PENERAPAN ALGORITMA *ROBUSTLY OPTIMIZED BERT PRETRAINING APPROACH* (ROBERTA) UNTUK ANALISIS EMOSI DALAM KONTEKS KESEHATAN MENTAL

LAPORAN TUGAS AKHIR

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

Disusun Oleh :

OCTABRIANA ANGGUN RACHMAWATI

32602100107

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNOLOGI INDUSTRI

UNIVERSITAS ISLAM SULTAN AGUNG

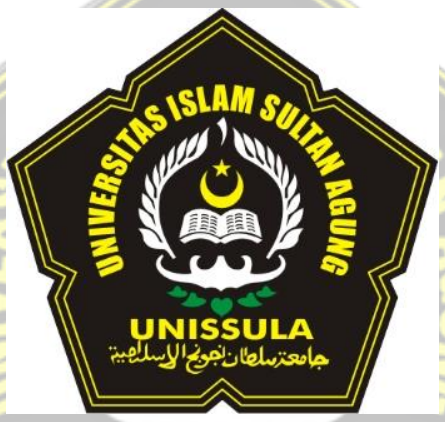
SEMARANG

2025

FINAL PROJECT

IMPLEMENTATION OF THE ROBUSTLY OPTIMIZED BERT PRETRAINING APPROACH (ROBERTA) FOR EMOTION ANALYSIS IN THE CONTEXT OF MENTAL HEALTH

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



Arranged By :

OCTABRIANA ANGGUN RACHMAWATI

32602100107

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNOLOGI INDUSTRI

UNIVERSITAS ISLAM SULTAN AGUNG

SEMARANG

2025

LEMBAR PENGESAHAN
TUGAS AKHIR

PENERAPAN ALGORITMA *ROBUSTLY OPTIMIZED BERT PRETRAINING APPROACH* (ROBERTA) UNTUK ANALISIS EMOSI DALAM KONTEKS KESEHATAN MENTAL

OCTABRIANA ANGGUN RACHMAWATI
NIM 32602100107

Telah dipertahankan di depan tim penguji ujian sarjana tugas akhir
Program Studi Teknik Informatika
Universitas Islam Sultan Agung
Pada tanggal : 18 Februari 2025

TIM PENGUJI SEMINAR PROPOSAL:

Moch. Taufik, ST., MIT NIK. 210604034 (Ketua Penguji)	.....	25 02 25
Bagus Satrio W.P, S.Kom., M.Cs NIK. 210616051 (Anggota Penguji)	.....	20-02-2025
Badie'ah, S.T,. M.Kom NIK. 210615044 (Pembimbing)	.....	20 - 02 - 2025

Semarang, 25 Febmari 2025

Mengetahui,

Kaprodi Teknik Informatika
Universitas Islam Sultan Agung

Moch. Taufik, ST., MIT
NIK. 210604034

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama	:	Octabariana Anggun Rachmawati
NIM	:	32602100107
Judul Tugas Akhir	:	PENERAPAN ALGORITMA *ROBUSTLY OPTIMIZED BERT PRETRAINING APPROACH* (ROBERTA) UNTUK ANALISIS EMOSI DALAM KONTEKS KESEHATAN MENTAL

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

Semarang, 06 Maret 2025

Yang Menyatakan,

Octabriana Anggun Rachmawati

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama	: Octabriana Anggun Rachmawati
NIM	: 32602100107
Program Studi	: Teknik Informatika
Fakultas	: Teknologi industri
Alamat Asal	: Pangkalan Bun, Kalimantan Tengah

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul PENERAPAN ALGORITMA *ROBUSTLY OPTIMIZED BERT PRETRAINING APPROACH* (ROBERTA) UNTUK ANALISIS EMOSI DALAM KONTEKS KESEHATAN MENTAL

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

Semarang, 06 Maret 2025

Yang menyatakan,

Octabriana Anggun Rahmawati

KATA PENGANTAR

Dengan mengucap syukur alhamdulillah atas kehadirat Allah SWT yang telah memberikan rahmat dan karunianya kepada penulis, sehingga dapat menyelesaikan Tugas Akhir dengan judul "PENERAPAN ALGORITMA *ROBUSTLY OPTIMIZED BERT PRETRAINING APPROACH* (ROBERTA) UNTUK ANALISIS EMOSI DALAM KONTEKS KESEHATAN MENTAL" ini untuk memenuhi salah satu syarat menyelesaikan studi serta dalam rangka memperoleh gelar sarjana (S-1) pada Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang. Tugas Akhir ini disusun dan dibuat dengan adanya bantuan dari berbagai pihak, materi maupun teknis, oleh karena itu saya selaku penulis mengucapkan terima kasih kepada:

1. Rektor UNISSULA Bapak Prof. Dr. H. Gunarto, S.H., M.H yang mengizinkan penulis menimba ilmu di kampus ini.
2. Dekan Fakultas Teknologi Industri Ibu Dr. Novi Marlyana, S.T., M.T., IPU., ASEAN.Eng
3. Dosen pembimbing penulis Ibu Badie'ah, S.T., M.Kom yang telah meluangkan waktu dan memberi ilmu.
4. Orang tua penulis yang telah memberi suport dan mengizinkan untuk menyelesaikan laporan ini.
5. Dan kepada semua pihak yang tidak dapat saya sebutkan satu persatu.

Dengan rendah hati, penulis menyadari bahwa laporan masih memiliki banyak kekurangan dalam hal kuantitas, kualitas, dan ilmu pengetahuan. Oleh karena itu, penulis mengharapkan kritikan dan saran yang membangun untuk membantu laporan ini menjadi lebih baik di masa depan.

Semarang, 18 Januari 2025

Octabriana Anggun Rachmawati

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

ABSTRAK

Kesehatan mental telah menjadi perhatian utama di seluruh dunia, terutama di tengah pandemi COVID-19. Penelitian ini bertujuan untuk menerapkan algoritma RoBERTa dalam analisis emosi, khususnya dalam konteks kesehatan mental. Data yang digunakan terdiri dari 37.685 baris data yang diambil dari platform Kaggle berupa teks komentar dan diproses melalui tahap *pre-processing*, pelabelan data, dan evaluasi model. Proses *pre-processing* meliputi pembersihan data dari simbol-simbol yang tidak diperlukan dan pengubahan huruf menjadi huruf kecil. Data kemudian diberi label berdasarkan kategori emosi seperti *happy, sad, surprise,* dan *neutral*. Model yang digunakan adalah RoBERTa dan DistilBERT, dengan hasil evaluasi menunjukkan bahwa model RoBERTa mencapai akurasi 98.75%, lebih tinggi dibandingkan dengan DistilBERT yang mencapai 98.65%. Sistem yang dibangun berupa sistem pendeteksi emosi. Penelitian ini tidak hanya berkontribusi pada pengembangan teknologi dalam analisis bahasa, tetapi juga memiliki dampak sosial yang signifikan dalam meningkatkan pemahaman dan penanganan masalah kesehatan mental di masyarakat. Diharapkan penelitian ini dapat menjadi dasar bagi pengembangan aplikasi lebih lanjut dalam bidang kesehatan.

Kata kunci : analisis emosi, kesehatan mental, BERT, RoBERTa, DistilBERT

Abstract

Mental health has become a major concern worldwide, especially amid the COVID-19 pandemic. This research aims to apply the RoBERTa algorithm in emotion analysis, particularly in the context of mental health. The data used consists of 37,685 rows commentar text obtained from the Kaggle platform and was processed through pre-processing, data labeling, and model evaluation stages. The pre-processing stage involved cleaning the data from unnecessary symbols and converting letters to lowercase. The data was then labeled based on emotion categories such as happy, sad, surprise, and neutral. The models used are RoBERTa and DistilBERT, with evaluation results showing that the RoBERTa model achieved an accuracy of 98.66%, higher than DistilBERT, which reached 98.17%. The developed system is a chatbot that can detect users' emotions and provide relevant suggestions. This research not only contributes to the development of technology in language analysis but also has a significant social impact in enhancing understanding and addressing mental health issues in society. It is hoped that this research can serve as a foundation for further application development in the field of mental health.

Keywords: *emotion analysis, mental health, BERT, RoBERTa, DistilBERT*

BAB I
PENDAHULUAN

1.1 Latar Belakang

Kesehatan mental merupakan salah satu aspek penting dalam kesehatan masyarakat global. Menurut Organisasi Kesehatan Dunia (*World Health Organization,* 2024) sejak tahun 2020, kesehatan mental telah menjadi perhatian utama, terutama di tengah pandemi COVID-19 yang telah mempengaruhi jutaan orang di seluruh dunia. Data menunjukkan bahwa prevalensi gangguan mental, seperti depresi dan kecemasan, meningkat secara signifikan selama periode ini.

WHO mencatat dalam laporan tahun 2021 bahwa kasus gangguan mental meningkat sebesar 25% di seluruh dunia selama tahun pertama pandemi. (*World Health Organization*, 2024). Hal ini menunjukkan bahwa situasi krisis dapat memperburuk kondisi kesehatan mental individu yang berdampak pada produktivitas dan kualitas hidup.

Di Indonesia, data dari Kementerian Kesehatan menunjukkan bahwa pada tahun 2022, sekitar 18% populasi mengalami masalah kesehatan mental, dengan angka yang lebih tinggi di kalangan remaja dan orang dewasa muda. Kemudian pada tahun 2023 (Kementrian Kesehatan Indonesia, 2023) terdapat gangguan kesehatan mental atau depresi menjadikan menjadikan masalah kejiwaan yang rentan terjadi pada remaja. Data di Indonesia 6,1 % warga Indonesia berusia 15 tahun ke atas mengalami gangguan kesehatan mental. Kondisi ini menunjukkan urgensi untuk memahami lebih dalam sentimen dan emosi masyarakat.

Dalam penelitian yang dilakukan (Sihombing *dkk,* 2024) Total sampel label sentimen yang diperoleh masing-masing berjumlah 9 untuk prediksi positif, 6 untuk prediksi negatif, dan 5 untuk netral. Nilai prediksi yang tinggi menunjukkan model klasifikasi sentimen reliabel dengan nilai rata-rata keseluruhan dari sampel prediksi adalah 97.70%. Hasil akurasi tersebut dipengaruhi oleh banyak data dan perbedaan label emosi yang dipakai. Peneliti menggunakan tiga label emosi diantaranya adalah positif, negativ dan netral. Penelitian ini mengarah pada klasifikasi emosi pada teks opini yang digunakan untuk mengetahui jenis emosi dari opini yang telah

diberikan Opini tersebut diambil dari media sosial Twitter. Metode yang digunakan dalam penelitian ini adalah RoBERTa.

Sebuah studi yang dilakukan (Vibriyanti dkk, 2024) menggunakan algoritma DistilBERT menghasilkan prediksi yang baik untuk mengklasifikasikan ulasan pada platform Duolingo. Hasil algoritma DistilBERT yang mengklasifikasikan 1000 data dari Google Play Store menunjukkan bahwa presisi, recall dan skor F1 untuk label kelas 0 masing-masing adalah 74%, 96% dan 84%.

Analisis emosi yang akurat dapat memberikan wawasan berharga tentang bagaimana perasaan seseorang berhubungan dengan kesehatan mental mereka. Dengan memahami emosi yang terekspresi dalam teks, kita dapat mengidentifikasi potensi masalah kesehatan mental lebih dini, memberikan dukungan yang lebih baik, dan merancang intervensi yang lebih efektif.

Dengan demikian, penelitian ini tidak hanya berkontribusi pada pengembangan teknologi dalam analisis bahasa, tetapi juga memiliki dampak sosial yang signifikan dalam meningkatkan pemahaman dan penanganan masalah kesehatan mental di masyarakat. Melalui penerapan algoritma BERT, diharapkan penelitian ini dapat menghasilkan model yang dapat diandalkan untuk analisis emosi, yang pada akhirnya dapat membantu dalam upaya pencegahan dan penanganan masalah kesehatan mental.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah di uraikan, rumusan masalah dalam penelitian ini dapat diidentifikasi sebagai berikut:

1. Penerapan Model RoBERTa dalam analisis emosi berbasis teks dalam konteks kesehatan.
2. Evaluasi tingkat akurasi model RoBERTa dalam mendeteksi emosi dan perbandingannya dengan model lain seperti DistilBERT.
3. Analisis efektifitas sistem deteksi emosi dalam memberikan respon yang sesuai kepada pengguna yang mengalami masalah kesehatan mental.

1.3 Batasan Masalah

Adapun Batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Fokus pada penerapan algoritma RoBERTa dalam konteks kesehatan mental.
2. Analisis akan terbatas menjadi empat kategori emosi seperti *happy*, *sad*, *surprice* dan *neutral*.
3. Bahasa yang digunakan hanya Bahasa Inggris.
4. Keluaran dari model bukanlah pengganti terapi klinis.

1.4 Tujuan

Tujuan dari penelitian ini adalah untuk menerapkan dan mengevaluasi kinerja algoritma RoBERTa dalam mendeteksi serta mengklasifikasikan berbagai jenis emosi yang muncul dalam konteks kesehatan mental, sehingga dapat memberikan pemahaman yang lebih mendalam mengenai pola ekspresi emosi dalam data teks yang dianalisis.

1.5 Manfaat

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut :

1. Menambah wawasan dan pengetahuan dalam bidang pemrosesan bahasa alami, khususnya dalam analisis emosi dalam konteks kesehatan mental.
2. Memberikan dasar bagi pengembangan aplikasi yang dapat membantu dalam memahami dan mengatasi masalah kesehatan mental melalui analisis emosi di lingkungan masyarakat.
3. Meningkatkan kesadaran tentang pentingnya perhatian terhadap kesehatan mental, serta bagaimana teknologi dapat berperan dalam mendukung kesejahteraan individu.

1.6 Sistematika Penulisan

Sistematika penulisan yang akan digunakan oleh penulis dalam pembuatan laporan tugas akhir adalah sebagai berikut :

BAB I : PENDAHULUAN

Pada Bab ini penulis mengutarakan latar belakang, rumusan masalah, tujuan, manfaat, ruang lingkup, dan sistematika penulisan.

BAB II : TINJAUAN PUSTAKA DAN DASAR TEORI

Pada Bab ini menguraikan teori-teori yang relevan dan penelitian sebelumnya yang berkaitan dengan penelitian ini.

BAB III : METODE PENELITIAN

Pada Bab ini menjelaskan metode yang digunakan dalam penelitian, termasuk pengumpulan data maupun perancangan sistem yang akan dilakukan.

BAB IV : HASIL DAN ANALISA

Pada bab ini berisi hasil dan pembahasan mengenai klasifikasi emosi pada teks menggunakan pendekatan RoBERTa dan DistilBERT

BAB V : KESIMPULAN DAN SARAN

Pada bab ini berisi kesimpulan dari seluruh bab dan saran penelitian selanjutnya.

BAB II
TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Analisis semantik digunakan untuk mengidentifikasi topik utama dan tema yang dibahas oleh buzzer di aplikasi X. Teknik yang umum digunakan dalam analisis semantik termasuk *Latent Dirichlet Allocation* (LDA), *word embeddings,* dan *Semantic Role Labeling*. LDA adalah metode yang popular untuk mengidentifikasi topik tersembunyi dalam teks dengan mengelompokkan kata-kata yang sering muncul bersama. *Word embedding*s, seperti *Word2Vec* dan *GloVe*, digunakan untuk merepresentasikan kata-kata dalam bentuk vektor yang menangkap makna kontekstualnya. (Amien, 2023)

Penelitian analisis emosi ini bertujuan untuk mengenali dan mengelompokkan perasaan yang terdapat dalam teks. Emosi merupakan bagian penting dari komunikasi manusia yang berdampak pada cara orang melihat dan merespons informasi.(Sutyo *dkk*, 2024)

Model BERT menerapkan dua arah pemodelan bahasa, tidak seperti model tradisional yang hanya menggunakan satu arah dari kiri ke kanan. Representasi kedua arah tersebut dapat diwujudkan melalui penggunaan *Masked Language Model* (MLM). MLM telah mengatur untuk menyamarkan token *input* secara acak sebesar 15% untuk keperluan prediksi. Ketika token ke-i dipilih, akan diubah menjadi token [MASK] sebesar 80%, sedangkan 10% lainnya akan digantikan dengan token acak. Sementara itu, 10% sisanya akan tetap tidak berubah. Token yang tidak diubah dimanfaatkan untuk mengatasi kelemahan pada tahap *fine-tuning*, karena token [MASK] hanya digunakan selama proses *pre-training*. Selanjutnya, digunakan model prediksi kalimat berikutnya (NSP) untuk memahami bagaimana kalimat A berhubungan dengan kalimat B. Separuh dari kalimat B dipilih sebagai label *Is Next* (kalimat berikutnya sebenarnya), sementara separuh lainnya diberi label *Not Next* (kalimat acak dari korpus). Di dalam loss function BERT, terdapat 3 jenis *loss function* yang digunakan: *Masked Language Modelling*

Loss (MLML), *Next Sentence Prediction Loss* (NSPL), serta cross entropy. Cross entropy berguna untuk menilai perbedaan antara 2 distribusi probabilitas pada sebuah rangkaian peristiwa atau variabel acak. Setelah perhitungan loss selesai, proses backpropagation dilakukan pada model RoBERTa. (Wamba *dkk*, 2021).

Pada penelitian hasil pengujian pada matriks evaluasi untuk Model-model yang telah diperbarui menunjukkan peningkatan total skor F1 antara 2,2 hingga 2,7 poin, serta peningkatan hingga 42,1 poin pada kategori soal yang paling menantang. Temuan dari penelitian ini mengungkapkan adanya variasi dalam cara RoBERTa, BERT, dan DistilBERT merepresentasikan komposisi dan informasi leksikal. Oleh karena itu, kemungkinan besar DistilBERT tetap mempertahankan sebagian besar informasi leksikal, seperti antonim, melalui proses penyulingan. Sementara itu, RoBERTa cenderung lebih banyak mempelajari fitur leksikal, termasuk antonim, dari jumlah data pelatihan yang lebih besar. (Ieva dkk, 2020)

Seperti halnya pada manusia, seringkali dapat terlihat kepribadian seseorang. Dengan jelas dapat disampaikan melalui kata-kata, entah dalam perkataan atau tulisan.Silakan tulis pesan singkat dan email. Hal tersebut juga berlaku untuk sistem AI. Model Bahasa Besar seperti Bard dan ChatGPT yang menyokong perkembangan terkini dalam bidang pemodelan bahasa. Tulisan satunya lagi perlu diubah dalam gaya penulisan yang lebih halus. Banyak orang telah mulai menyadari dari pengalaman yang mereka alami. Bercakap dengan AI, sistem ini memiliki sifat tertentu yang bersikap hati-hati. (Takwin, 2024)

2.2 Dasar Teori

2.2.1 Identifikasi emosi pada teks

Emosi adalah salah satu elemen yang memiliki dampak signifikan terhadap perilaku manusia terhadap lingkungan. Emosi terbagi menjadi dua kategori, yaitu emosi yang bersifat positif dan yang bersifat negatif. Untuk mengenali emosi ini, penting untuk memahami beberapa faktor yang memicunya. Adanya media sosial diera digital ini, membuat seseorang sering berkomunikasi di dunia maya. Identifikasi emosi dilakukan dengan mengetahui pemicu yang muncul dan mengetahui respon seseorang tersebut.(Putri dkk., 2023) Emosi sendiri dibagi

menjadi dua yaitu verbal dan non verbal, emosi Verbal bisa diartikan sebagai ekspresi emosi secara langsung dengan memanfaatkan lisan atau tulisan sebagai sarana. Sementara itu, non verbal umumnya memanfaatkan gerakan tubuh untuk menyampaikannya.

Sejauh ini, emosi dapat dideteksi melalui suara dan ekspresi wajah. Namun, emosi yang dituliskan sering kali menimbulkan arti ganda yang menjadi perdebatan. Selain itu, sulit untuk menyampaikan emosi dalam bentuk tulisan. Meskipun definisi struktur kata untuk emosi bisa dibagi menjadi lima kategori, yaitu marah (*anger*), takut (*fear*), senang (*happiness*), cinta (*love*), dan sedih (*sadness*). Namun, mengenali emosi di twitter kita tidak bisa hanya mengandalkan teknik pengolahan teks, karena karakter tweet yang singkat dan tata bahasa yang seringkali tidak teratur, yang membuat pemahaman terhadap emosi yang ingin didampaikan menjadi lebih sulit. Oleh karena itu, diperlukan teknik pengolahan teks yang lebih canggih, seperti penambahan teks, mengambil kata-kata yang tidak teratur. Dengan demikian, kita dapat membangun model pembelajaran mesin yang lebih efisien dalam mengklasifikasikan atau mengelompokkan emosi. (Sudianto, 2022)

2.2.2 Kesehatan Mental

Pada pertengahan abad ke-20, penelitian mengenai kesehatan mental mengalami perkembangan yang signifikan seiring dengan kemajuan ilmu pengetahuan dan teknologi modern. Kesehatan mental kini dipahami sebagai suatu ilmu yang diterapkan secara praktis dalam kehidupan sehari-hari manusia, melalui berbagai bentuk bimbingan dan konseling yang dilakukan di semua aspek kehidupan individu, termasuk di dalam keluarga, sekolah-sekolah, lembaga pendidikan, dan masyarakat.Pada awalnya, kesehatan mental hanya dipandang sebagai hal yang relevan bagi individu dengan gangguan kejiwaan, sehingga tidak diperuntukkan bagi setiap orang secara umum. (Diana, 2020). Namun, seiring berjalannya waktu, pandangan ini berubah kesehatan mental kini tidak hanya mencakup individu yang mengalami gangguan kejiwaan, tetapi juga mereka yang berada dalam kondisi mental yang sehat. Hal ini berfokus pada kemampuan

individu untuk mengeksplorasi dirinya sendiri dan cara mereka berinteraksi dengan lingkungan sekitar.

2.2.3 *Deep learning*

Deep Learning merupakan teknik pembelajaran mesin yang memanfaatkan jaringan saraf tiruan berlapis-lapis untuk mengekstraksi fitur yang rumit dan mewakili data *input* yang kompleks secara hierarkis. Yang membedakan *Deep Learning* dari pembelajaran mesin konvensional adalah jumlah dan kompleksitas lapisan neuron. Kemampuan *Deep Learning* dalam memperoleh pola dari data yang rumit dan abstrak telah membawa kesuksesan yang luar biasa dalam bidang seperti pengenalan citra, pemrosesan bahasa alami, pengenalan suara, dan banyak lagi. *Recurrent Neural Network* (RNN) adalah metode dalam algoritma *Deep Learning*. (Lubis *dkk*, 2024).

Artificial Neural Network digarap dengan rupa yang menyerupai otak manusianya. Di sini, neuron-neuron saling terhubung membentuk jaringan neuron yang amat rumit.(Nugroho *dkk,* 2020). *Deep Learning* merupakan bagian dari ilmu *Machine Learning* yang berfokus pada kecerdasan buatan. Metode *Deep Learning* termasuk dalam jaringan saraf yang terkenal dengan penggunaan struktur multilayer, sering digunakan karena kemampuannya untuk menyelesaikan berbagai masalah secara bersamaan dan memberikan solusi yang berbeda. *Deep Learning* mampu memahami data sehingga dapat merepresentasikan informasi tersebut dengan efektif, yang memungkinkan untuk melakukan prediksi dengan akurat.(Amigo, 2021)

2.2.4 Algoritma BERT

BERT merupakan model representasi yang berbasis fine-tuning pertama yang berhasil memperoleh hasil terbaik dalam tugas NLP. Data yang digunakan dalam model BERT adalah kalimat yang diubah menjadi vektor per kata dalam kalimat tersebut, yang sering disebut sebagai token. Pada model BERT, token memiliki representasi embedding yang unik.

BERT menggunakan WordPiece *embedding* yang mengandung 30.000 suku kata. Di awal dari setiap *sequence* terdapat token [CLS]. Untuk membedakan kalimat di setiap *sequence*, di setiap akhir kalimat ditempatkan token [SEP].

Kemudian ditambahkan *embedding* di setiap token untuk membedakan apakah token termasuk di kalimat yang mana yang disebut segment embedding. Hasil dari penjumlahan ini akhirnya akan ditambahkan dengan position embeddingyang memiliki jumlah dimensi sama dengan token *embedding*. *Position embedding* dihitung dengan rumus sebagai berikut.

$$\vec{p}2k = \sin(1100002kd\ /\cdot t) \quad (1)$$

$$\vec{p}2k + 1 = \cos(1100002kd\ /\cdot t) \quad (2)$$

Setiap *embedding* memiliki ukuran dimensi 768.

BERT berdasarkan pada transformer (Adelani dkk., 2021) yaitu mekanisme untuk memahami konteks kalimat dengan mempelajari keterkaitan kata-kata dalam teks. Transformer mampu mempelajari dan memodifikasi pemahaman yang diperoleh melalui mekanisme *self-attention*. Mekanisme *self-attention* mrupakan metode yang digunakan oleh transformator untuk mengubah kata-kata yang terhubung dengannya. Transformer terbentuk dari dua mekanisme, yaitu *Encoder* dan *Decoder*.

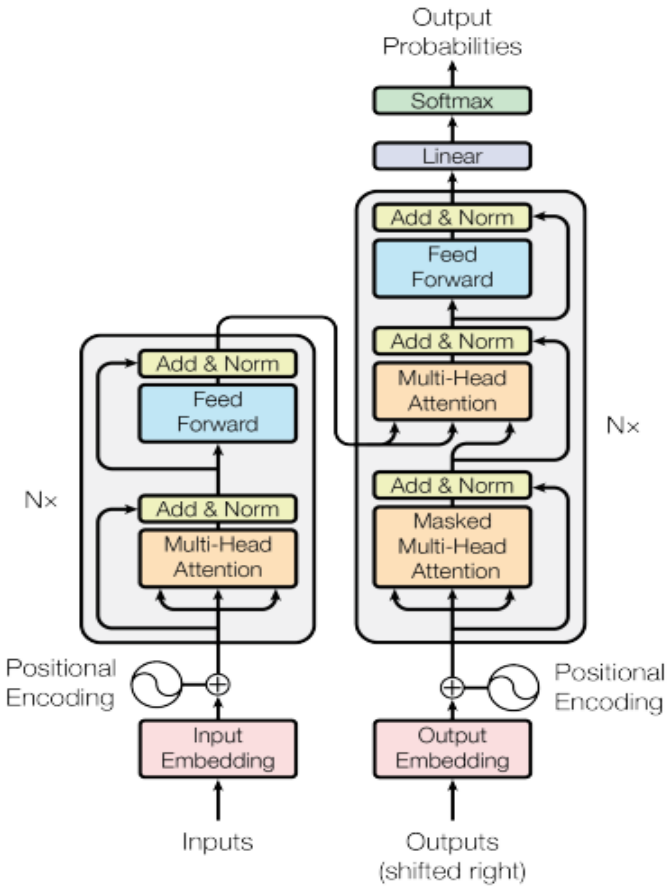
1. *Encoder*

   Encoder berfungsi untuk memproses teks yang diberikan. Saat ini, *Encoder* terdiri dari N = 6 lapisan yang serupa. Setiap lapisan memiliki dua bagian, yaitu sublapisan self-attention dan jaringan saraf yang maju. Melalui sublapisan *self-attention*, *Encoder* memungkinkan *node* untuk tidak hanya memperhatikan kata yang sedang dianalisis, tetapi juga mengerti konteks yang berkaitan dengan kata tersebut. Setiap posisi dalam *Encoder* dapat menganalisis semua posisi dari lapisan-lapisan sebelumnya di dalam *Encoder*.

2. *Decoder*

   Penggunakan *Decoder* adalah untuk menghasilkan urutan *Output* yang telah diprediksi. *Decoder* dilengkapi dengan tumpukan 6 lapisan yang bisa dikenali. Setiap lapisan terdiri dari dua sublapisan yang sama dengan lapisan *Encoder*. Dengan menambahkan lapisan perhatian di antara keduanya, hal ini akan membantu node saat ini dalam mengakses konten utama yang diinginkan dengan melakukan perhatian *multi-head* pada hasil keluaran *Encoder*. Di *Decoder*, sama seperti di *Encoder*, lapisan self-attention dapat membantu

setiap posisi dalam *Decoder* memproses semua posisi sebelumnya serta saat ini.



Gambar 2.1 Arsitektur transformer

(Vaswani, 2017)

Encoder (kiri) dan *Decoder* (kanan) terdiri dari tumpukan (stack) 6 layer yang identik.

1. *Encoder* dan *Decoder*: Ini adalah dua komponen utama dalam arsitektur Transformer. *Encoder* bertugas untuk memproses *input* (misalnya, kalimat) dan menghasilkan representasi numerik yang menangkap makna dari *input* tersebut. *Decoder* kemudian menggunakan representasi ini untuk menghasilkan *Output* (misalnya, terjemahan atau ringkasan).
2. Tumpukan (*stack*) 6 layer: Setiap *Encoder* dan *Decoder* terdiri dari 6 lapisan yang identik. Setiap lapisan ini melakukan pemrosesan tertentu pada data.

Setiap layer *Encoder* dan *Decoder* memiliki dua sub-layer:

1. *Self-attention layer*, Lapisan ini memungkinkan model untuk memperhatikan bagian-bagian penting dari *input*. Misalnya, ketika memproses kata "bank", model dapat memperhatikan kata-kata lain di sekitar "bank" untuk menentukan apakah yang dimaksud adalah "bank sungai" atau "bank tempat menyimpan uang".
2. *Feed-forward neural network* adalah jaringan saraf tiruan yang melakukan transformasi *non-linear* pada data.

Dengan tambahan attention layer pada *Decoder* di antara dua layers tersebut untuk membantu node mendapatkan *key content*. *Attention layer* pada *Decoder* Selain *self-attention*, *Decoder* juga memiliki lapisan *attention* yang memungkinkan model untuk fokus pada bagian-bagian relevan dari *Output Encoder* saat menghasilkan *Output*. yang membutuhkan attention dengan melakukan multi-head attention pada *Output* dari *Encoder*. Multi-head attention: Ini adalah mekanisme yang memungkinkan model untuk memperhatikan beberapa aspek dari *input* secara simultan.

Lapisan *self-attention*, serta *Encoder* dan *Decoder*, memfasilitasi *node* agar tidak hanya berfokus pada kata yang sedang dianalisis, tetapi juga untuk menangkap konteks semantik dari kata tersebut. Konteks semantik ini merujuk pada arti dari kata dalam konteks kalimat atau paragraf yang lebih besar. *Self-attention* mendukung model dalam memahami makna kata dengan mengevaluasi kata-kata di sekelilingnya. Setiap posisi pada *Encoder* mampu berinteraksi dengan semua posisi dari lapisan sebelumnya pada *Encoder* dan *Decoder*. Interaksi antar posisi ini memungkinkan setiap elemen dalam lapisan *Encoder* untuk terhubung dengan seluruh posisi dari lapisan sebelumnya, baik dalam *Encoder* maupun *Decoder*. Dengan demikian, model dapat mengidentifikasi hubungan yang rumit di antara kata-kata dalam sebuah kalimat.

2.2.5 Algoritma RoBERTa

RoBERTa adalah versi yang ditingkatkan dari model bahasa BERT. Model ini dikembangkan dengan melakukan perubahan pada pengaturan hyperparameter dan melatihnya menggunakan data dalam jumlah yang jauh lebih banyak. (Liu dkk, 2019) Penelitian menunjukkan bahwa meskipun BERT sudah menunjukkan

performa yang baik, masih ada ruang untuk perbaikan. Perbedaan utama antara RoBERTa dan BERT adalah sebagai berikut:

1. RoBERTa menghilangkan tugas NSP untuk meningkatkan efisiensi *pre-training.*
2. Menggunakan lebih banyak data dan waktu *pre-training* yang lebih lama.
3. Menerapkan dynamic masking dibandingkan static masking pada BERT. Dengan optimasi ini, RoBERTa menunjukkan peningkatan performa dibandingkan BERT dalam berbagai benchmark NLP.

Upaya peningkatan pada RoBERTa berhasil meningkatkan kinerjanya secara signifikan, mengindikasikan bahwa metode pelatihan sebelumnya yang digunakan pada BERT belum sepenuhnya optimal. Facebook AI kemudian meluncurkan RoBERTa dengan harapan dapat mencapai hasil yang lebih baik dibandingkan BERT, terutama dengan memanfaatkan dataset yang lebih besar.(Kurnia, 2023) Perbedaan antara BERT dan RoBERTa cukup mencolok. Penelitian tambahan juga mengonfirmasi bahwa RoBERTa memberikan hasil yang lebih unggul dibandingkan dengan BERT.

2.2.5 DistilBERT

DistilBERT adalah versi ringan dari BERT yang dikembangkan oleh *Hugging Face*. Model ini bertujuan untuk mengurangi kompleksitas komputasi tanpa mengorbankan terlalu banyak performa. Beberapa perbedaannya adalah:

1. Menggunakan teknik *distillation* untuk mentransfer pengetahuan dari BERT ke model yang lebih kecil.
2. Mengurangi jumlah lapisan dari 12 menjadi 6, sehingga lebih cepat dan lebih hemat sumber daya.
3. Tidak menggunakan NSP dalam proses pre-training. DistilBERT tetap mampu mencapai sekitar 97% performa BERT dengan ukuran yang jauh lebih kecil, sehingga cocok untuk implementasi dalam lingkungan dengan keterbatasan komputasi.

Distillation BERT atau DistilBERT dirancang untuk mengurangi ukuran dan meningkatkan kecepatan pelatihan representasi *encoder* dua arah dari algoritma transformer(BERT), dimana algoritma DistilBERT menggunakan

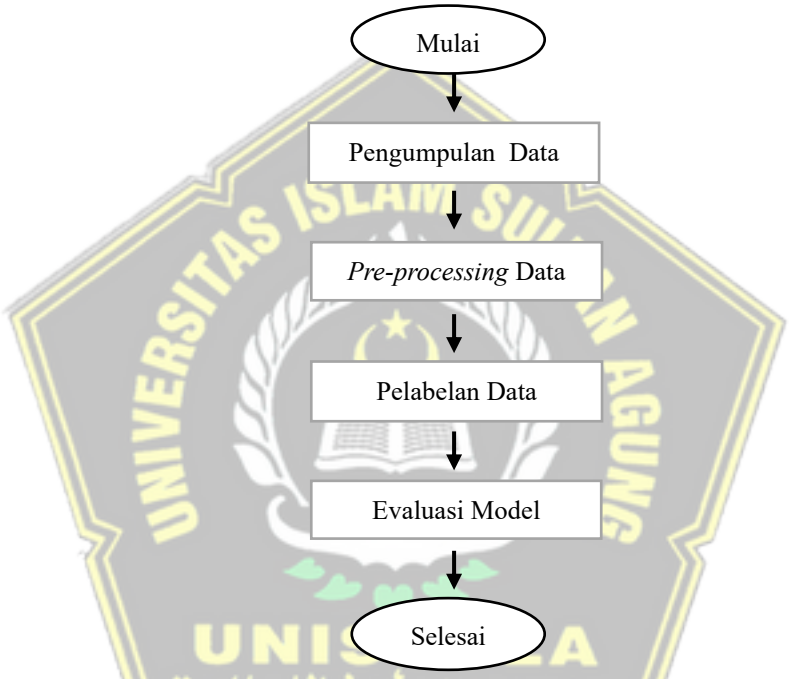
pengetahuan tentang distilasi (penyulingan) yang dapat mereduksi ukuran parameter BERT hingga 40%, serta mempercepat proses *inferensi* hingga 60%. (Ehsan et al, 2020).

Pada tahap tokenisasi, simbol atau token tertentu (CLS) dihasilkan untuk kebutuhan klasifikasi. Berikutnya, *input* ID dan *attention* mask dimasukkan sebagai input pertama ke dalam model. Ketika diproses dalam arsitektur DistilBERT, setiap kata menghasilkan vektor unik yang mencakup token khusus [CLS]. Hanya vektor yang dihasilkan dari token [CLS] yang digunakan sebagai *input* untuk klasifikasi. Vektor tersebut merepresentasikan keseluruhan teks.(Mauliddiyah dkk, 2024)

BAB III
METODE PENELITIAN

3.1 Tahapan penelitian

Alur penelitian dalam kajian ini ditampilkan pada ilustrasi 3.1. Penelitian ini mencakup serangkaian tahap, termasuk pengumpulan Data, *Pre-processing*, pelabelan data, dan evaluasi.



Gambar 3. 1 tahap penelitian

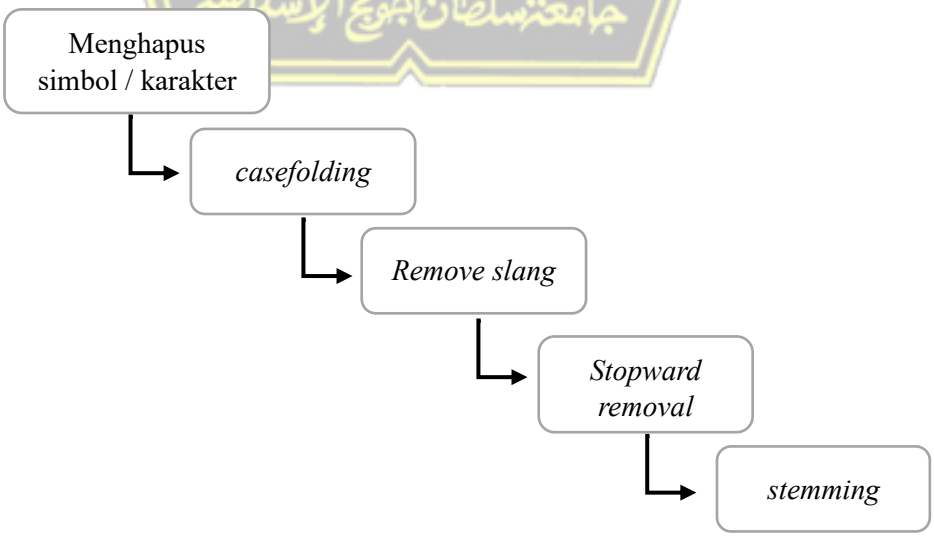
3.1.1 Pengumpulan data

Data yang digunakan dalam penelitian ini sebanyak 37.685 baris data yang diambil dari Kaggle. File ini di unduh dalam bentuk exel dengan format .CSV untuk mempermudah pengolahan data. Kaggle merupakan sebuah platform *online* yang menyediakan berbagai dataset resmi. Kaggle adalah tempat yang sangat berguna bagi peneliti, data *scientist*, dan pelajar untuk mengakses berbagai jenis data yang dapat digunakan untuk analisis dan pengembangan model pembelajaran mesin. Berikut merupakan jumlah baris data yang telah di ambil berdasarkan label.

3.1.2 *Pre-processing*

Pada tahap ini, informasi yang telah terkumpul akan diolah untuk mengekstrak pengetahuan yang terkandung di dalamnya, karena masih ada banyak data yang tidak teratur dan muncul simbol-simbol yang tidak diperlukan. Data akan diolah sebelumnya dengan cara menghapus informasi yang tidak relevan atau tidak sesuai, sehingga proses pengolahan data menjadi lebih mudah dan menghasilkan informasi yang sesuai dengan tujuan penelitian. Persiapan data dilakukan melalui berbagai metode yang berbeda. Selanjutnya, hal ini disesuaikan dengan kebutuhan dataset yang sedang digunakan. Setiap proses pencarian mungkin memerlukan teknik *pre-processing* yang berbeda.

Dalam tahap penelusuran ini, strategi yang diterapkan meliputi penghapusan simbol-simbol, perubahan semua huruf menjadi huruf kecil, pembersihan kata-kata slang untuk diganti dengan istilah standar, serta pemrosesan stemming untuk menyederhanakan variasi kata dalam dataset sehingga kata-kata yang memiliki akhiran atau awalan yang sama dapat diringkas. Selain itu, terdapat juga pengelompokan kata-kata lain yang memiliki fondasi dan makna yang sepadan, meskipun tampilannya berbeda akibat penggunaan imbuhan yang bervariasi. Proses ini sangat penting dilakukan secara sistematis agar data yang diperoleh dapat konsisten dan optimal. Rincian tahapan *pre-processing* yang dilakukan dalam penelitian ini bisa dilihat pada Gambar 3.2.



Gambar 3. 2 *pre-processing*

3.1.3 Proses pelabelan data

Dataset yang telah melalui tahap *pre-processing* dan memenuhi kriteria yang ditetapkan oleh peneliti kemudian diberi label berdasarkan kategori emosi yang sesuai. Dalam penelitian ini, terdapat tiga belas jenis emosi utama yang diidentifikasi, yaitu *Neutral, Love, Happiness, Sadness, Relief, Hate, Anger, Fun, Enthusiasm, Surprise, Empty, Worry,* dan *Boredom.* Namun, untuk mempermudah proses pengolahan data serta analisis yang lebih terfokus, tiga belas emosi tersebut dan untuk menghindari bias model akibat distribusi data yang tidak merata, label-label dengan makna serupa digabungkan yaitu :

1. *Happy* mencakup emosi positif seperti *joy, love, relief*.
2. *Sad* mencakup emosi negatif seperti *anger, fear, worry*.
3. *Surprise* berdiri sendiri karena memiliki karakteristik unik dalam ekspresi emosi.
4. *Neutral* mencakup teks yang tidak mengandung emosi spesifik atau netral.

Dengan melakukan pengelompokan ini, distribusi data menjadi lebih seimbang sehingga model dapat belajar dengan lebih baik tanpa dipengaruhi oleh ketimpangan jumlah data dalam setiap kategori. Selain itu, pendekatan ini juga memungkinkan analisis emosi yang lebih luas tanpa kehilangan esensi dari ekspresi emosi yang dianalisis dalam dataset.

Pengelompokan Berdasarkan Dimensi Emosi Utama Klasifikasi ini didasarkan pada model psikologi emosi dasar seperti yang dikemukakan oleh Paul Ekman (Kowalska dkk, 2017), yang mengidentifikasi emosi utama yang sering muncul dalam ekspresi manusia. Dari berbagai emosi yang ada, kategori *Happy, Sad, Surprise*, dan *Neutral* mencakup spektrum utama emosi yang sering ditemukan dalam analisis teks. Penggabungan Kategori yang Mirip Dataset awal memiliki banyak label emosi, tetapi beberapa kategori memiliki distribusi data yang tidak seimbang misalnya, beberapa emosi hanya memiliki sedikit data dibandingkan lainnya.

Agar data yang digunakan lebih seimbang di setiap kategori, dilakukan proses penyesuaian jumlah data per kategori. Hasilnya, setiap kategori label memiliki 4.000 baris data, sehingga total data yang digunakan menjadi 16.000 baris, dari

jumlah awal sebanyak 37.586 baris data. Langkah ini dilakukan untuk memastikan distribusi data yang merata pada setiap kategori, yang penting untuk mendukung akurasi analisis model yang digunakan dalam penelitian.

3.1.4 Evaluasi Model

Evaluasi adalah hasil akhir dari rangkaian aktivitas yang telah dilaksanakan atau hasil yang diperoleh setelah melalui proses yang panjang, sehingga metode yang dipilih adalah *Confusion matrix* untuk menilai tingkat akurasi dari proses klasifikasi. *Confusion matrix* mengandung nilai atau rumus yang diperlukan untuk menghitung tingkat keakuratan sistem tersebut. Oleh karena itu, penerapan *Confusion matrix* menjadi sangat krusial, sehingga sistem yang telah dirancang mempunyai kualitas yang baik dalam proses klasifikasi..

Confusion matrix adalah tabel yang menunjukkan jumlah prediksi yang benar dan salah untuk masing-masing kelas. Kolom menunjukkan prediksi model, sedangkan baris menunjukkan label asli (*ground truth*).

Tabel 3.1 struktur *confusion matrix*

	Pred: Anger	***Pred: Happy***	***Pred: Sad***	***Pred: Neutral***
True: Anger	*True Positives* (TP)	*False Negatives* (FN)	...	...
True: Happy	*False Positives* (FP)	*True Positives* (TP)	...	...

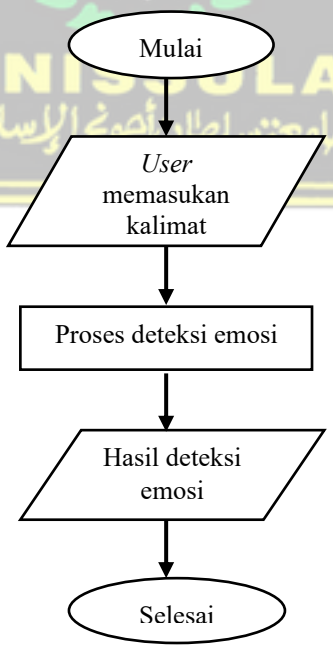
Setiap komponen dihitung sesuai dengan ketentuan prediksi dan label yang benar. Data matriks kebingungan mencakup empat nilai, yaitu :

1. *Accuracy,* matriks ini mengukur proposisi prediksi yang benar dibuat oleh model di seluruh kumpulan data. matriks ini dihitung sebagai rasio positif benar(TP) dan negatif benar(FP) terhadap jumlah sampel.
2. *Precision,* mengukur akurasi prediksi model pada suatu kelas tertentu, yaitu seberapa banyak prediksi positif benar dibandingkan dengan seluruh prediksi positif. *Precision* tinggi berarti ketika model memprediksi suatu kelas,

kemungkinan besar prediksi itu benar. Cocok digunakan jika kesalahan prediksi positif memiliki dampak besar.

3. *Recall* mengacu pada proporsi prediksi yang benar berbanding dengan seluruh hasil pendeteksian yang aktual. Rumus ini akan memberikan jumlah objek yang berhasil terdeteksi dengan akurat dari keseluruhan objek yang dikenali oleh sistem. *Recall*, yang sering disebut juga sebagai sensitivitas atau tingkat positif benar, adalah sebuah metrik yang menilai kemampuan model dalam mendeteksi semua contoh dari suatu kategori tertentu. *Recall* sangat krusial dalam keadaan di mana kita ingin menghindari kehilangan (*miss*) sampel positif dari kumpulan data.
4. F1-*Score*, sebagai tambahan dari accuracy, *precision* dan *recall* juga dihitung selama evaluasi untuk menilai kinerja model dengan menggabungkan nilai *precision* dan *recall*. Nilai 1 menunjukkan bahwa model yang diterapkan menunjukkan kinerja yang sangat baik dalam melakukan prediksi, sementara nilai 0 mengindikasikan bahwa model memiliki kinerja yang kurang memuaskan. Dalam klasifikasi biner, F1-*Score*.

3.2 Analisis Sistem



Gambar 3.3 *Alur Sistem*

Keterangan :

1. Mulai : Sistem atau proses dimulai.
2. User memasukkan *input* : Pengguna memberikan teks atau kalimat yang akan dianalisis.
3. Proses deteksi emosi: Model menganalisis *input* untuk mendeteksi emosi.
4. Sistem memberikan *Output* Hasil deteksi emosi.
5. Selesai.

3.3 Analisis Kebutuhan

Dalam tahap ini, peneliti menganalisis perangkat lunak apa saja yang diperlukan untuk membuat sistem deteksi ini bekerja dengan baik dan menghasilkan hasil yang diinginkan. Sistem ini dibuat menggunakan program-program berikut:

1. Python 3.11.0: Python adalah bahasa pemrograman tingkat tinggi dengan sintaks sederhana yang mudah dibaca. Python dapat digunakan untuk berbagai aplikasi dan pada berbagai platform. Python versi 3.11.5 dipilih untuk penelitian ini karena bersifat *open source*, memiliki komunitas yang besar, dan banyak sumber daya online.
2. *Library* Transformers: Transformers menyediakan akses ke model transformer yang kuat dan *Natural Language Processing* (NLP). Dalam penelitian ini, memungkinkan penggunaan model NLP tingkat lanjut seperti BERT untuk pemahaman bahasa yang lebih baik dan generasi jawaban yang lebih kontekstual.
3. *Library* PyTorch: PyTorch adalah perpustakaan *open-source* untuk komputasi tensor dan pembelajaran mesin, dan merupakan kerangka kerja pembelajaran mendalam yang populer. Dapat digunakan untuk membangun dan melatih model transformer dalam penelitian ini, dengan fitur dinamis dan dukungan GPU untuk meningkatkan efisiensi dan kinerja model.
4. *Library* Pandas: Pandas menyediakan struktur data seperti DataFrame untuk analisis data yang efisien. Dapat digunakan dalam penelitian ini untuk

mengelola dan menganalisis data seperti riwayat percakapan atau dataset yang digunakan untuk pelatihan model.

5. Library JSON: JSON (*Java Script Object Notation*) digunakan untuk menyimpan dan bertukar data dalam format yang mudah dibaca manusia dan mudah diproses oleh komputer. Dalam penelitian ini, JSON dapat digunakan untuk menyimpan dan mengelola konfigurasi, dataset, atau hasil percakapan.
6. *Visual Studio Code: Visual Studio Code* dipilih sebagai editor kode untuk pengembangan aplikasi dalam penelitian ini karena ramah pengguna, memiliki berbagai ekstensi, integrasi Git, terminal terintegrasi, dukungan untuk berbagai kerangka kerja dan bahasa pemrograman, dan pengenalan bahasa pemrograman cerdas untuk memudahkan proses pengembangan.
7. Google *Colaboratory*: Google Colab menyediakan lingkungan cloud untuk mengembangkan dan melatih model dengan akses GPU atau CPU, yang penting untuk model berat seperti Transformers. Digunakan dalam penelitian ini untuk menulis dan menjalankan kode, serta menyimpan dan mendokumentasikan.
8. Streamlit: Streamlit adalah kerangka kerja *open-source* yang memungkinkan Anda membuat antarmuka pengguna interaktif dalam Python. Dalam penelitian ini, Streamlit digunakan untuk dengan cepat mengubah skrip Python menjadi aplikasi web interaktif tanpa memerlukan pengetahuan mendalam tentang pengembangan web atau UI, menghemat waktu dalam prosesnya.

BAB IV
HASIL dan ANALISA

4.1 Hasil Analisa

4.1.1 Persiapan Data

Penelitian ini menggunakan dataset yang berisi 37.586 baris data yang berupa teks komentar tweeter, yang diperoleh dari Kaggle. Data tersebut diunduh dalam format Excel dengan ekstensi .CSV agar lebih mudah untuk diolah. File CSV ini kemudian diunggah ke dalam Google Colab untuk memulai proses analisis lebih lanjut. Langkah pertama yang dilakukan adalah memeriksa jumlah total data yang tersedia, guna memastikan bahwa dataset tersebut lengkap dan siap untuk dianalisis.

Tabel 4. 1 isi dari dataset

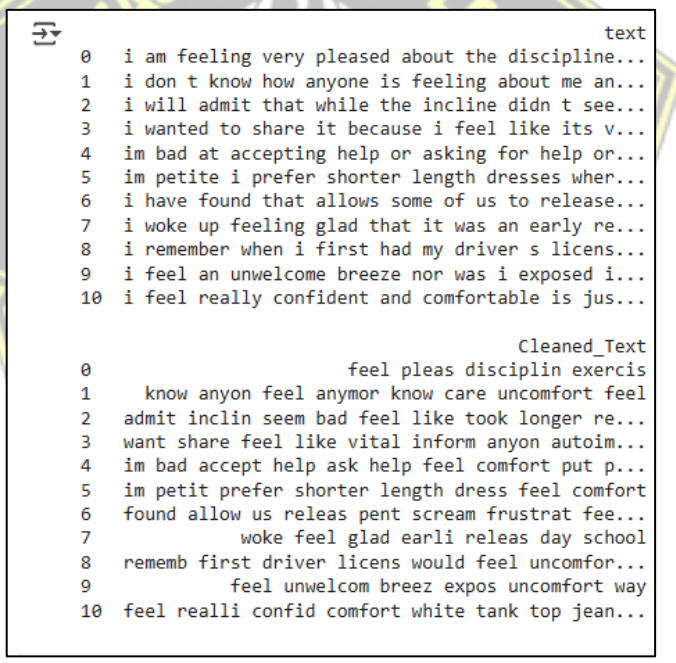
No	Text	Emotion
1	*I absolutely despise one subject, but now I feel hesitant to drop it*	*hate*
2	*i fall out of standing bow if i feel akward in nia or if my chain comes off on the trail*	*neutral*
3	*i am super exciting to be pursuing something that i feel really passionate about and i feel lucky to have been accepted into this program*	*neutral*
4	*i dunno how it feels to be completely happy the real world has taught me about struggle but what i m going thru is nothing close to struggle*	*happiness*
5	*i feel like she has taken on the role of a grandmother to me since my beloved grandma is no longer with me*	*love*
6	*i was feeling she determined that im feeling about of his movements possibly due to the way hes positioned*	*neutral*

No	Text	Emotion
7	*I could sit here and dwell on it all, but I'm feeling way too rebellious for that today. Basically, I'm angry at the world and myself all at once.*	*anger*
8	*i just couldnt help feeling a little bit bitter towards his great big happy grin*	*happiness*
9	*i would feel joyful*	*fun*
10	*I looked at what had changed for us in two generations and what hadn't changed for them in two or three, and instead of feeling outraged by their history of aggression, I felt privileged because of it.*	*anger*
11	*i feel and some is just a hateful of hollow yes i hear many smiths these days*	*empty*
12	*i feel morally outraged and furious more often than i d like*	*anger*
13	*i don t feel unimportant so much as i feel like one fast armored animal among many different kinds as entitled to this place as they are*	*neutral*
14	*i feel honoured a href http thepamperedsparrow*	*neutral*
15	*i feel so comfortable around him*	*relief*
16	*I also hope you understand why I feel so frustrated with you when you don't support the hat rule or when you show up to a school event without a hat yourself.*	*anger*
17	*i look at that god the god of abraham i feel im near a real god not the sort of dignified businesslike rotary club god we chatter about here on sunday mornings*	*neutral*
18	*i was going crazy thank god i have a craving for fruits and chocolate it made me go out in the cold with a gross wind blowing in my neck feeling mad and angry and crappy*	*anger*

No	Text	Emotion
19	*"I feel like I'm back in my element and very pleased to be surrounded by adorable tiny garments," I said without emotion, while a deep, unsettling anxiety swelled in my chest.*	*relief*
20	*"I feel like I'm back in my element and very pleased to be surrounded by adorable tiny garments," I said without emotion, while a deep, unsettling anxiety swelled in my chest.*	*worry*

4.1.2 *Pre-processing*

4.1.2.1 *Data Cleaning*



	text
0	i am feeling very pleased about the discipline...
1	i don t know how anyone is feeling about me an...
2	i will admit that while the incline didn t see...
3	i wanted to share it because i feel like its v...
4	im bad at accepting help or asking for help or...
5	im petite i prefer shorter length dresses wher...
6	i have found that allows some of us to release...
7	i woke up feeling glad that it was an early re...
8	i remember when i first had my driver s licens...
9	i feel an unwelcome breeze nor was i exposed i...
10	i feel really confident and comfortable is jus...

	Cleaned_Text
0	feel pleas disciplin exercis
1	know anyon feel anymor know care uncomfort feel
2	admit inclin seem bad feel like took longer re...
3	want share feel like vital inform anyon autoim...
4	im bad accept help ask help feel comfort put p...
5	im petit prefer shorter length dress feel comfort
6	found allow us releas pent scream frustrat fee...
7	woke feel glad earli releas day school
8	rememb first driver licens would feel uncomfor...
9	feel unwelcom breez expos uncomfort way
10	feel realli confid comfort white tank top jean...

Gambar 4. 1 hasil teks sebelum dan sesudah *cleaning*

Gambar 4.1 merupakan hasil dari *cleaning data* yang dilakukan pada kolom text dengan mengubah seluruh teks menjadi huruf kecil agar tidak ada perbedaan antara huruf kapital dan huruf kecil. Menghapus spasi ekstra dengan mengganti beberapa spasi berturut-turut menjadi satu spasi biasa dan kemudian menghapus spasi di awal dan akhir teks Selanjutnya menghapus *stopwords* dengan membagi teks menjadi kata-kata, dan kata-kata yang ada dalam daftar *stopwords* akan

dihapus. Setelah itu melakukan *stemming* pada kata-kata yang telah difilter, yaitu mengubah kata ke bentuk dasarnya menggunakan metode *steam* dari objek PorterStemmer, dan menggabungkan kembali kata-kata yang sudah diproses menjadi satu kalimat dengan menggunakan `join(stemmed_words)`. Setelah itu, menyimpan hasil *cleaning* terbaru dengan nama kolom `Cleaned_Text`.

4.1.2.2 Mapping data

Tabel 4. 2 Jumlah data awal

Emotion	***Count***	***Presentace***
Neutral	13536	35,63%
Relief	2728	7,18%
Love	2551	6,71%
Empety	2542	6,69%
Sadness	2481	6,53%
Worry	2474	6,51%
Enthusiasm	2304	6,06%
Happiness	2174	5.72%
Hate	2136	5,62%
Fun	2075	5,46%
Surprice	1532	4,03%
Anger	1335	3.51%
Boredom	126	0,33%

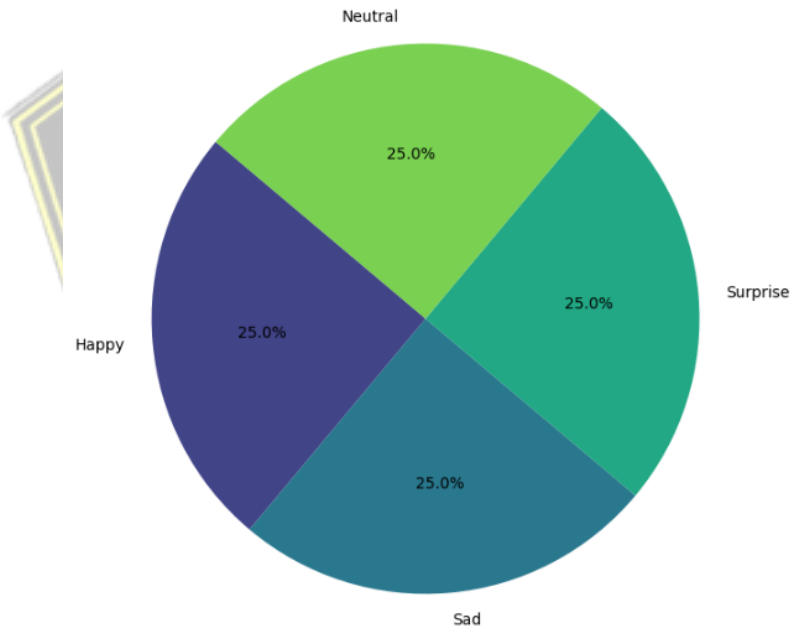
Tabel 4. 3 Jumlah kelas setelah di *mapping*

MappedEmotion	***Count***
Neutral	13536
Happy	11832
Sad	8552
Surprise	4074

Hasil *mapping* yang digunakan untuk memudahkan pemetaan emosi dengan memetakan tiga belas kategori seperti pada tabel 4.2 menjadi empat kategori pada

tabel 4.3, untuk *happy* gabungan dari kategori *fun, happiness, love, relief,* dan *enthusiasm*. Kategori *sad* berisi *sadness, anger, hate worry dan boredom*. Kategori *surprise* berisi *empty* dan *surprise*. Kemudian untuk *Neutral* tetap pada kategori *Neutral*. Kategori yang baru di gabungkan kemudian disimpan dengan nama kolom `Mapped_Emotion`. Proses ini dilakukan karena beberapa kategori memiliki jumlah sampel yang sangat kecil, sehingga sulit untuk dilatih atau menghasilkan model yang andal. Menggabungkan kategori serupa dapat membantu memperbaiki distribusi data. Ketika jumlah kategori terlalu banyak, model bisa kesulitan membedakan antara kategori yang mirip. Dengan mengelompokkan kategori yang mirip model dapat lebih fokus pada perbedaan utama.

4.1.2.3 Distribusi Data

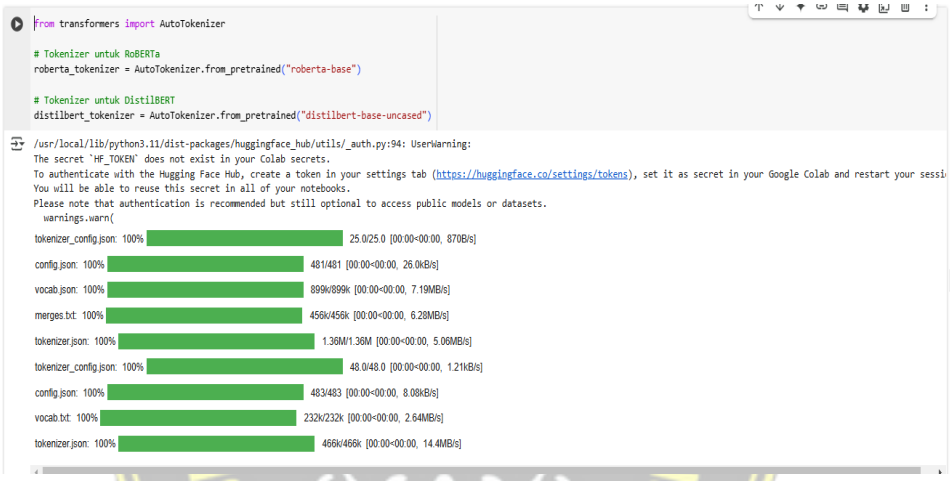


Gambar 4. 2 Distribusi data

Gambar 4.4 merupakan diagram yang telah dilakukan pendistribusian data. Pada kolom `Mapped_Emotion` data di bagi masing – masing kategori menjadi 4000 baris data, sehingga dapat memastikan model dapat membuat prediksi yang akurat dan tidak bias. Distribusi emosi ini penting untuk mengidentifikasi apakah ada ketidakseimbangan antar kelas emosi, yang bisa memengaruhi kualitas model yang akan dibangun. Proses ini juga mencakup pemeriksaan panjang setiap entri data, untuk mengetahui apakah ada teks yang terlalu panjang atau terlalu pendek. Data

yang terlalu panjang atau pendek mungkin perlu disesuaikan atau dibersihkan agar sesuai dengan standar yang dibutuhkan. Setelah itu, dilakukan identifikasi terhadap jumlah label emosi yang ada di dalam dataset, serta pencatatan frekuensi kemunculannya, untuk membantu dalam tahap pra-pemrosesan dan pemilihan teknik penyeimbangan data jika diperlukan. Semua informasi ini akan digunakan untuk merencanakan langkah-langkah selanjutnya dalam proses pembersihan data dan pelatihan model analisis emosi.

4.1.3 Tokenisasi Model



Gambar 4. 3 tokenisasi model

Gambar 4.3 ini menampilkan cuplikan kode Python yang menggunakan pustaka *Transformers* dari *Hugging Face* untuk memuat *tokenizer* dari dua model pemrosesan bahasa alami, yaitu RoBERTa dan DistilBERT. *Tokenizer* ini berfungsi untuk mengonversi teks menjadi token yang dapat dipahami oleh model. Pada kode tersebut, *tokenizer* RoBERTa dipanggil menggunakan `AutoTokenizer.from_pretrained ("RoBERTa-base")`, sementara *tokenizer* DistilBERT dipanggil dengan `AutoTokenizer.from_pretrained("distilbert-base-uncased")`. Token ini biasanya diperlukan untuk mengakses model atau dataset yang bersifat privat, tetapi tidak wajib untuk model publik seperti yang digunakan dalam kode ini. Setelah itu, log menunjukkan proses pengunduhan berbagai file konfigurasi dan *tokenizer*, termasuk config.json, vocab.json, merges.txt, serta tokenizer.json. Semua file berhasil diunduh dengan status 100%, yang menunjukkan bahwa *tokenizer* siap digunakan tanpa kendala.

4.2 Hasil Evaluasi

4.2.1 Model RoBERTa

Penelitian ini menggunakan *library Transformers* yang disediakan oleh *HuggingFace*, yang menyediakan ribuan model *pre-trained*. Setelah proses *pre-training*, langkah berikutnya adalah melakukan *fine-tuning*. Pada tahap *fine-tuning*, parameter yang digunakan meliputi jumlah *epoch* sebanyak 5, ukuran *batch* (*batch size*) sebesar 16, dan *learning rate* tertentu untuk mengoptimalkan performa model.

Tabel 4. 4 proses *fine tuning* RoBERTa

Epoch	*Training loss*	*Validation loss*	*Accurasion*
1	0.079900	0.185005	0.961625
2	0.067400	0.110659	0.978437
3	0.073800	0.092016	0.982187
4	0.042500	0.098623	0.984375
5	0.074700	0.058168	0.986563

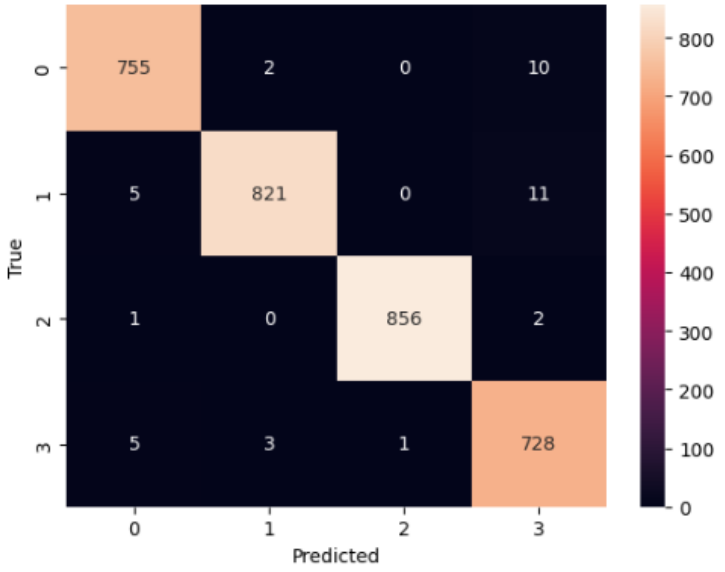
Dari hasil yang ditampilkan pada tabel, dapat dianalisis bahwa *Training loss* mengalami sedikit fluktuasi tetapi tetap relatif kecil (berkisar antara 0.042500 hingga 0.079900). *Validation loss* mengalami penurunan yang cukup stabil, dari 0.185005 pada *epoch* pertama menjadi 0.085168 pada *epoch* kelima. Hal ini menunjukkan bahwa model terus mengalami perbaikan dalam memahami pola dari data validasi. *Accuracy* meningkat dari 96.06% (*epoch* 1) menjadi 98.66% (*epoch* 5), yang menunjukkan, yang menunjukkan bahwa model memiliki keseimbangan yang baik antara mendeteksi kelas positif dan meminimalkan kesalahan klasifikasi.

Tabel 4. 5 hasil evaluasi model RoBERTa

	precision	recall	F1-score	support
Class 0	0.99	0.98	0.98	767
Class 1	0.99	0.98	0.99	873
Class 2	1.00	1.00	1.00	859
Class 3	0.97	0.99	0.98	737
Accuracy			0.99	3200
Macro avg	0.99	0.99	0.99	3200
Weight avg	0.99	0.99	0.99	3200
Accuracy RoBERTa : 0.9875				

Tabel 4.5 menunjukkan hasil evaluasi model RoBERTa setelah dilakukan fine-tuning untuk tugas klasifikasi emosi. Hasil evaluasi ini ditampilkan dalam bentuk matriks *precision, recall,* F1-*score*, dan *support* untuk masing-masing kelas. Berdasarkan tabel, model memiliki performa yang sangat baik dengan akurasi keseluruhan sebesar 98.75%. Untuk masing-masing kelas, *Class* 2 memiliki performa terbaik dengan *precision, recall,* dan F1-*score* sebesar 1.00, yang berarti model mampu mengklasifikasikan kelas ini dengan sempurna tanpa kesalahan. Sementara itu, *Class* 0, *Class* 1, dan *Class* 3 juga memiliki matriks yang sangat tinggi, dengan precision dan recall mendekati 1.

Selain itu, *macro average* dan *weighted average* untuk *precision, recall, serta* F1-*score* semuanya bernilai 0.99 yang menunjukkan bahwa model mempertahankan performa tinggi di seluruh kelas tanpa ada bias signifikan terhadap kelas tertentu. Dengan nilai ini, dapat disimpulkan bahwa model RoBERTa sangat efektif dalam melakukan klasifikasi emosi dengan tingkat kesalahan yang sangat rendah.



Gambar 4. 4 confusion matrix model RoBERTa

Gambar 4.4 merupakan *confusion matrix*, yang digunakan untuk mengevaluasi performa model RoBERTa dalam mengklasifikasikan emosi berdasarkan teks. *Confusion matrix* ini menunjukkan jumlah prediksi yang benar dan salah yang dibuat oleh model dibandingkan dengan label sebenarnya. Semakin banyak angka yang muncul di diagonal utama, semakin baik performa model karena itu menunjukkan bahwa prediksi yang dibuat sesuai dengan label yang benar.

Pada *confusion matrix* ini, model berhasil mengklasifikasikan sebagian besar data dengan benar. Misalnya, sebanyak 755 sampel yang sebenarnya termasuk dalam kelas 0 berhasil diprediksi dengan benar sebagai kelas 0, sementara terdapat sedikit kesalahan di mana beberapa sampel dari kelas 0 diklasifikasikan sebagai kelas lain, seperti kelas 1 dan kelas 3. Hal yang sama terjadi pada kelas lainnya, di mana sebagian besar prediksi berada pada diagonal utama, tetapi masih ada sejumlah kecil kesalahan klasifikasi.

4.2.2 Model DistilBERT

Tabel 4. 6 proses *fine tuning* DistilBERT

Epoch	*Training loss*	*Validation loss*	*Accurasion*
1	0.069700	0.131253	0.969375
2	0.071800	0.086152	0.891250

Epoch	*Training loss*	*Validation loss*	*Accurasion*
3	0.050300	0.092782	0.982500
4	0.031800	0.093724	0.981875
5	0.006100	0.080794	0.986563

Tabel ini menampilkan hasil *fine-tuning* model DistilBERT pada tugas klasifikasi, dengan matriks yang dievaluasi selama 5 *epoch*. Setiap *epoch* menunjukkan nilai *Training loss, validation loss, accuracy* yang memberikan gambaran bagaimana model belajar dan meningkatkan performanya seiring waktu.

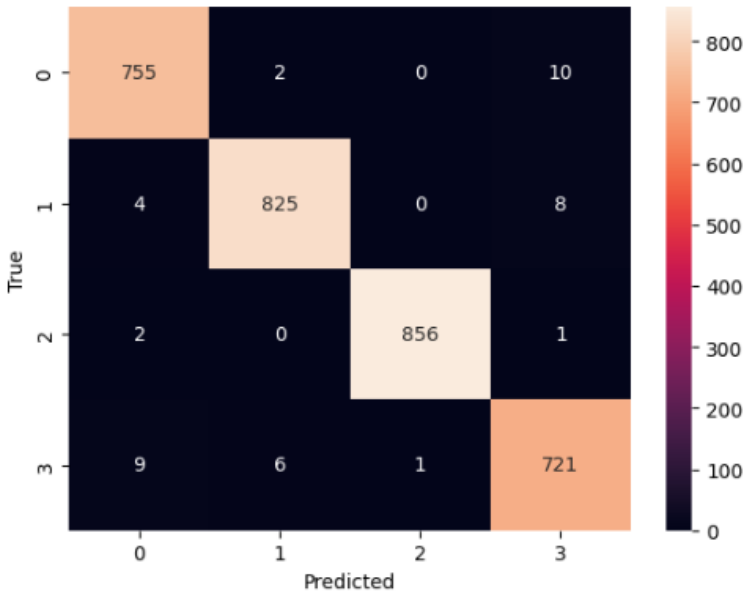
Pada *epoch* pertama, *Training loss* sebesar 0.0697 dan *Validation loss* sebesar 0.131253, menunjukkan bahwa model masih dalam tahap awal pembelajaran. Namun, akurasi sudah cukup tinggi, yaitu 96.94%, yang berarti model mampu mengklasifikasikan data dengan baik sejak awal. Seiring bertambahnya *epoch*, *Training loss* terus menurun, yang berarti model semakin memahami pola dalam data. *Validation loss* juga cenderung menurun, meskipun ada sedikit fluktuasi pada *epoch* ke-3 dan ke-4. Akurasi terus meningkat, dengan nilai tertinggi sebesar 98.65% pada *epoch* ke-5, yang menunjukkan bahwa model telah mencapai performa terbaiknya.

Tabel 4. 7 hasil evaluasi model DistilBERT

	precision	*recall*	F1-score	*support*
Class 0	0.98	0.98	0.98	767
Class 1	0.99	0.99	0.99	873
Class 2	1.00	1.00	1.00	859
Class 3	0.97	0.98	0.98	737
Accuracy			0.99	3200
Macro avg	0.99	0.99	0.99	3200
Weight avg	0.99	0.99	0.99	3200
Accuracy DistlisBERT : 0.9865625				

Tabel 4.7 menampilkan hasil evaluasi model DistilBERT setelah dilakukan *fine-tuning* untuk klasifikasi. Evaluasi dilakukan menggunakan matriks *precision, recall,* F1-*score*, dan *support* untuk masing-masing kelas. Dari hasil yang ditampilkan, model mencapai akurasi keseluruhan sebesar 98.65%, yang menunjukkan performa yang sangat baik dalam mengklasifikasikan data. *Class* 2 memiliki performa terbaik dengan *precision, recall*, dan F1-*score* sebesar 1.00, yang berarti model mampu mengenali kelas ini dengan sempurna tanpa kesalahan. Kelas lainnya, yaitu *Class* 0, *Class* 1, dan *Class* 3, juga memiliki nilai precision dan recall yang sangat tinggi, berkisar antara 0.97 hingga 0.99.

Selain itu, *macro average* dan *weighted average* untuk *precision, recall*, dan F1-*score* semuanya memiliki nilai 0.99, yang menunjukkan bahwa model dapat bekerja dengan baik di semua kelas secara seimbang tanpa adanya bias signifikan terhadap satu kelas tertentu. Dengan akurasi dan nilai matriks evaluasi yang tinggi, dapat disimpulkan bahwa model DistilBERT sangat efektif dalam menangani tugas klasifikasi ini, dengan tingkat kesalahan yang sangat rendah.



Gambar 4. 5 *confusion matrix* model DistilBERT

Gambar 4.5 merupakan *confusion matrix* yang menunjukkan hasil klasifikasi model DistilBERT terhadap empat kelas. Setiap angka dalam matriks menunjukkan

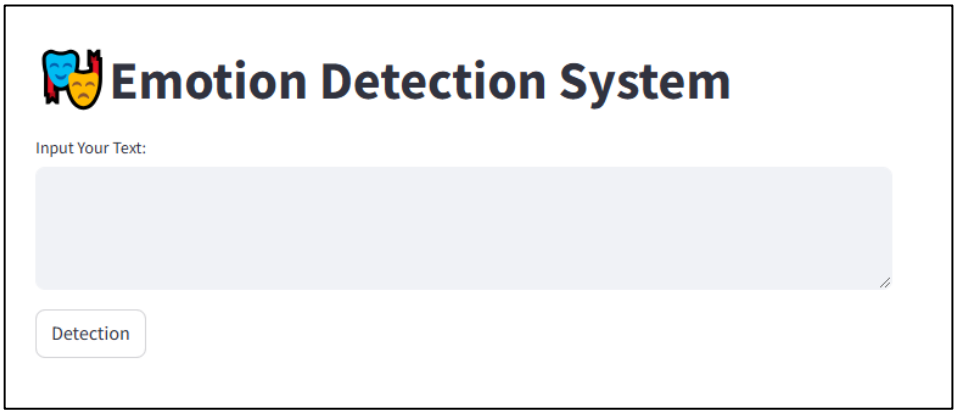
jumlah sampel yang diklasifikasikan ke dalam kelas tertentu oleh model. Dari matriks ini, terlihat bahwa sebagian besar prediksi berada di diagonal utama, yang berarti model mampu mengklasifikasikan data dengan sangat baik. Sebagai contoh, kelas 0 memiliki 755 sampel yang diklasifikasikan dengan benar, sedangkan hanya 12 sampel yang salah diklasifikasikan ke kelas lain (2 ke kelas 1 dan 10 ke kelas 3).

Kelas 1 juga memiliki performa yang baik, dengan 825 prediksi benar, sementara hanya 4 sampel salah diklasifikasikan ke kelas 0 dan 8 ke kelas 3. Demikian pula, kelas 2 memiliki 856 prediksi benar, dengan kesalahan yang sangat minim, hanya 2 sampel ke kelas 0 dan 1 ke kelas 3. Kelas 3 memiliki 721 prediksi benar, tetapi sedikit lebih banyak kesalahan dibandingkan kelas lainnya, dengan 9 sampel diklasifikasikan sebagai kelas 0, 6 sebagai kelas 1, dan 1 sebagai kelas 2.

4.3 Hasil Perancangan *User Interface*

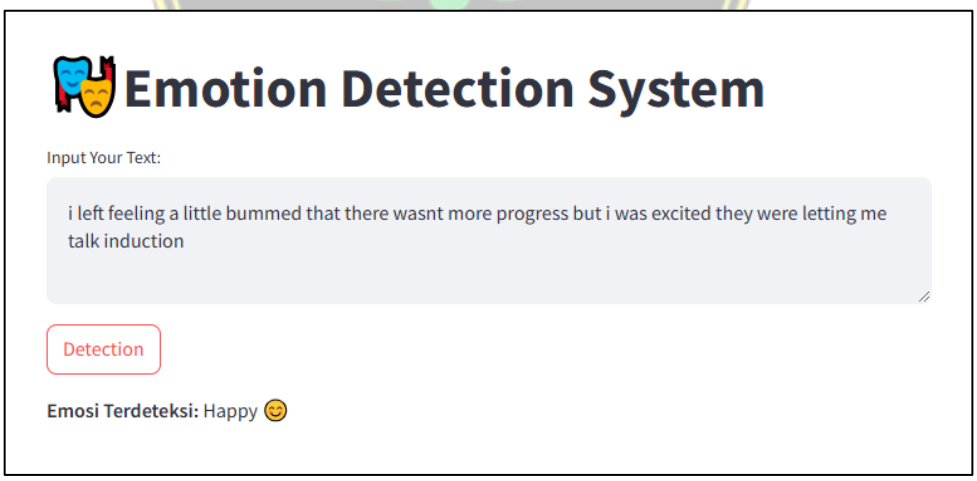
Tahap ini merupakan tahap implementasi pada platform website, di mana sistem yang telah dikembangkan berfungsi sebagai deteksi emosi. Sistem ini dirancang untuk menganalisis teks yang dimasukkan oleh pengguna dan mengidentifikasi emosi yang terkandung di dalamnya, seperti senang, sedih, marah, atau takut. Pada tahap ini, website telah selesai dibuat dan diuji untuk memastikan bahwa fitur deteksi emosi dapat berjalan dengan baik sesuai dengan yang diharapkan. Pengguna dapat memasukkan teks ke dalam sistem, kemudian model akan memprosesnya dan menampilkan hasil analisis emosi secara real-time.

Berikut ini adalah hasil dari demo website yang telah dikembangkan, menunjukkan bagaimana sistem mampu mengenali dan mengklasifikasikan berbagai emosi berdasarkan data yang diberikan.



Gambar 4. 6 tampilan halaman awal

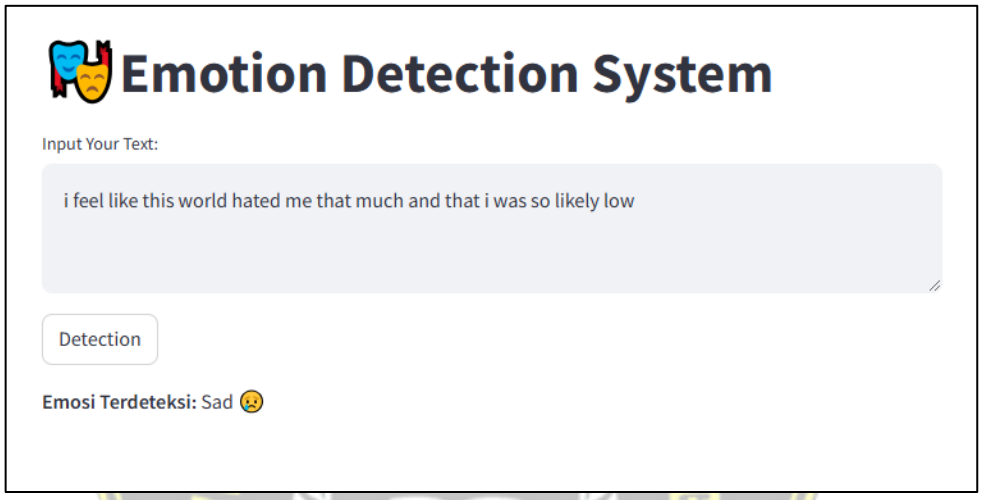
Gambar 4.6 menunjukkan antarmuka aplikasi berbasis web untuk analisis emosi (*Emotion Analysis*). Aplikasi ini dirancang untuk menerima masukan berupa teks dari pengguna melalui kolom *input* yang tersedia. Setelah teks dimasukkan, pengguna dapat menekan tombol "*Detections*" untuk memproses *input* tersebut. Tujuan utama aplikasi ini adalah untuk menganalisis teks yang dimasukkan dan mengidentifikasi emosi yang terkandung di dalamnya, seperti senang, sedih, marah, atau emosi lainnya. Dengan desain antarmuka yang sederhana dan intuitif, aplikasi ini memudahkan pengguna untuk memahami cara kerja dan menjalankan analisis emosi secara efisien.



Gambar 4. 7 hasil deteksi emosi *Happy*

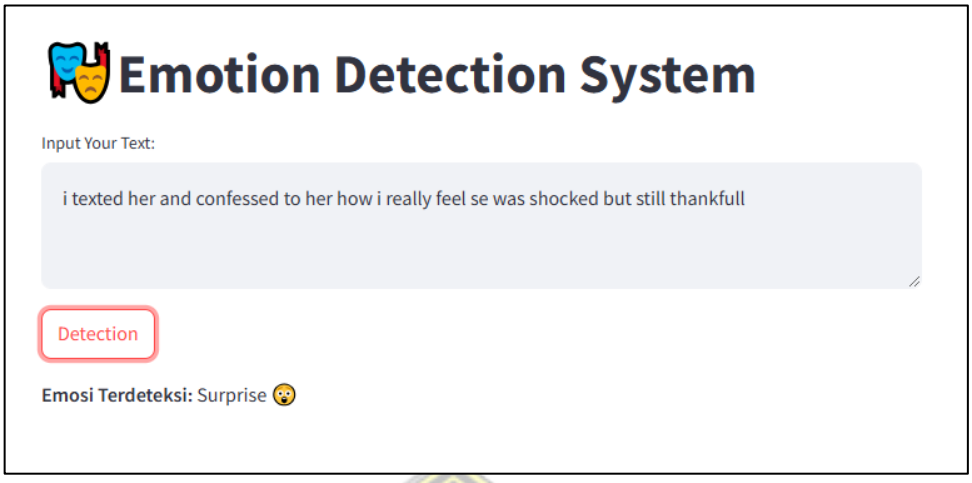
Gambar 4.7 merupakan hasil setelah pengguna memasukkan teks yang ingin dianalisis. Pada gambar, pengguna telah mengetikkan kalimat "*i left feeling a little bummed that there wasnt more progress but i was excited they were letting me talk*

induction" ke dalam kotak teks yang tersedia. Setelah itu, terdapat sebuah tombol bertuliskan "*Detections*", yang kemungkinan besar berfungsi untuk memproses teks dan mendeteksi emosi. Setelah tombol ditekan, sistem menampilkan hasil analisis dalam sebuah kotak hijau dengan teks "Emosi yang terdeteksi: *Happy*", yang menunjukkan bahwa sistem mengenali teks sebagai ekspresi kebahagiaan.



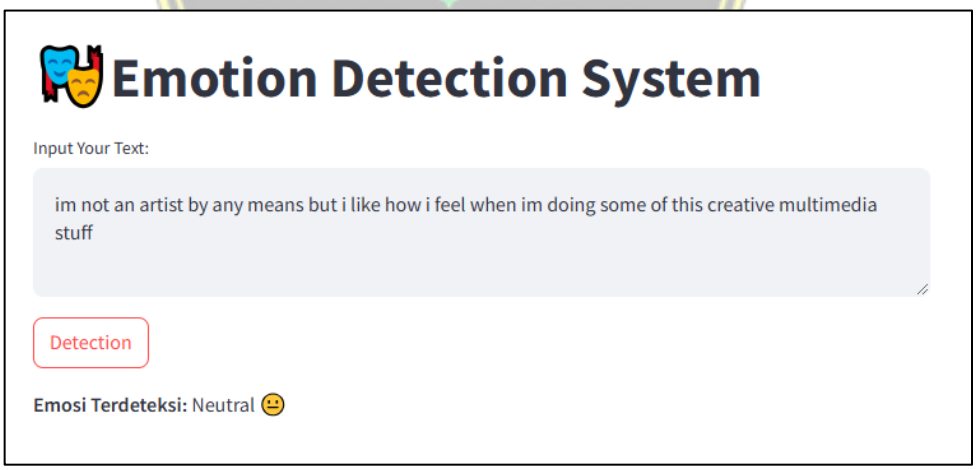
Gambar 4. 8 hasil deteksi emosi *Sad*

Gambar 4.8 merupakan hasil setelah pengguna memasukkan teks yang ingin dianalisis. Pada gambar, pengguna telah mengetikkan kalimat " *i feel like this world hated me that much and that i was so likely low*" ke dalam kotak teks yang tersedia. Setelah itu, terdapat sebuah tombol bertuliskan " *Detections* ", yang kemungkinan besar berfungsi untuk memproses teks dan mendeteksi emosi. Setelah tombol ditekan, sistem menampilkan hasil analisis dalam sebuah kotak hijau dengan teks "Emosi yang terdeteksi: *Sad*", yang menunjukkan bahwa sistem mengenali teks sebagai ekspresi kesedihan.



Gambar 4. 9 hasil deteksi emosi *surprise*

Gambar 4.9 merupakan hasil setelah pengguna memasukkan teks yang ingin dianalisis. Pada gambar, pengguna telah mengetikkan kalimat " *i texted her and confessed to her how i really feel se was shocked but still thankfull* " ke dalam kotak teks yang tersedia. Setelah itu, terdapat sebuah tombol bertuliskan "*Detections*", yang kemungkinan besar berfungsi untuk memproses teks dan mendeteksi emosi. Setelah tombol ditekan, sistem menampilkan hasil analisis dalam sebuah kotak hijau dengan teks "Emosi yang terdeteksi: *Surprise*", yang menunjukkan bahwa sistem mengenali teks sebagai ekspresi terkejut.



Gambar 4. 10 hasil deteksi emosi *neutral*

Gambar 4.10 merupakan hasil setelah pengguna memasukkan teks yang ingin dianalisis. Pada gambar, pengguna telah mengetikkan kalimat " *im not an artist by any means but i like how i feel when im doing some of this creative multimedia stuff*

" ke dalam kotak teks yang tersedia. Setelah itu, terdapat sebuah tombol bertuliskan "*Detections*", yang kemungkinan besar berfungsi untuk memproses teks dan mendeteksi emosi. Setelah tombol ditekan, sistem menampilkan hasil analisis dalam sebuah kotak hijau dengan teks "Emosi yang terdeteksi: *neutral*", yang menunjukkan bahwa sistem mengenali teks sebagai ekspresi netral.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

1. Penelitian ini berhasil menerapkan algoritma *Bidirectional Encoder Representations From Transformers* (RoBERTa) untuk analisis emosi dalam konteks kesehatan mental.
2. Hasil evaluasi menunjukkan bahwa model RoBERTa memiliki akurasi yang lebih tinggi dibandingkan dengan DistilBERT, dengan akurasi mencapai 98.66% pada pengujian.
3. Sistem yang dibangun berupa deteksi emosi yang diharapkan dapat membantu dalam memahami dan menangani masalah kesehatan mental.

5.2 Saran

1. Disarankan untuk melakukan penelitian lebih lanjut dengan menggunakan dataset yang lebih besar dan beragam untuk meningkatkan akurasi model.
2. Penelitian selanjutnya dapat mengeksplorasi penggunaan algoritma lain dalam analisis emosi untuk membandingkan performa dan efektivitasnya.
3. Diperlukan kolaborasi dengan profesional kesehatan mental untuk memastikan bahwa saran yang diberikan sesuai dan bermanfaat bagi pengguna yang mengalami masalah kesehatan mental.

DAFTAR PUSTAKA

Adelani, D. I., Abbott, J., Neubig, G., Daniel, D., Kreutzer, J., Lignos, C., Palen-michel, C., Buzaaba, H., Rijhwani, S., Ruder, S., Mayhew, S., dan Azime, I. A. (2021). *MasakhaNER : Named Entity Recognition for African Languages*. *9*, 1116–1131.

Amigo, J. M. (2021). *Data mining, machine learning, deep learning, chemometrics: Definitions, common points and trends* (Spoiler Alert: VALIDATE your models!). *Brazilian Journal of Analytical Chemistry*, *8*(32), 22–38. https://doi.org/10.30744/BRJAC.2179-3425.AR-38-2021

Diana, V. (2020). Kesehatan Mental (Sejarah Kesehatan Mental). In *Halodoc.Com* (Nomor March). https://www.researchgate.net/profile/Diana-Fakhriyani/publication/348819060_Kesehatan_Mental/links/60591b56458515e834643f66/Kesehatan-Mental.pdf

Ehsan, M., Nemati, S., Abdar, M., dan Asadi, S. (2020). *Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ' s public news and information . January*.

Ieva, S., dan Ignacio, I. (2020). Compositional and lexical semantics in RoBERTa, BERT and DistilBERT: A case study on CoQA. *EMNLP 2020 - 2020 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 7046–7056. https://doi.org/10.18653/v1/2020.emnlp-main.573

Kementrian Kesehatan Indonesia. (2023). *Menjaga Kesehatan Mental Para Penerus Bangsa*. Kemenkes. https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20231012/3644025/menjaga-kesehatan-mental-para-penerus-bangsa/

Kowalska, M., dan Wróbel, M. (2017). Basic Emotions. In *Encyclopedia of Personality and Individual Differences* (hal. 1–6). https://doi.org/10.1007/978-3-319-28099-8_495-1

Kurnia, P. Y. (2023). Analisis Sentimen Berita Vaksin COVID-19 Dengan Robustly Optimized BERT Pre-Training Approach (RoBERTa). *AT-TAWASSUTH: Jurnal Ekonomi Islam*, *VIII*(I), 1–19.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., dan Stoyanov, V. (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. *1*. http://arxiv.org/abs/1907.11692

Lubis, N., Siambaton, M. Z., dan Rachmat, A. (2024). Implementasi Algoritma Deep Learning pada Aplikasi Speech to Text Online dengan Metode Recurrent Neural Network (RNN). *sudo Jurnal Teknik Informatika*, *3*(3), 113–126. https://doi.org/10.56211/sudo.v3i3.583

Mauliddiyah, S. M. N. F. H. F. R. (2024). *Analisis sentiment ulasan aplikasi pembelajaran duolinggo di play store menggunakan distilbert*. *7*, 502–511. https://doi.org/10.37600/tekinkom.v7i1.1395

Nugroho, P. A., Fenriana, I., dan Arijanto, R. (2020). Implementasi Deep Learning Menggunakan Convolutional Neural Network (CNN) Pada Ekspresi Manusia. *Algor*, *2*(1), 12–21.

Putri, A. W., Wibhawa,(2023). *KESEHATAN MENTAL MASYARAKAT INDONESIA (PENGETAHUAN , DAN KETERBUKAAN MASYARAKAT TERHADAP GANGGUAN KESEHATAN MENTAL)*. 252–258.

Sihombing, Y. O. N. V. S. (2024). *Prediksi Sentimen Pada Teks Media Sosial Corporate University*. 302–316.

Sudianto, S. (2022). Analisis Kinerja Algoritma Machine Learning Untuk Klasifikasi Emosi. *Building of Informatics, Technology and Science (BITS)*, *4*(2), 1027–1034. https://doi.org/10.47065/bits.v4i2.2261

Sutyo, Febrian R.A. Harahap, Nazruddin S., D. (2024). *Implementasi Question Answering Berbasis Chatbot Telegram Pada Tafsir*. *4*(5), 2464–2472. https://doi.org/10.30865/klik.v4i5.1784

Takwin, B. (2024). *PERAN PSIKOLOGI DALAM PENGEMBANGAN DAN PENGGUNAAN KECERDASAN BUATAN*.

Vaswani, A. (2017). *Attention Is All You Need*. *Nips*.

Wamba, S. F., Bawack, R. E., Guthrie, C., Queiroz, M. M., dan Carillo, K. D. A., D. (2021). *Are we preparing for a good AI society ? A bibliometric review and research agenda*. *19*.

World Health Organization. (2024). *WORLD HEALTH ORGANIZATION - World health statistics 2024. ISBN 9789240094703. tatistics 2024*.