelum dilakukan *preprocessing* data dan kolom *preprocessing text* berisikan contoh data *text* setelah dilakukan *preprocessing* data. Hasil *preprocessing* data :

- *Case folding* dapat dilihat bahwa semua huruf yang ada pada kalimat sudah menjadi huruf kecil semua.

- *Filtering* dapat dilihat bahwa kata @username, retweet, #hashtags, emoticon, dan URL sudah terhapus.
- *Stopword* dapat dilihat bahwa kata tambahan depan sudah terhapus seperti : dipancang : pancang, tergesa : gesa, ditanam : tanam,
- *Stemming* dapat dilihat bahwa kata imbuhan sudah terhapus seperti : pangkalmu : pangkal, kutuki : kutuk.

Preprocessing data akhir mengubah label menjadi model *one hot encoding* dapat dilihat pada Tabel 4.3. Kolom label berisikan label yang digunakan, dan kolom *one hot encoding* berisikan hasil dari label dari data *text* dan dirubah menjadi model *one hot encoding*.

Tabel 4.2 Hasil *Preprocessing* Data

<i>Text</i>	<i>Preprocessing Text</i>
TAMSIL ISLAM NUSANTARA. Karena kau cuma tiang yang dipancang tergesa-gesa kemarin sore dan pangkalmu cuma dangkal-dangkal saja ditanam, maka engkau jadi takut setengah mati pada angin. Bahkan semilir yang segar pun kau caci dan kau kutuki	tamsil islam nusantara tiang pancang gesa gesa kemarin sore pangkal dangkal dangkal tanam engkau takut mati angin milir segar caci kutuk
Kalo kamuuuuuu, kamuu itu speciallll. Cintanya juga ga kalah kok sama yang lain. Karena sebenarnya bagi aku kalian itu satu. Jadi, cinta satu cinta semua. [URL]	kamu spesial cinta kalah sebenarnya cinta cinta
temen2 masa kecil yang turned out being asshole itu kalo cewe jadi ukhti2 hijrah kalo cowo jadi abdi negara, rasanya semua statement ku di sosmed pasti yang menyangkal orang2 kek gitu, parahnya lagi kalo mereka pacaran bangsat	teman turned makhluk asshole cewek ukhti hijrah pria abdi negara nyata sosial sangkal orang parah hubung bangsat
Buat saya, apa yang dilakukan Via Valen simbol perlawanan yang perlu didukung, apapun profesinya wanita tetap harus dihormati dan dihargai [URL]"	valen simbol lawan dukung apa profesi wanita hormat harga
Ini adalah hal yang paling membahagiakan saat biasku foto bersama ELF #ReturnOfTheLittlePrince #HappyHeeChulDay	bahagia foto

Tabel 4.3 Hasil *One Hot Encoding* Pada Label

Label	<i>One Hot Encoding</i>
<i>Anger</i> [0]	[1., 0., 0., 0., 0.,]
<i>Fear</i> [1]	[0., 1., 0., 0., 0.,]
<i>Happy</i> [2]	[0., 0., 1., 0., 0.,]
<i>Love</i> [3]	[0., 0., 0., 1., 0.,]
<i>Sadness</i> [4]	[0., 0., 0., 0., 1.,]

4.3 Hasil *Hyperparameter-tuning* dan *Fine-tuning* per Model

Tabel 4.4 menampilkan arsitektur *hyperparameter-tuning* pada model BERT, RoBERTa, DistilBERT. Kolom model berisikan model yang digunakan, kolom *batch size* berisikan ukuran *batch size* yang digunakan pada model tersebut, kolom *dropout* berisikan ukuran *dropout* yang digunakan pada model tersebut, kolom *att dropout* berisikan *att dropout* yang digunakan pada model tersebut, kolom *epoch* berisikan *epoch* yang digunakan pada model tersebut, dan kolom *learning rate* berisikan *learning-rate* yang digunakan pada model tersebut.

Tabel 4.4 Arsitektur *Hyperparameter-tuning* Pada Model Bert, Roberta, Distilbert

Model	<i>Batch Size</i>	<i>Dropout</i>	<i>Att Dropout</i>	<i>Epoch</i>	<i>Learning Rate</i>
BERT	16	0.1	0.1	10	1e-5
RoBERTa	16	0.1	0.1	10	1e-5
DistilBERT	16	0.1	0.1	10	1e-5

Tabel 4.5 menampilkan arsitektur *fine-tuning* pada model BERT, RoBERTa, DistilBERT. Kolom no berisikan nomor model pada tabel, kolom model berisikan nama model, dan kolom *layer* berisikan informasi tentang isi yang ada pada *layer*.

Tabel 4.5 Arsitektur *Fine-tuning* Pada Model Bert, Roberta, Distilbert

No	Model	Layer	
1.	BERT	4	Input_ids = 256
			Attention_mask = 256
			Bert = 768
			Output_layer = 5
2.	RoBERTa	4	Input_ids = 256
			Attention_mask = 256
			Roberta = 768
			Output_layer = 5
3.	DistilBERT	4	Input_ids = 256
			Attention_mask = 256
			Distilbert = 768
			Output_layer = 5

Kolom *Layer* Tabel 4.5 menampilkan hasil struktur yang sudah dibangun seperti pada Gambar 3.3 :

1. Input_ids adalah *layer input* data *text* yang sudah di rubah dengan *word embedding* dengan panjang 256.
2. Attention_mask adalah *layer input* data *text* yang sudah dirubah dengan *word embedding* dan di tutup dengan angka 1 dengan panjang 256.
3. Bert, Roberta, Distilbert adalah layer yang menerima data dari Input_ids dan Attention_mask dan diolah dengan struktur masing-masing model dengan ukuran *hidden size* 768.
4. Output_layer adalah layer yang menerima hasil dari model dan memproses dengan *neural network* untuk mendapatkan hasil klasifikasi dengan 5 label.

4.4 Hasil Total Parameter Setiap *Hyperparameter-tuning* dan *Fine-tuning* per Model

Tabel 4.6 menampilkan hasil total parameter dari arsitektur *hyperparameter-tuning* dan *fine-tuning* pada model BERT, RoBERTa, DistilBERT . Kolom no berisikan nomor model pada tabel, kolom model berisikan nama model, kolom *layer* berisikan informasi tentang isi yang ada pada *layer*, dan kolom parameter berisikan tentang total parameter yang dimiliki setiap model.

Tabel 4.6 Hasil Total Parameter Dari Arsitektur *Hyperparameter-tuning* dan *Fine-tuning* Pada Model Bert, Roberta, Distilbert

No	Model	Layer		Parameter
1.	BERT	4	Input_ids = 256	0
			Attention_mask = 256	0
			Bert = 768	109482240
			Output_layer = 5	3845
			T	109486085
2.	RoBERTa	4	Input_ids = 256	0
			Attention_mask = 256	0
			Roberta = 768	124645632
			Output_layer = 5	3845
			T	124649477
3.	DistilBERT	4	Input_ids = 256	0
			Attention_mask = 256	0
			Distilbert = 768	66362880
			Output_layer = 5	3845
			T	66366725

Kolom parameter Tabel 4.6 menampilkan hasil total parameter/ukuran setiap masing-masing model :

1. Model BERT memiliki parameter/ukuran = 109486085.
2. Model RoBERTa memiliki parameter/ukuran = 124649477.
3. Model DistilBERT memiliki parameter/ukuran = 66366725.

Hasil dari parameter/ukuran model DistilBERT terbukti bahwa model DistilBERT melakukan pengurangan sebesar 40% (43119360) pada model BERT dan hasil parameter/ukuran model RoBERTa terbukti bahwa memiliki parameter/ukuran yang lebih besar pada model BERT sebanyak 15163392.

4.5 Hasil Waktu *Training* Setiap *Hyperparameter-tuning* dan *Fine-tuning* per Model

Tabel 4.7 menampilkan hasil waktu *training* dari arsitektur *hyperparameter-tuning* dan *fine-tuning* pada model BERT, RoBERTa, DistilBERT . Kolom no berisikan nomor model pada

tabel, kolom model berisikan nama model, dan kolom waktu *training* berisikan waktu pada model dalam melakukan *training*.

Tabel 4.7 Hasil Waktu *Training* Dari Arsitektur *Hyperparameter-tuning* dan *Fine-tuning* Pada Model Bert, Roberta, Distilbert

No	Model	Waktu <i>training</i>
1.	BERT	24 min 00 s
2.	RoBERTa	25 min 39 s
3.	DistilBERT	13 min 34 s

Kolom waktu *training* Tabel 4.7 menampilkan hasil waktu *training* setiap masing-masing model :

1. Model BERT memiliki waktu *training* = 24 min 00 s.
2. Model RoBERTa memiliki waktu *training* = 25 min 39 s.
3. Model DistilBERT memiliki waktu *training* = 13 min 34 s.

Hasil dari waktu *training* model DistilBERT terbukti bahwa model DistilBERT memiliki waktu *training* yang lebih cepat 60% dari model BERT dengan selisih 10 min 26 s dan model RoBERTa memiliki waktu *training* yang lebih lama dari model BERT dengan selisih 1 min 39 s, proses perhitungan waktu *training* menggunakan kode “%%time” pada google colab.

4.6 Hasil Akurasi Setiap *Hyperparameter-tuning* dan *Fine-tuning* per Model

Tabel 4.8 menampilkan hasil akurasi dari arsitektur *hyperparameter-tuning* dan *fine-tuning* pada model BERT, RoBERTa, DistilBERT. Kolom no berisikan nomor model pada tabel, kolom model berisikan nama model, dan kolom akurasi berisikan hasil akurasi pada model yang di cek dengan data *testing*.

Tabel 4.8 Hasil Akurasi Dari Arsitektur *Hyperparameter-tuning* dan *Fine-tuning* Pada Model Bert, Roberta, Distilbert

No	Model	Akurasi
1.	BERT	89,60
2.	RoBERTA	90,83
3.	DistilBERT	88,99

Kolom akurasi Tabel 4.8 menampilkan hasil akurasi setiap masing-masing model :

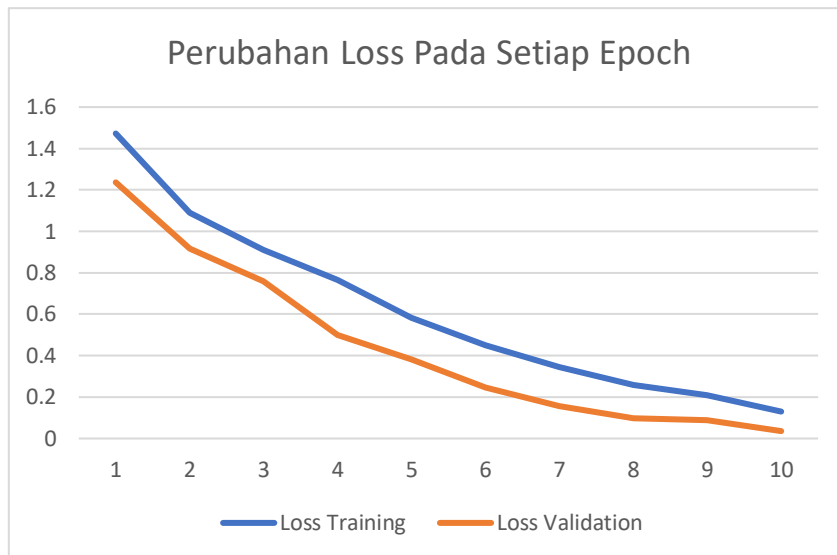
1. Model BERT memiliki akurasi = 89,60.
2. Model RoBERTa memiliki akurasi = 90,83.
3. Model DistilBERT memiliki akurasi = 88,99.

Hasil dari akurasi model DistilBERT = 88,99 terbukti bahwa model DistilBERT memiliki akurasi yang hampir sama (97%) dengan model BERT = 89,60 meskipun memiliki parameter/ukuran pengurangan sebesar 40%, waktu *training* yang lebih cepat 60% dan model RoBERTa memiliki akurasi = 90,83 terbukti bahwa memiliki akurasi yang lebih besar dari model BERT.

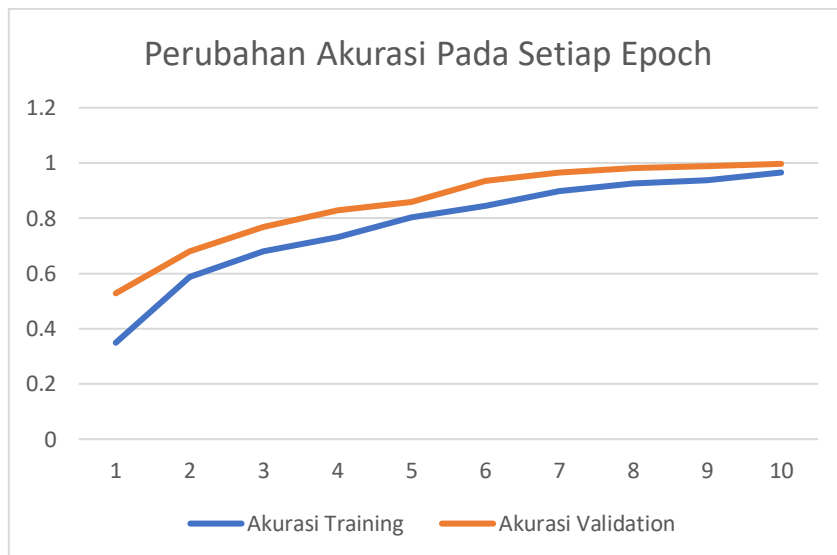
4.7 Hasil Analisa Peningkatan Performa Training Setiap Hyperparameter-tuning dan Fine-tuning per Model

Tabel 4.8 menampilkan hasil akurasi dari arsitektur *hyperparameter-tuning* dan *fine-tuning* pada model BERT, RoBERTa, DistilBERT. Model BERT, RoBERTa, DistilBERT akan dilakukan analisa peningkatan performa model ketika dilakukan proses *training* model. Analisa peningkatan performa model pada model BERT dengan *hyperparameter-tuning* *batch size* = 16, *dropout* = 0.1, *attention dropout* = 0.1, *epoch* = 10, dan *learning-rate* = 1e-5 dan *fine-tuning* yang ditetapkan melakukan peningkatan performa model dapat dilihat pada hasil *loss* setiap *epoch* saat *training* pada Gambar 4.2 dan hasil akurasi setiap *epoch* saat *training* pada Gambar 4.3.

Hasil dari Gambar 4.2 pada *line* berwarna biru = *loss training* dan *line* berwarna jingga = *loss validation* sudah mendapatkan nilai yang bagus karena *loss validation* lebih kecil dari *loss training* dan memiliki hasil *loss* yang hampir sepadan pada *loss training* dan *loss validation* dan Hasil dari Gambar 4.3 pada *line* berwarna biru = akurasi *training* dan *line* berwarna jingga = akurasi *validation* sudah mendapatkan nilai yang bagus karena akurasi *validation* lebih besar dari akurasi *training* dan memiliki hasil akurasi yang hampir sepadan pada akurasi *training* dan akurasi *validation*.



Gambar 4.2 Perubahan *loss* pada setiap *epoch* model bert

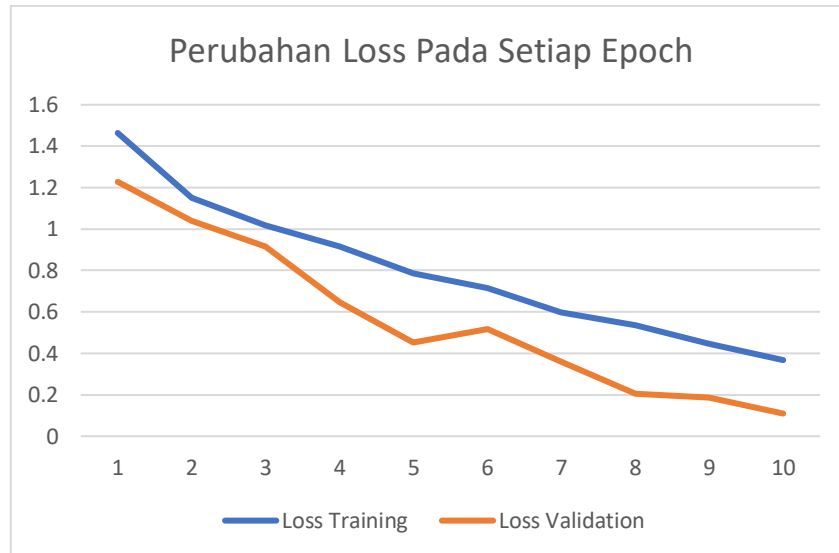


Gambar 4.3 Perubahan akurasi pada setiap *epoch* model bert

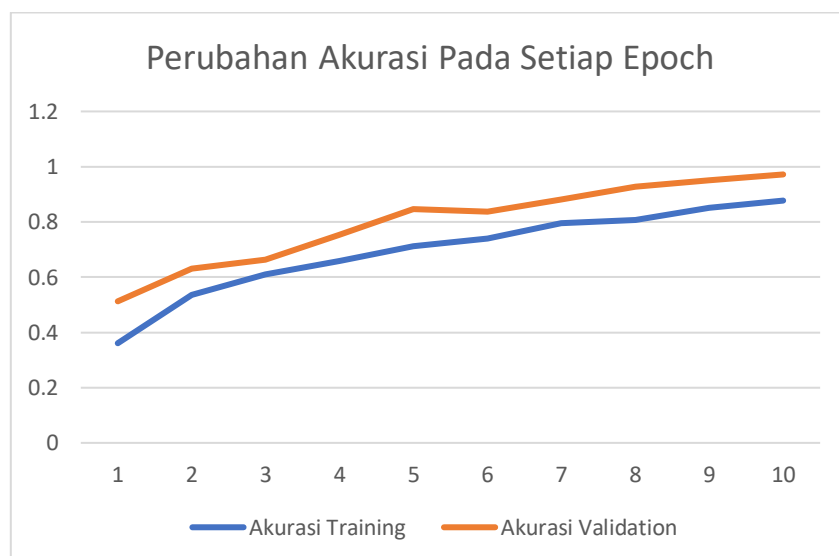
Analisa peningkatan performa model pada model RoBERTa dengan *hyperparameter-tuning* $batch\ size = 16$, $dropout = 0.1$, $attention\ dropout = 0.1$, $epoch = 10$, dan $learning-rate = 1e-5$ dan *fine-tuning* yang ditetapkan melakukan peningkatan performa model dapat dilihat pada hasil *loss* setiap *epoch* saat *training* pada Gambar 4.4 dan hasil akurasi setiap *epoch* saat *training* pada Gambar 4.5.

Hasil dari Gambar 4.4 pada *line* berwarna biru = *loss training* dan *line* berwarna jingga = *loss validation* sudah mendapatkan nilai yang bagus karena *loss validation* lebih kecil dari *loss training* dan memiliki hasil *loss* yang hampir sepadan pada *loss training* dan *loss validation* dan Hasil dari Gambar 4.5 pada *line* berwarna biru = akurasi *training* dan

line berwarna jingga = akurasi *validation* sudah mendapatkan nilai yang bagus karena akurasi *validation* lebih besar dari akurasi *training* dan memiliki hasil akurasi yang hampir sepadan pada akurasi *training* dan akurasi *validation*.



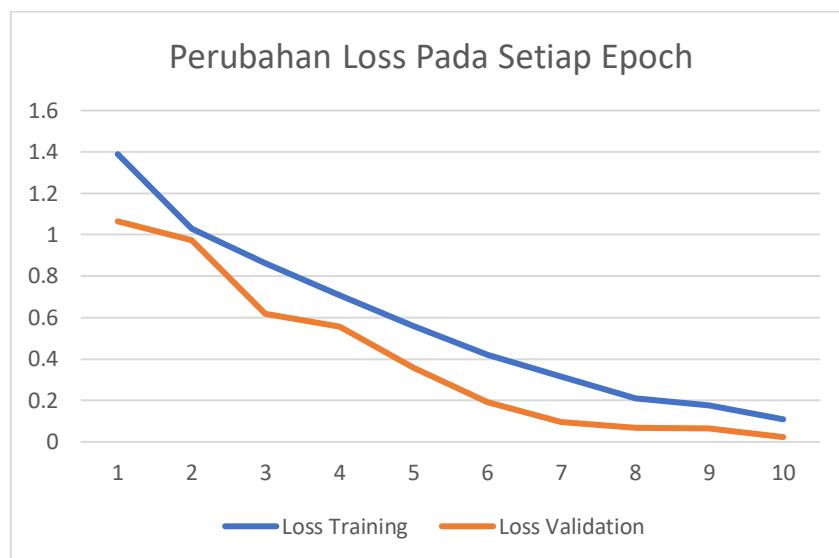
Gambar 4.4 Perubahan *loss* pada setiap *epoch* model Roberta



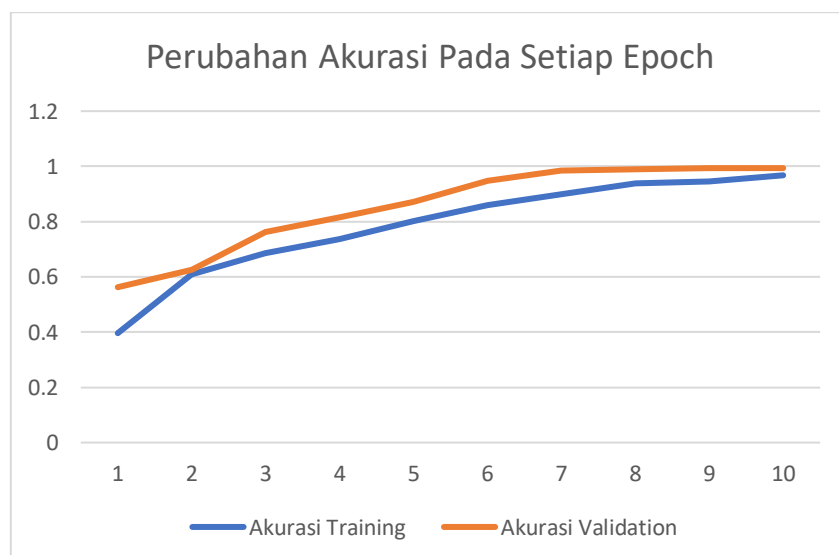
Gambar 4.5 Perubahan akurasi pada setiap *epoch* model Roberta

Analisa peningkatan performa model pada model DistilBERT dengan *hyperparameter-tuning* $batch\ size = 16$, $dropout = 0.1$, $attention\ dropout = 0.1$, $epoch = 10$, dan $learning-rate = 1e-5$ dan *fine-tuning* yang ditetapkan melakukan peningkatan performa model dapat dilihat pada hasil *loss* setiap *epoch* saat *training* pada Gambar 4.6 dan hasil akurasi setiap *epoch* saat *training* pada Gambar 4.7.

Hasil dari Gambar 4.6 pada *line* berwarna biru = *loss training* dan *line* berwarna jingga = *loss validation* sudah mendapatkan nilai yang bagus karena *loss validation* lebih kecil dari *loss training* dan memiliki hasil *loss* yang hampir sepadan pada *loss training* dan *loss validation* dan Hasil dari Gambar 4.7 pada *line* berwarna biru = akurasi *training* dan *line* berwarna jingga = akurasi *validation* sudah mendapatkan nilai yang bagus karena akurasi *validation* lebih besar dari akurasi *training* dan memiliki hasil akurasi yang hampir sepadan pada akurasi *training* dan akurasi *validation*.



Gambar 4.6 Perubahan *loss* pada setiap *epoch* model distilbert



Gambar 4.7 Perubahan akurasi pada setiap *epoch* model distilbert

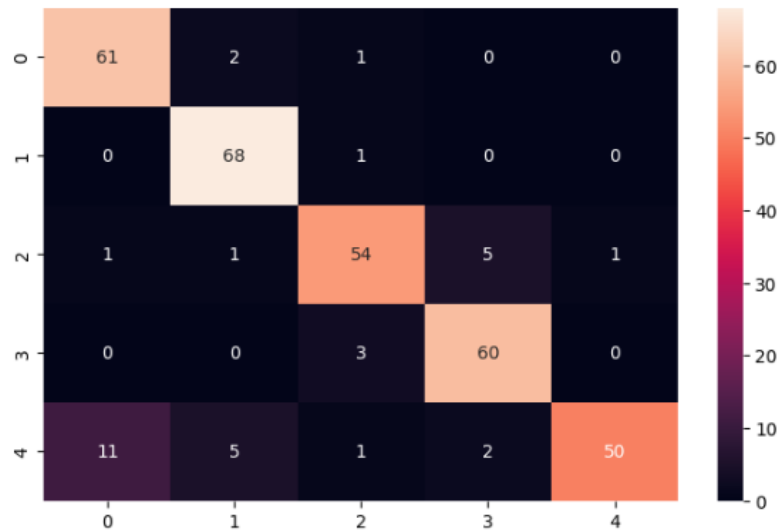
4.8 Hasil Prediksi Klasifikasi Emosi dan *F1-score* Setiap *Hyperparameter-tuning* dan *Fine-tuning* per Model

Tahapan dari pembangunan model yang telah dilakukan seperti : melihat ukuran parameter, waktu *training*, hasil akurasi, dan menganalisa berbagai model yang telah digunakan, selanjutnya akan dilakukan proses mengklasifikasi emosi dari data *testing* yang telah dibangun. Gambar 4.8 menampilkan hasil prediksi klasifikasi pada model BERT. Kode sebuah label dapat dilihat pada Tabel 4.3.

Kesalahan prediksi pada model BERT :

- Kesalahan pada prediksi label *anger* memiliki kesalahan pada label *fear* = 2, label *happy* = 1, label *love* = 0, dan label *sadness* = 0, total kesalahan prediksi pada label *anger* = 3.
- Kesalahan pada prediksi label *fear* memiliki kesalahan pada label *anger* = 0, label *happy* = 1, label *love* = 0, dan label *sadness* = 0, total kesalahan prediksi pada label *fear* = 1.
- Kesalahan pada prediksi label *happy* memiliki kesalahan pada label *anger* = 1, label *fear* = 1, label *love* = 5, dan label *sadness* = 1, total kesalahan prediksi pada label *happy* = 8.
- Kesalahan pada prediksi label *love* memiliki kesalahan pada label *anger* = 0, label *fear* = 0, label *happy* = 3, dan label *sadness* = 0, total kesalahan prediksi pada label *love* = 3.
- Kesalahan pada prediksi label *sadness* memiliki kesalahan pada label *anger* = 11, label *fear* = 5, label *happy* = 1, dan label *love* = 2, total kesalahan prediksi pada label *anger* = 19.
- Kesalahan prediksi terbesar ada pada label *sadness*, kesalahan prediksi terkecil ada pada label *fear*, dan total kesalahan pada seluruh label = 34. Kesalahan prediksi terbesar pada label *sadness* disebabkan dengan adanya berbagai macam kata yang dapat diungkapkan dalam rasa emosional kesedihan seperti : menangis, kecewa, sakit, sedih, dll, dan kesalahan prediksi terkecil pada label *fear* disebabkan dengan hanya beberapa kata yang dapat diungkapkan dalam rasa emosional ketakutan seperti : takut.

Tahap akhir penelitian ini melakukan perhitungan kualitas model menggunakan *confusion matrix* yang menghasilkan nilai *f1-score*, *f1-score* pada model BERT mendapatkan nilai 90. Hasil *f1-sccore* dapat dilihat pada Gambar 4.9.



Gambar 4.8 Hasil prediksi klasifikasi pada model bert

	precision	recall	f1-score	support
0	0.84	0.95	0.89	64
1	0.89	0.99	0.94	69
2	0.90	0.87	0.89	62
3	0.90	0.95	0.92	63
4	0.98	0.72	0.83	69
accuracy			0.90	327
macro avg	0.90	0.90	0.89	327
weighted avg	0.90	0.90	0.89	327

Gambar 4.9 Hasil *f1-score* model bert

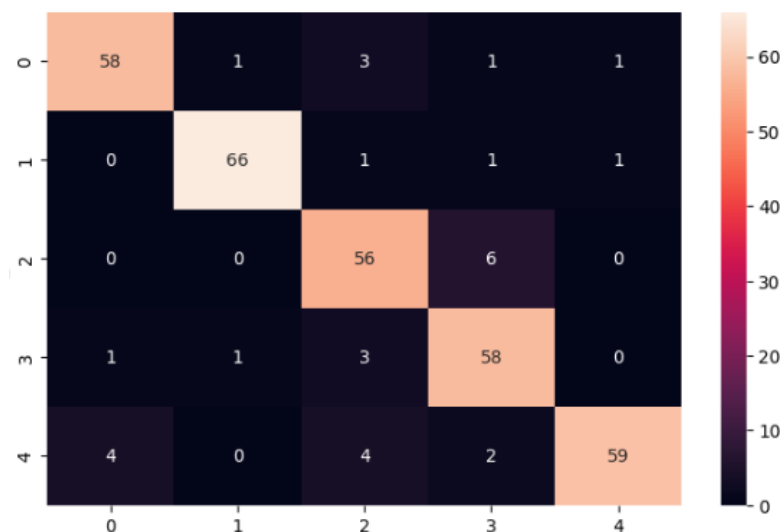
Gambar 4.10 menampilkan hasil prediksi klasifikasi pada model RoBERTa. Kode sebuah label dapat dilihat pada Tabel 4.3.

Kesalahan prediksi pada model RoBERTa :

- Kesalahan pada prediksi label *anger* memiliki kesalahan pada label *fear* = 1, label *happy* = 3, label *love* = 1, dan label *sadness* = 1, total kesalahan prediksi pada label *anger* = 6.
- Kesalahan pada prediksi label *fear* memiliki kesalahan pada label *anger* = 0, label *happy* = 1, label *love* = 1, dan label *sadness* = 1, total kesalahan prediksi pada label *fear* = 3.
- Kesalahan pada prediksi label *happy* memiliki kesalahan pada label *anger* = 0, label *fear* = 0, label *love* = 6, dan label *sadness* = 0, total kesalahan prediksi pada label *happy* = 6.
- Kesalahan pada prediksi label *love* memiliki kesalahan pada label *anger* = 1, label *fear* = 1, label *happy* = 3, dan label *sadness* = 0, total kesalahan prediksi pada label *love* = 5.

- Kesalahan pada prediksi label *sadness* memiliki kesalahan pada label *anger* = 4, label *fear* = 0, label *happy* = 4, dan label *love* = 2, total kesalahan prediksi pada label *anger* = 10.
- Kesalahan prediksi terbesar ada pada label *sadness*, kesalahan prediksi terkecil ada pada label *fear*, dan total kesalahan pada seluruh label = 30. Kesalahan prediksi terbesar pada label *sadness* disebabkan dengan adanya berbagai macam kata yang dapat diungkapkan dalam rasa emosional kesedihan seperti : menangis, kecewa, sakit, sedih, dll, dan kesalahan prediksi terkecil pada label *fear* disebabkan dengan hanya beberapa kata yang dapat diungkapkan dalam rasa emosional ketakutan seperti : takut.

Tahap akhir pada penelitian ini melakukan perhitungan kualitas model menggunakan *confusion matrix* yang menghasilkan nilai *f1-score*, *f1-score* pada model RoBERTa mendapatkan nilai 91. Hasil *f1-score* dapat dilihat pada Gambar 4.11.



Gambar 4.10 Hasil prediksi klasifikasi pada model roberta

	precision	recall	f1-score	support
0	0.92	0.91	0.91	64
1	0.97	0.96	0.96	69
2	0.84	0.90	0.87	62
3	0.85	0.92	0.89	63
4	0.97	0.86	0.91	69
accuracy			0.91	327
macro avg	0.91	0.91	0.91	327
weighted avg	0.91	0.91	0.91	327

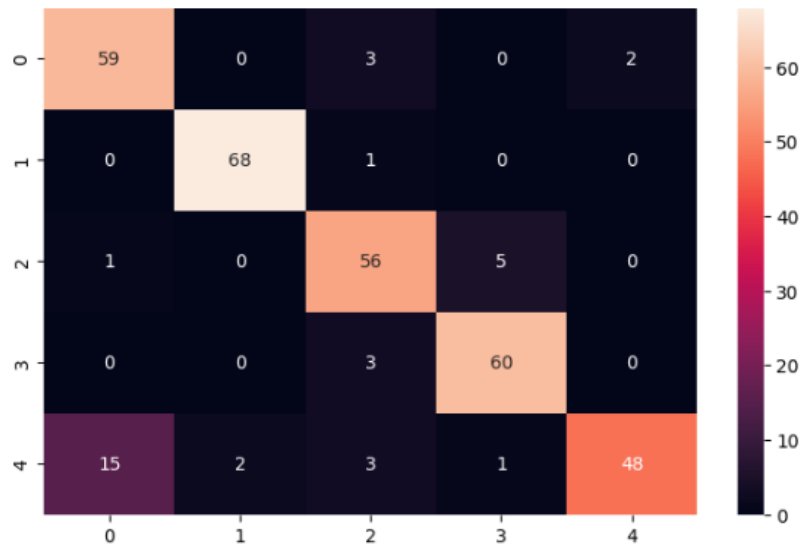
Gambar 4.11 Hasil *f1-score* model roberta

Gambar 4.12 menampilkan hasil prediksi klasifikasi pada model distilBERT. Kode sebuah label dapat dilihat pada Tabel 4.3.

Kesalahan prediksi pada model distilBERT :

- Kesalahan pada prediksi label *anger* memiliki kesalahan pada label *fear* = 0, label *happy* = 3, label *love* = 0, dan label *sadness* = 2, total kesalahan prediksi pada label *anger* = 5.
- Kesalahan pada prediksi label *fear* memiliki kesalahan pada label *anger* = 0, label *happy* = 1, label *love* = 0, dan label *sadness* = 0, total kesalahan prediksi pada label *fear* = 1.
- Kesalahan pada prediksi label *happy* memiliki kesalahan pada label *anger* = 1, label *fear* = 0, label *love* = 5, dan label *sadness* = 0, total kesalahan prediksi pada label *happy* = 6.
- Kesalahan pada prediksi label *love* memiliki kesalahan pada label *anger* = 0, label *fear* = 0, label *happy* = 3, dan label *sadness* = 0, total kesalahan prediksi pada label *love* = 3.
- Kesalahan pada prediksi label *sadness* memiliki kesalahan pada label *anger* = 15, label *fear* = 2, label *happy* = 3, dan label *love* = 1, total kesalahan prediksi pada label *sadness* = 21.
- Kesalahan prediksi terbesar ada pada label *sadness*, kesalahan prediksi terkecil ada pada label *fear*, dan total kesalahan pada seluruh label = 36. Kesalahan prediksi terbesar pada label *sadness* disebabkan dengan adanya berbagai macam kata yang dapat diungkapkan dalam rasa emosional kesedihan seperti : menangis, kecewa, sakit, sedih, dll, dan kesalahan prediksi terkecil pada label *fear* disebabkan dengan hanya beberapa kata yang dapat diungkapkan dalam rasa emosional ketakutan seperti : takut.

Tahap akhir pada penelitian ini melakukan perhitungan kualitas model menggunakan *confusion matrix* yang menghasilkan nilai *f1-score*, *f1-score* pada model distilBERT mendapatkan nilai 89. Hasil *f1-score* dapat dilihat pada Gambar 4.13.



Gambar 4.12 Hasil prediksi klasifikasi pada model distilbert

	precision	recall	f1-score	support
0	0.79	0.92	0.85	64
1	0.97	0.99	0.98	69
2	0.85	0.90	0.88	62
3	0.91	0.95	0.93	63
4	0.96	0.70	0.81	69
accuracy			0.89	327
macro avg	0.90	0.89	0.89	327
weighted avg	0.90	0.89	0.89	327

Gambar 4.13 Hasil *f1-score* model distilbert

4.9 Hasil Dari Semua Aspek Setiap *Hyperparameter-tuning* dan *Fine-tuning* per Model

Hasil parameter, waktu *training*, akurasi, dan *f1-score* akan dibandingkan dari keseluruhan model yang telah dibangun pada penelitian ini agar mengetahui kualitas masing-masing model pada kasus klasifikasi emosi data *text* Bahasa Indonesia dengan menggunakan proses *fine-tuning* dan *hyperparameter-tuning*. Tabel 4.9 menampilkan hasil parameter, waktu *training*, akurasi, dan *f1-score* setiap *hyperparameter-tuning* dan *fine-tuning* per model. Kolom no berisikan nomor model pada tabel, kolom model berisikan nama model, kolom parameter berisikan tentang total parameter yang dimiliki setiap model, kolom waktu *training* berisikan waktu pada model dalam melakukan *training*, akurasi berisikan hasil akurasi pada model yang di cek dengan data *testing*, dan kolom *f1-score* berisikan hasil *f1-score* pada model.

Tabel 4.9 Hasil Parameter, Waktu *Training*, Akurasi, dan *F1-Score* Setiap *Hyperparameter-tuning* dan *Fine-tuning* per Model

No	Model	Parameter	Waktu <i>training</i>	Akurasi	<i>F1-score</i>
1.	BERT	109486085	24 min 00 s	89,60	90
2.	RoBERTa	124649477	25 min 39 s	90,83	91
3.	DistilBERT	66366725	13 min 34 s	88,99	89

Penelitian ini bertujuan dalam membangun model BERT, RoBERTa, dan DistilBERT untuk mengetahui kualitas masing-masing model pada kasus klasifikasi emosi data *text* Bahasa Indonesia dengan menggunakan proses *fine-tuning* dan *hyperparameter-tuning*. Hasil penelitian ini membuktikan kualitas masing-masing model pada kasus klasifikasi emosi data *text* Bahasa Indonesia dengan menggunakan proses *fine-tuning* dan *hyperparameter-tuning* pada model BERT, RoBERTa, dan DistilBERT bahwa :

- Hasil dari parameter/ukuran model DistilBERT terbukti melakukan pengurangan sebesar 40% (43119360) pada model BERT dan hasil parameter/ukuran model RoBERTa terbukti bahwa memiliki parameter/ukuran yang lebih besar pada model BERT sebanyak 15163392.
- Hasil dari waktu *training* model DistilBERT terbukti bahwa model DistilBERT memiliki waktu *training* yang lebih cepat 60% dari model BERT dengan selisih 10 min 26 s dan model RoBERTa memiliki kecepatan waktu *training* yang lebih lama dari model BERT dengan selisih 1 min 39 s.
- Hasil dari akurasi model DistilBERT = 88,99 terbukti bahwa model DistilBERT memiliki akurasi yang hampir sama (97%) dengan model BERT = 89,60 meskipun memiliki parameter/ukuran pengurangan sebesar 40%, kecepatan waktu *training* yang lebih cepat 60% dan model RoBERTa memiliki akurasi = 90,83 terbukti bahwa memiliki akurasi yang lebih besar dari model BERT.
- Kesalahan prediksi terbesar pada model BERT, RoBERTa, DistilBERT ada pada label *sadness* disebabkan dengan adanya berbagai macam kata yang dapat diungkapkan dalam rasa emosional kesedihan seperti : menangis, kecewa, sakit, sedih, dll, dan kesalahan prediksi terkecil ada pada label *fear* disebabkan dengan adanya beberapa kata saja yang dapat diungkapkan dalam rasa emosional ketakutan seperti : takut.

Penelitian ini juga dibandingkan hasil *f1-score* dengan penelitian sebelumnya. Tabel 4.10 menampilkan hasil *f1-score* dengan penelitian sebelumnya pada model BERT, RoBERTa, DistilBERT. Kolom judul dan penulis berisikan tentang judul pada penelitian dan penulis

yang melakukan penelitian, kolom metode berisikan metode yang digunakan pada penelitian, kolom data berisikan tentang data yang digunakan pada penelitian, dan kolom *f1-score* berisikan hasil *f1-score* yang diperoleh dari model pada penelitian.

Tabel 4.10 Perbandingan Hasil *F1-score* Dengan Penelitian Sebelumnya Pada Model BERT, RoBERTa, DistilBERT

Judul dan Penulis	Metode	Data	<i>F1-score</i>
Classification Performance Comparison of BERT and IndoBERT on Self-Report of COVID-19 Status On Social Media (Budiman et al., 2024)	BERT	<i>Twitter</i>	(82)
Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging (Christian et al., 2021)	BERT, RoBERTa	<i>Facebook</i>	(71,8, 72)
Multi-label text classification of Indonesian customer reviews using bidirectional encoder representations from transformers language model (Nissa & Yulianti, 2023)	BERT, DistilBERT	<i>Airy rooms</i> <i>hotel</i> <i>reviews</i>	(90,9, 87,8)
Emotion Classification of Indonesian Text Data Using the BERT, RoBERTA, and DistilBERT Algorithms (Penelitian ini)	BERT, RoBERTa, DistilBERT	Indo4B	(89, 90, 91)

BAB 5

Kesimpulan dan Saran

5.1 Kesimpulan

1. Model BERT, RoBERTa, dan DistilBERT dapat membantu dalam mengklasifikasi emosi pada data *text* Bahasa Indonesia dengan berbagai macam tahapan. Beberapa tahapan yang dilakukan yaitu : pengumpulan data, *preprocessing* data, *hyperparameter-tuning* dan *fine-tuning*, melihat total parameter, melihat waktu *training*, melihat hasil akurasi, menganalisa peningkatan performa model saat *training*, melihat hasil prediksi klasifikasi emosi per label dan nilai *convusion matrix* dengan *f1-score*, dan hasil dari semua aspek pada seluruh model. Model BERT, RoBERTa, dan DistilBERT dapat mengklasifikasi emosi dari 5 label seperti *anger*, *fear*, *happy*, *love*, *sadness*, 5 label ini diolah dengan model *single label* dan data tersebut diolah dengan cara *undersample*. Klasifikasi emosi dirancang dengan model BERT, RoBERTa, dan DistilBERT dengan melakukan *fine-tuning* dan *hyperparameter-tuning*. *Fine-tuning* : penambahan *output layer* dan *loss function* : *cross entropy loss*. *Output layer* berfungsi untuk mengetahui hasil prediksi klasifikasi dan penambahan *loss cross entropy* yang digunakan untuk menghitung perbedaan antara keluaran yang dibuat oleh model dan keluaran yang diharapkan dan *Hyperparameter-tuning* : *batch size* = 16, *dropout* = 0.1, *attention dropout* = 0.1, *epoch* = 10, *learning rate* = $1e-5$.
2. Penelitian ini menghasilkan model yang dapat menerapkan klasifikasi emosi dengan hasil yang baik. Skenario yang telah diterapkan pada penelitian ini menghasilkan model BERT mendapatkan nilai parameter = 109486085, waktu *training* = 24 min 00 s, akurasi = 89,60, dan nilai *f1-score* = 90. Model RoBERTa mendapatkan nilai parameter = 124649477, waktu *training* = 25 min 39 s, akurasi = 90,83, dan nilai *f1-score* = 91. Model DistilBERT mendapatkan nilai parameter = 66366725, waktu *training* = 13 min 34 s, akurasi = 88,99, dan nilai *f1-score* = 89.

5.2 Saran

Berdasarkan penelitian ini, peneliti berharap agar penelitian ini dapat dikembangkan lebih baik lagi dengan cara :

1. Menggunakan sumber data yang lebih banyak dan *preprocessing* data yang lebih lengkap terhadap kata pada Bahasa Indonesia, sehingga entitas pada data bertambah.

2. Menggunakan model yang lebih banyak dalam membandingkan setiap kualitas masing-masing model yang terpercaya pada data *text* Bahasa Indonesia.
3. Melakukan *pretraining* dari awal dan memodifikasi *distillation* pada DistilBERT. DistilBERT menggunakan model *adversial distillation* yang mempunyai tujuan untuk model siswa mengikuti model guru dengan baik, dan hanya menggunakan 1 guru. Ada berbagai macam model mengenai *distillation* seperti model *multi-teacher distillation*, pada *muti-teacher distillation* juga mempunyai tujuan yang sama seperti *adverdsial distillation* tetapi pada *multi-teacher distillation* menggunakan lebih dari 1 guru.

DAFTAR PUSTAKA

- Abas, A. R., Elhenawy, I., Zidan, M., & Othman, M. (2022). Bert-cnn: A deep learning model for detecting emotions from text. *Computers, Materials and Continua*, 71(2), 2943–2961. <https://doi.org/10.32604/cmc.2022.021671>
- Acheampong, F. A., Nunoo-Mensah, H., & Chen, W. (2021). Transformer Models for Text-based Emotion Detection: A Review of BERT-based Approaches. *Springer*. <https://doi.org/https://doi.org/10.1007/s10462-021-09958-2>
- Acheampong, F. A., Nunoo-Mensah, H., Chen, W., & Wenyu, C. (2020). *Comparative Analyses of BERT, RoBERTa, DistilBERT, and XLNet for Text-based Emotion Recognition*. <https://www.researchgate.net/publication/346443459>
- Aman, S., & Szpakowicz, S. (2007). Identifying Expressions of Emotion in Text. *LNAI*, 4629, 196–205. https://doi.org/10.1007/978-3-540-74628-7_27
- Avvaru, A., & Vobilisetty, S. (2020). *BERT at SemEval-2020 Task 8: Using BERT to analyse meme emotions*. Online. https://en.wikipedia.org/wiki/Internet_meme
- Budiman, I., Faisal, M. R., Faridhah, A., Farmadi, A., Mazdadi, M. I., Saragih, H., & Abadi, F. (2024). *Classification Performance Comparison of BERT and IndoBERT on Self-Report of COVID-19 Status on Social Media*.
- Christian, H., Suhartono, D., Chowanda, A., & Zamli, K. Z. (2021). Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging. *Journal of Big Data*, 8(1). <https://doi.org/10.1186/s40537-021-00459-1>
- Cortiz, D. (2020). *Exploring Transformers in Emotion Recognition: a comparison of BERT, DistillBERT, RoBERTa, XLNet and ELECTRA*.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *ArXiv*. <https://doi.org/http://arxiv.org/abs/1810.04805>
- Dong, M. (2018). Emotion Classification on Indonesian Twitter Dataset. *IEEE*. <https://doi.org/10.1109/IALP.2018.8629262>
- Fathan Hidayatullah, A., Anna Apong, R., Teck Ching Lai, D., Qazi, A., Tungku Link, J., Seri Begawan, B. B., & Darussalam, B. (n.d.). *Adult Content Detection on Indonesian Tweets by Fine-tuning Transformer-based Models*.

- González-Carvajal, S., & Garrido-Merchán, E. C. (2020). Comparing BERT against traditional machine learning text classification. *ArXiv*. <https://doi.org/https://doi.org/10.48550/arXiv.2005.13012>
- Hastomo, W., Bayangkari Karno, A. S., Kalbuana, N., Meiriki, A., & Sutarno. (2021). Characteristic Parameters of Epoch Deep Learning to Predict Covid-19 Data in Indonesia. *Journal of Physics: Conference Series*, 1933(1). <https://doi.org/10.1088/1742-6596/1933/1/012050>
- He, F., Liu, T., & Tao, D. (2019). *Control Batch Size and Learning Rate to Generalize Well: Theoretical and Empirical Evidence*.
- Hidayatullah, A. F., & Ma'Arif, M. R. (2017). Pre-processing Tasks in Indonesian Twitter Messages. *Journal of Physics: Conference Series*, 801(1). <https://doi.org/10.1088/1742-6596/801/1/012072>
- Hinton, G., Vinyals, O., & Dean, J. (2015). *Distilling the Knowledge in a Neural Network*. <http://arxiv.org/abs/1503.02531>
- Huan, H., Guo, Z., Cai, T., & He, Z. (2022). A text classification method based on a convolutional and bidirectional long short-term memory model. *Connection Science*, 34(1), 2108–2124. <https://doi.org/10.1080/09540091.2022.2098926>
- Huang, J., Li, Y.-F., Xie, M., Huang, J., Xie, M., & Li, Y. F. (2015). A Systematic Analysis of Data Preprocessing for Machine Learning- based Software Cost Estimation. *HAL*, 67. <https://doi.org/10.1016/j.infsof.2015.07.004>
- Kaur, H., Pannu, H. S., & Malhi, A. K. (2019). A systematic review on imbalanced data challenges in machine learning: Applications and solutions. *ACM Computing Surveys*, 52(4). <https://doi.org/10.1145/3343440>
- Lee, K., Palsetia, D., Narayanan, R., Patwary, M. M. A., Agrawal, A., & Choudhary, A. (2011). Twitter trending topic classification. *Proceedings - IEEE International Conference on Data Mining, ICDM*, 251–258. <https://doi.org/10.1109/ICDMW.2011.171>
- Li, W., & Xu, H. (2014). Text-based emotion classification using emotion cause extraction. *Expert Systems with Applications*, 41(4 PART 2), 1742–1749. <https://doi.org/10.1016/j.eswa.2013.08.073>
- Lin, C. H., & Nuha, U. (2023). Sentiment analysis of Indonesian datasets based on a hybrid deep-learning strategy. *Journal of Big Data*, 10(1). <https://doi.org/10.1186/s40537-023-00782-9>

- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *ArXiv*. <https://doi.org/https://doi.org/10.48550/arXiv.1907.11692>
- Markoulidakis, I., Rallis, I., Georgoulas, I., Kopsiaftis, G., Doulamis, A., & Doulamis, N. (2021). Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem. *Technologies*, 9(4), 81. <https://doi.org/10.3390/technologies9040081>
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113. <https://doi.org/10.1016/j.asej.2014.04.011>
- Nimmi, K., Janet, B., Selvan, A. K., & Sivakumaran, N. (2022). Pre-trained ensemble model for identification of emotion during COVID-19 based on emergency response support system dataset. *Applied Soft Computing*, 122. <https://doi.org/10.1016/j.asoc.2022.108842>
- Nissa, N. K., & Yulianti, E. (2023). Multi-label text classification of Indonesian customer reviews using bidirectional encoder representations from transformers language model. *International Journal of Electrical and Computer Engineering*, 13(5), 5641–5652. <https://doi.org/10.11591/ijece.v13i5.pp5641-5652>
- Putri, F. P. (2023). DistilBERT with Adam Optimizer Tuning for Text-based Emotion Detection. *International Journal of New Media Technology*, 10(1), 30.
- Rahmad, F., Suryanto, Y., & Ramli, K. (2020). Performance Comparison of Anti-Spam Technology Using Confusion Matrix Classification. *IOP Conference Series: Materials Science and Engineering*, 879(1). <https://doi.org/10.1088/1757-899X/879/1/012076>
- Rezapour, M. (2021). *Emotion Detection with Transformers: A Comparative Study*. https://huggingface.co/LawalAfeez/emotion_detection/tree/main,
- Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2020). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *ArXiv*. <https://doi.org/https://doi.org/10.48550/arXiv.1910.01108>
- Saputra, D. F., Pratama, I. F., & Nathaniel, A. R. (2024). Psychological Stress Detection Using a Transformer-Based Model. *ComTech: Computer, Mathematics and Engineering Applications*, 15(1), 65–71. <https://doi.org/10.21512/comtech.v15i1.11105>

- Shaver, P. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/0022-3514.52.6.1061>
- Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019). How to Fine-Tune BERT for Text Classification? *Springer*. https://doi.org/10.1007/978-3-030-32381-3_16
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need*. <http://arxiv.org/abs/1706.03762>
- Wiciaputra, Y. K., Young, J. C., & Rusli, A. (2021). Bilingual text classification in english and indonesian via transfer learning using XLM-RoBERTa. *International Journal of Advances in Soft Computing and Its Applications*, 13(3), 72–87. <https://doi.org/10.15849/ijasca.211128.06>
- Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. (2020). IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. *ArXiv*. <https://doi.org/10.48550/arXiv.2009.05387>
- Xu, J., Zhang, Y., & Miao, D. (2020). Three-way confusion matrix for classification: A measure driven view. *Information Sciences*, 507, 772–794. <https://doi.org/10.1016/j.ins.2019.06.064>
- Yanuar, M., & Shiramatsu, S. (2020). *Aspect Ectractonfor Tourist Spot Review in Indonesian Language Using BERT*.