

TESIS

**OPTIMALISASI MODEL TRANSFORMER UNTUK
IDENTIFIKASI CYBERBULLYING DI MEDIA SOSIAL
PADA KONDISI KETERBATASAN SUMBER DAYA**

KOMPUTASI



Oleh :

Luky Vianika Sari

NIM: 23206052010

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA**

2025

HALAMAN PENGESAHAN



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-1896/Un.02/DST/PP.00.9/08/2025

Tugas Akhir dengan judul : OPTIMALISASI MODEL TRANSFORMER UNTUK IDENTIFIKASI CYBERBULLYING DI MEDIA SOSIAL PADA KONDISI KETERBATASAN SUMBER DAYA KOMPUTASI

yang dipersiapkan dan disusun oleh:

Nama : LUKY VIANIKA SARI, S.kom
Nomor Induk Mahasiswa : 23206052010
Telah diujikan pada : Kamis, 07 Agustus 2025
Nilai ujian Tugas Akhir : A

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang
Dr. Agung Fatwanto, S.Si., M.Kom.
SIGNED

Valid ID: 68a7b9b1c48a0



Penguji I
Muhammad Mustakim, S.T. M.T.
SIGNED

Valid ID: 68a7d4576ad8f



Penguji II
Dr. Agus Mulyanto, S.Si., M.Kom., ASEAN
Eng.
SIGNED

Valid ID: 68a4ff5c53585



Yogyakarta, 07 Agustus 2025
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi
Prof. Dr. Hj. Kharul Wardati, M.Si.
SIGNED

Valid ID: 68a7d06c1512c

SURAT PERNYATAAN KEASLIAN

PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama : Luky Vianika Sari
NIM : 23206052010
Jenjang : Magister
Program Studi : Informatika

menyatakan bahwa naskah tesis ini secara keseluruhan adalah hasil penelitian/karya saya sendiri, kecuali pada bagian-bagian yang dirujuk sumbernya.

Yogyakarta, 4 Agustus 2025

Saya yang menyatakan,


10000
TEMPER
89AF1ANX006935523

Luky Vianika Sari
NIM: 23206052010

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

PERNYATAAN BEBAS PLAGIASI

PERNYATAAN BEBAS PLAGIASI

Yang bertanda tangan di bawah ini:

Nama : Luky Vianika Sari
NIM : 23206052010
Jenjang : Magister
Program Studi : Informatika

menyatakan bahwa naskah tesis ini secara keseluruhan benar-benar bebas dari plagiasi. Jika di kemudian hari terbukti melakukan plagiasi, maka saya siap ditindak sesuai ketentuan hukum yang berlaku.

Yogyakarta, 4 Agustus 2025

Saya yang menyatakan,


Luky Vianika Sari
NIM: 23206052010

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

NOTA DINAS PEMBIMBING

NOTA DINAS PEMBIMBING

Kepada Yth.,
Dekan Fakultas Sains dan Teknologi
UIN Sunan Kalijaga
Yogyakarta

Assalamu 'alaikum wr. wb.

Setelah melakukan bimbingan, arahan, dan koreksi terhadap penulisan tesis yang berjudul:

OPTIMALISASI MODEL TRANSFORMER UNTUK IDENTIFIKASI CYBERBULLYING DI MEDIA SOSIAL PADA KONDISI KETERBATASAN SUMBER DAYA KOMPUTASI

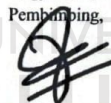
Yang ditulis oleh:

Nama	: Luky Vianika Sari
NIM	: 23206052010
Jenjang	: Magister
Program Studi	: Informatika

Saya berpendapat bahwa tesis tersebut sudah dapat diajukan kepada Magister Informatika UIN Sunan Kalijaga untuk diuji dalam rangka memperoleh gelar Magister Informatika.

Wassalamu 'alaikum wr. wb.

Yogyakarta, 4 Agustus 2025
Pembimbing,



Dr. Agung Fatwanto, S.Si., M.Kom.
19770103 200501 1 003

ABSTRAK

Kasus *cyberbullying* di media sosial terus meningkat dan berdampak serius pada kesehatan mental, khususnya di kalangan remaja. Meskipun berbagai pendekatan berbasis kecerdasan buatan telah dikembangkan untuk mendeteksi *cyberbullying*, banyak di antaranya masih bergantung pada arsitektur model yang kompleks, proses prapemrosesan yang rumit, serta kebutuhan komputasi yang tinggi, sehingga kurang ideal untuk diterapkan pada lingkungan dengan keterbatasan sumber daya.

Studi ini bertujuan untuk mengoptimalkan model berbasis *transformer* untuk klasifikasi *cyberbullying* multikelas pada konten media sosial dalam kondisi keterbatasan sumber daya komputasi. Dalam penelitian ini, peneliti mengevaluasi dan membandingkan kinerja dan upaya pelatihan dari beragam model berbasis *transformer* (BERT, RoBERTa, XLNet, HateBERT), arsitektur hibrida (RoBERTa+BiLSTM, RoBERTa+CNN, BERT+BiLSTM), dan model *ensemble majority voting*. Dataset yang digunakan berisi lebih dari 47.000 tweet berbahasa Inggris yang diberi label ke dalam enam kategori: agama, jenis kelamin, etnis, usia, *cyberbullying* lainnya, dan non-*cyberbullying*.

Seluruh model yang digunakan pada penelitian ini disempurnakan dengan pengaturan hiperparameter default, tanpa penyetelan kompleks, rekayasa fitur eksternal, maupun teknik *ensemble* lanjutan seperti *stacking*. Model *ensemble* dengan pendekatan *majority voting* yang menggabungkan RoBERTa,

BERT, dan XLNet menghasilkan performa terbaik, dengan akurasi 85,01%, *macro F1 score* 85,04%, dan *weighted F1 score* 84,88%. Meskipun tidak melampaui akurasi sistem yang sangat terspesialisasi dari penelitian sebelumnya, pendekatan ini menonjolkan kesederhanaan dan efisiensi operasional. Hasil ini menunjukkan bahwa ensemble berbasis transformer dapat secara efektif mengklasifikasikan berbagai jenis *cyberbullying* dengan konfigurasi yang ringan dan praktis, sehingga layak diterapkan pada lingkungan dengan keterbatasan sumber daya.

Kata kunci: Cyberbullying, Text Classification, Transformers Model, Hybrid Models, Ensemble Learning



ABSTRACT

Cases of cyberbullying on social media continue to rise and have serious impacts on mental health, particularly among teenagers. Although various artificial intelligence-based approaches have been developed to detect cyberbullying, many of them still rely on complex model architectures, complicated preprocessing steps, and high computational demands, making them less ideal for use in resource-constrained environments.

This study aims to optimize a transformer-based model for multi-class cyberbullying classification on social media content under limited computational resources. In this research, the authors evaluate and compare the performance and training effort of various transformer-based models (BERT, RoBERTa, XLNet, HateBERT), hybrid architectures (RoBERTa+BiLSTM, RoBERTa+CNN, BERT+BiLSTM), and a majority-voting ensemble model. The dataset used contains more than 47,000 English-language tweets labeled into six categories: religion, gender, ethnicity, age, other cyberbullying, and non-cyberbullying.

All models in this study were fine-tuned using default hyperparameter settings, without complex tuning, external feature engineering, or advanced ensemble techniques such as stacking. The majority-voting ensemble combining RoBERTa, BERT, and XLNet achieved the best performance, with an accuracy of 85.01%, a macro F1 score of 85.04%, and a weighted F1 score of 84.88%. Although it does not surpass the accuracy of highly

specialized systems from previous studies, this approach emphasizes simplicity and operational efficiency. These results indicate that transformer-based ensembles can effectively classify various types of cyberbullying with a lightweight and practical configuration, making them suitable for deployment in resource-limited environments.

Keyword: Cyberbullying, Text Classification, Transformers Model, Hybrid Models, Ensemble Learning



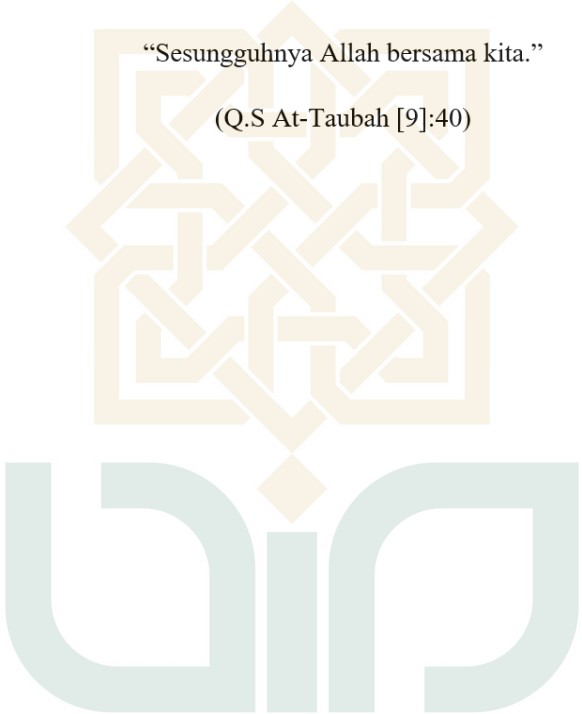
MOTTO

إِنَّ اللَّهَ مَعَنَا

Innallâha ma'anâ

“Sesungguhnya Allah bersama kita.”

(Q.S At-Taubah [9]:40)



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN PERSEMBAHAN

Segala puji dan syukur saya panjatkan kepada Allah SWT atas limpahan rahmat dan hidayah-Nya dalam setiap langkah pencarian ilmu. Ucapan terima kasih saya sampaikan kepada diri saya sendiri atas kerja keras dan ketekunan yang telah membawa saya hingga ke titik ini. Rasa hormat dan terima kasih yang mendalam juga saya tujukan kepada kedua orang tua dan saudara-saudara yang senantiasa memberikan doa, restu, serta dukungan tanpa henti. Penghargaan saya sampaikan kepada seluruh sivitas akademika UIN Sunan Kalijaga Yogyakarta atas segala bantuan, baik secara langsung maupun tidak langsung, selama saya menempuh pendidikan di jenjang magister. Tak lupa, saya juga mengucapkan terima kasih kepada teman dekat saya yang selalu menjadi tempat bertukar pikiran, memberikan masukan berharga, serta mendampingi dalam proses panjang penyusunan karya ini.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Segala puji bagi Allah SWT, Tuhan semesta alam. Dengan limpahan rahmat dan petunjuk-Nya, penulis akhirnya dapat menyelesaikan laporan tugas akhir tesis yang berjudul *“Optimalisasi Model Transformer untuk Identifikasi Cyberbullying di Media Sosial pada Kondisi Keterbatasan Sumber Daya Komputasi”* dengan lancar. Sholawat serta salam senantiasa tercurah kepada Nabi Muhammad SAW, pembawa cahaya peradaban bagi umat manusia. Semoga kita semua mendapatkan syafa'at beliau di hari akhir. Tesis ini disusun sebagai hasil akhir dari kegiatan penelitian untuk memenuhi syarat memperoleh gelar Magister pada Program Studi Informatika, Fakultas Sains dan Teknologi, UIN Sunan Kalijaga Yogyakarta. Selain sebagai bentuk akademik, penulis berharap karya ini dapat memberikan kontribusi pengetahuan dan wawasan bagi para pembaca.

Penyusunan tesis ini tidak lepas dari peran dan dukungan banyak pihak. Oleh karena itu, dengan segala kerendahan hati, penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT yang selalu memberikan kekuatan, semangat, dan kemudahan dalam setiap langkah penulis.
2. Kedua orang tua tercinta, Bapak Burhanudin S.Ag dan Ibu Sri Puji Astuti S.Pdi, serta seluruh saudara saya, atas cinta, doa, dan dukungan baik secara moral maupun materi.

3. Bapak Dr. Agung Fatwanto, S.Si., M.Kom., selaku pembimbing tesis, atas arahan, masukan, dan bimbingan yang sangat berarti selama proses penulisan ini.
4. Ibu Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D., sebagai pembimbing akademik, atas dukungan dan motivasi yang diberikan selama masa studi.
5. Seluruh dosen dan staf di Program Studi Magister Informatika, Fakultas Sains dan Teknologi, UIN Sunan Kalijaga, yang telah membekali penulis dengan ilmu dan pengalaman berharga.
6. Rekan-rekan seperjuangan angkatan 2023 Magister Informatika, yang selalu memberikan semangat, kerja sama, dan kebersamaan selama masa perkuliahan.
7. Secara khusus, ucapan terima kasih penulis sampaikan kepada sahabat saya, Nadya Fitri Fajriyah, S.Kom., atas dukungannya selama proses penyusunan, terutama dalam berbagi ide, diskusi, dan pertukaran pemikiran yang sangat membantu.
8. Serta semua pihak yang telah turut memberikan kontribusi, baik secara langsung maupun tidak langsung, yang tidak dapat disebutkan satu per satu.

Semoga segala kebaikan yang telah diberikan mendapatkan balasan terbaik dari Allah SWT. Penulis menyadari bahwa masih terdapat kekurangan dalam penyusunan tesis ini. Oleh karena itu, saran dan kritik yang membangun sangat diharapkan guna dapat

memberikan manfaat bagi pengembangan ilmu pengetahuan dan masyarakat pada umumnya. Akhir kata, penulis menyampaikan permohonan maaf atas segala kekurangan dan mengucapkan terima kasih atas segala perhatian yang telah diberikan.

Wassalamualaikum Warahmatullahi Wabarakatuh.

Yogyakarta, 4 Agustus 2025

Penulis

Luky Vianika Sari

23206052010

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR ISI

HALAMAN PENGESAHAN.....	i
SURAT PERNYATAAN KEASLIAN	ii
PERNYATAAN BEBAS PLAGIASI	iii
NOTA DINAS PEMBIMBING.....	iv
ABSTRAK	v
ABSTRACT	vii
MOTTO	ix
HALAMAN PERSEMBAHAN	x
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiv
DAFTAR GAMBAR.....	xvii
DAFTAR TABEL	xviii
BAB I PENDAHULUAN.....	1
A. Latar Belakang Masalah	1
B. Rumusan Masalah.....	4
C. Tujuan Penelitian	4
D. Batasan Masalah	5
E. Manfaat Penelitian	6
BAB II KAJIAN PUSTAKA DAN LANDASAN TEORI	7
A. Kajian Pustaka	7
B. Landasan Teori	13

1. Klasifikasi.....	13
2. Transformer-Based Model.....	14
3. Hybrid Model	15
4. Ensemble Learning.....	15
5. Cyberbullying.....	16
BAB III METODE PENELITIAN	18
A. Pendekatan Penelitian	18
B. Alat dan Lingkungan Eksperimen	19
C. Data Penelitian.....	21
D. Tahapan Penelitian.....	24
1. Pra-pemrosesan.....	28
2. Tokenisasi.....	29
3. Klasifikasi.....	30
4. Evaluasi Kinerja	30
BAB IV HASIL DAN PEMBAHASAN	33
A. Hasil.....	33
1. Pra-Pemrosesan	33
2. Tokenisasi.....	34
3. Klasifikasi.....	35
4. Evaluasi Kinerja	38
B. Pembahasan	41
BAB V PENUTUP.....	48
A. Kesimpulan	48
B. Saran	49
DAFTAR PUSTAKA	50
LAMPIRAN I SKEMA RANCANGAN	60

LAMPIRAN II CONFUSION MATRIX.....	63
CURICULUM VITAE.....	68



DAFTAR GAMBAR

Gambar 3.1 Tahapan Penelitian Klasifikasi Cyberbullying.....	27
Gambar 4.1 Confusion Matrix dari prediksi Ensemble learning (RoBERTa, BERT, XLNet).....	37
Gambar 4.2 Grafik Perbandingan Akurasi Training dan Validasi Model Transformer	38
Gambar 4.3 Perbandingan Akurasi Model Hybrid	39
Gambar 4.4 Perbandingan Akurasi Model Klasifikasi	40



DAFTAR TABEL

Tabel 2.1 Perbandingan Penelitian Terdahulu dengan Penelitian Saat Ini	11
Tabel 3.1 Library dan Fungsinya dalam Penelitian	20
Tabel 3.2 Model yang Digunakan dalam Penelitian	21
Tabel 3.3 Distribusi Jumlah Data Berdasarkan Kategori Cyberbullying.....	22
Tabel 3.4 Pembagian Dataset untuk Eksperimen.....	23
Tabel 3.5 Teknik Pra-Pemrosesan Teks yang Digunakan dalam Penelitian.....	28
Tabel 4.1 Contoh Hasil Pra-Pemrosesan pada Dataset Twitter ..	33
Tabel 4.2 Contoh Hasil Sebelum dan Sesudah Tokenisasi pada Model RoBERTa.....	34
Tabel 4.3 Contoh Hasil Tokenisasi Sebelum dan Sesudah pada Model HateBERT.....	34
Tabel 4.4 Hasil Tokenisasi Sebelum dan Sesudah Model XLNet	35
Tabel 4.5 Contoh Hasil Tokenisasi Sebelum dan Sesudah pada Model BERT.....	35
Tabel 4.6 Hasil Perbandingan Kinerja Klasifikasi.....	36
Tabel 4.7 Perbandingan Kompleksitas Model dan Kelayakan Dunia Nyata	44

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Perkembangan pesat teknologi informasi telah mendorong penggunaan media sosial secara masif di seluruh dunia (Kang, H., Wang, Y., Wang, M., Al Imran Yasin, M., Osman, M. N., & Ang, 2024). Platform seperti Twitter, Facebook, Instagram, dan YouTube telah mengubah pola komunikasi masyarakat modern, memungkinkan pertukaran informasi yang cepat dan tanpa batas. Namun, di balik manfaatnya, media sosial juga membuka ruang bagi perilaku negatif, salah satunya adalah *cyberbullying*. Berbeda dengan perundungan konvensional, *cyberbullying* bersifat anonim, tersebar luas, dan dapat terjadi kapan saja tanpa batas ruang dan waktu, sehingga memberikan dampak psikologis yang serius terhadap korbannya, terutama pada kalangan remaja (Samarasinghe dkk., 2022). Studi global menunjukkan bahwa insiden *cyberbullying* terus meningkat dari tahun ke tahun dan menjadi salah satu faktor utama yang memengaruhi kesehatan mental generasi muda, seperti munculnya gangguan kecemasan, depresi, hingga keinginan untuk bunuh diri (Ningsih dkk., 2023)(Lesmana dkk., 2025)(Kurniawan dkk., 2024).

Upaya deteksi otomatis terhadap *cyberbullying* menjadi penting untuk melindungi pengguna media sosial dari dampak negatif yang ditimbulkannya (Rejeb dkk., 2025). Sayangnya, sistem deteksi saat ini masih mengandalkan pelaporan manual dari

pengguna atau pendekatan berbasis pencocokan kata kunci (*keyword-based*) (Zhang, H., & Shang, 2025), yang tidak efektif dalam menangkap makna implisit, sarkasme, maupun ujaran kebencian terselubung (Prabhu & Seethalakshmi, 2025). Tantangan lain terletak pada sifat bahasa media sosial yang sangat informal, penuh singkatan, slang, dan sangat kontekstual, sehingga menyulitkan sistem konvensional dalam melakukan klasifikasi secara akurat dan real-time (Philipo dkk., 2024). Meskipun berbagai penelitian telah mengusulkan metode deteksi otomatis yang lebih canggih, sebagian besar pendekatannya masih terlalu kompleks untuk diimplementasikan secara luas di dunia nyata.

Beberapa penelitian sebelumnya menunjukkan capaian akurasi yang tinggi dalam deteksi *cyberbullying*, namun dengan metode yang relatif kompleks. Alqahtani dkk. (Alqahtani & Ilyas, 2024) menerapkan pendekatan *ensemble stacking* yang mengintegrasikan *Decision Tree*, *Random Forest*, dan XGBoost, dengan akurasi 90,71%. Hadi dkk. (Hadi et al., 2024) menggunakan kombinasi Bag-of-Words dan seleksi fitur Chi-Square, yang menghasilkan akurasi 83% dengan SVM dan *Logistic Regression*. Jamjoom dkk. (Jamjoom et al., 2024) mengembangkan RoBERTaNet, yaitu integrasi model RoBERTa dengan fitur GloVe, Word2Vec, dan FastText, dan mencatatkan akurasi tertinggi sebesar 95,9%. Sementara itu, Vivekananth dkk. (Vivekananth & Sharma, 2025) menerapkan MTF-IDF yang diklasifikasikan dengan berbagai model *machine learning* serta

hyperparameter tuning menggunakan Optuna, menghasilkan akurasi sebesar 83,82%. Meskipun menjanjikan, pendekatan-pendekatan tersebut melibatkan arsitektur yang rumit, ekstraksi fitur tambahan, serta proses *tuning* yang memerlukan sumber daya komputasi besar, sehingga sulit untuk diadopsi di lingkungan dengan keterbatasan infrastruktur, seperti sekolah, universitas, atau instansi publik.

Dalam bidang pemrosesan bahasa alami NLP (*Natural Language Processing*), model berbasis *transformer* telah menjadi standar baru dalam berbagai tugas klasifikasi teks, termasuk deteksi *cyberbullying* (Approach, 2024). Arsitektur seperti BERT, RoBERTa, XLNet, dan HateBERT menawarkan keunggulan dalam memahami konteks kalimat secara dua arah dan menyeluruh (Gardazi dkk., 2025). Kemampuan ini jauh melampaui pendekatan tradisional seperti TF-IDF, Naive Bayes, atau SVM yang hanya memproses teks secara leksikal tanpa memperhitungkan konteks makna (Areshey, A., & Mathkour, 2024). *Transformer* mampu menangkap relasi antar-kata dalam kalimat secara global melalui mekanisme *self-attention*, sehingga cocok untuk mengidentifikasi ujaran yang kompleks dan implisit dalam media sosial (Uyun et al., 2025).

Penelitian ini bertujuan untuk mengoptimalkan model berbasis *transformer* untuk klasifikasi *cyberbullying* multikelas pada konten media sosial dalam kondisi keterbatasan sumber daya komputasi. Penelitian ini melakukan evaluasi terhadap model

transformer murni (BERT, RoBERTa, XLNet, dan HateBERT), model hibrida (RoBERTa-BiLSTM, RoBERTa-CNN, dan BERT-BiLSTM), serta pendekatan *ensemble* berbasis *majority voting*. Seluruh model disempurnakan dengan pengaturan hiperparameter default, tanpa penyetelan kompleks, rekayasa fitur eksternal, maupun teknik *ensemble* lanjutan seperti *stacking*. Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score, untuk memastikan penilaian performa yang menyeluruh. Fokus penelitian juga diarahkan pada skenario realistis dengan keterbatasan sumber daya komputasi, agar hasilnya relevan dan dapat diimplementasikan secara luas dalam konteks dunia nyata.

B. Rumusan Masalah

Berdasarkan latar belakang yang dijabarkan sebelumnya, maka rumusan masalah pada penelitian ini adalah:

1. Bagaimana kinerja dan perbandingan model *transformer*, hibrida, dan *ensemble* dalam klasifikasi *cyberbullying* multikelas pada media sosial dengan keterbatasan sumber daya komputasi?
2. Apakah model klasifikasi *cyberbullying* yang paling optimal dan efisien untuk diterapkan dalam konteks nyata dengan keterbatasan sumber daya komputasi?

C. Tujuan Penelitian

Berdasarkan rumusan masalah dan batasan masalah yang sudah ditetapkan sebelumnya, maka ditetapkan tujuan penelitian sebagai berikut:

1. Penelitian ini bertujuan untuk mengevaluasi dan membandingkan kinerja model *transformer* (BERT, RoBERTa, XLNet, dan HateBERT), model hibrida (RoBERTa-BiLSTM, RoBERTa-CNN, dan BERT-BiLSTM), serta pendekatan *ensemble* (*majority voting*) dalam melakukan klasifikasi *cyberbullying* multikelas pada konten media sosial, khususnya dalam skenario dengan keterbatasan sumber daya komputasi.
2. Mengidentifikasi model klasifikasi *cyberbullying* yang paling optimal dan efisien untuk diterapkan dalam konteks nyata dengan keterbatasan sumber daya komputasi.

D. Batasan Masalah

Penelitian ini memiliki lingkup bahasan yang terbatas agar pembahasan tidak meluas, maka batasan masalah dalam penelitian ini adalah:

1. Dataset yang digunakan merupakan data publik berbahasa Inggris dari platform Twitter, yang telah dilabeli ke dalam enam kategori *cyberbullying*, yaitu: berdasarkan agama, usia, jenis kelamin, etnis, *non-cyberbullying*, dan kategori *cyberbullying* lainnya.
2. Model klasifikasi yang dievaluasi meliputi model berbasis *transformer* (BERT, RoBERTa, HateBERT, dan XLNet), model hibrida (RoBERTa+BiLSTM, RoBERTa+CNN, dan BERT+LSTM), serta pendekatan *ensemble learning* menggunakan metode *majority voting*.

3. Penelitian ini tidak mencakup penyetelan hiperparameter (*hyperparameter tuning*), penyeimbangan data (*data balancing*), maupun penerapan teknik optimasi lanjutan lainnya, guna memastikan proses evaluasi mencerminkan skenario realistis dalam kondisi keterbatasan sumber daya komputasi.
4. Evaluasi kinerja model dilakukan dengan empat metrik utama, yaitu: akurasi, presisi, recall, dan F1-score untuk menilai performa dalam klasifikasi multikelas.
5. Penelitian ini dilakukan menggunakan lingkungan komputasi berbasis cloud (Google Colab) dengan spesifikasi perangkat keras terbatas, untuk mensimulasikan keterbatasan sumber daya komputasi dalam penerapan nyata.

E. Manfaat Penelitian

Manfaat penelitian ini adalah memberikan kontribusi dalam pengembangan sistem klasifikasi *cyberbullying* berbasis kecerdasan buatan yang lebih akurat dan aplikatif melalui evaluasi komprehensif terhadap model-model *transformer*, hibrida, dan pendekatan *ensemble learning*. Hasil evaluasi kinerja model menggunakan berbagai metrik memberikan pemahaman menyeluruh mengenai kelebihan dan keterbatasan setiap pendekatan, yang dapat dijadikan referensi untuk penelitian lebih lanjut dalam bidang pemrosesan bahasa alami dan klasifikasi teks di media sosial.

BAB V

PENUTUP

A. Kesimpulan

Berdasarkan hasil penelitian, dapat ditarik beberapa kesimpulan yang menjawab rumusan masalah. Pertama, terkait kinerja dan perbandingan model, penelitian ini mengevaluasi berbagai pendekatan meliputi model transformer (BERT, RoBERTa, XLNet, HateBERT), model hibrida (RoBERTa+BiLSTM, RoBERTa+CNN, BERT+BiLSTM), serta metode ensemble berbasis majority voting. Semua model diuji dalam kondisi standar, tanpa penyetelan hiperparameter dan rekayasa fitur yang kompleks, untuk memastikan reproduktibilitas dan kesesuaian dengan keterbatasan sumber daya komputasi. Hasil evaluasi menunjukkan bahwa model transformer murni memberikan kinerja stabil pada tiap kategori, model hibrida mampu meningkatkan kapasitas representasi namun tidak secara konsisten melampaui transformer, sedangkan pendekatan ensemble berhasil menggabungkan keunggulan tiap model sehingga memberikan hasil yang lebih seimbang.

Kedua, mengenai model yang paling optimal dan efisien untuk konteks nyata, penelitian ini menemukan bahwa ensemble transformer yang menggabungkan RoBERTa, BERT, dan XLNet memberikan performa terbaik. Model ini mencapai akurasi 85,01% dengan presisi makro 85,26%, recall makro 85,15%, skor F1 makro 85,04%, dan weighted F1-score 84,91%. Meskipun studi

sebelumnya melaporkan akurasi lebih tinggi melalui sistem dengan rekayasa intensif, penelitian ini menunjukkan bahwa pendekatan ensemble yang ringan dan modular tetap mampu menghasilkan performa kuat serta dapat digeneralisasi pada klasifikasi cyberbullying multikelas.

Hal ini membuktikan bahwa model dapat diterapkan secara praktis dan efisien, khususnya dalam lingkungan dengan keterbatasan sumber daya seperti institusi pendidikan, pemerintahan, atau organisasi nirlaba. Dengan demikian, kontribusi utama penelitian ini adalah memberikan bukti empiris bahwa strategi berbasis transformer, khususnya melalui ensemble learning, mampu menyeimbangkan aspek performa, kesederhanaan, dan kelayakan operasional. Pendekatan ini dapat menjadi dasar pengembangan sistem deteksi cyberbullying yang efektif sekaligus terjangkau untuk kebutuhan nyata di berbagai sektor.

B. Saran

Penelitian selanjutnya dapat mempertimbangkan penerapan strategi penyetelan minimal, augmentasi data, atau pembobotan ensemble adaptif untuk lebih meningkatkan performa dalam pengaturan dunia nyata yang sangat tidak seimbang atau terus berkembang.

DAFTAR PUSTAKA

- Al Sukhni, H., Sakr, F., Alsmadi, M. K., Abd-Elghany, M., Gomaa, I. A., & Abdallah, S. (2025). Data-Driven Weather Prediction: Integrating Deep Learning and Ensemble Models for Robust Weather Forecasting. *Journal of Cybersecurity & Information Management*, 15(2). (2025). Data-Driven Weather Prediction: Integrating Deep Learning and Ensemble Models for Robust Weather Forecasting. *Journal of Cybersecurity & Information Management*, 15(2).
- Alabdulwahab, A., Haq, M. A., & Alshehri, M. (2023). Cyberbullying Detection using Machine Learning and Deep Learning. *International Journal of Advanced Computer Science and Applications*, 14(10), 424–432. <https://doi.org/10.14569/IJACSA.2023.0141045>
- Alqahtani, A. F., & Ilyas, M. (2024). An Ensemble-Based Multi-Classification Machine Learning Classifiers Approach to Detect Multiple Classes of Cyberbullying. *Machine Learning and Knowledge Extraction*, 6(1), 156–170. <https://doi.org/10.3390/make6010009>
- Approach, B. M. A. F. I. (2024). *Real-Time Cyberbullying Detection in Online Social Networks with CNN- Real-Time Cyberbullying Detection in Online Social Networks with CNN- BiLSTM-GRU Models : A Fog-Based IoT Approach* Date : September , 2024. September.

- Areshey, A., & Mathkour, H. (2024). Exploring transformer models for sentiment classification: A comparison of BERT, RoBERTa, ALBERT, DistilBERT, and XLNet. *Expert Systems with Applications*, 41(11).
- Basha, S. M., & Fathima, A. S. (2023). *Natural language processing: Practical approach*. MileStone Research Publications.
- Cabot, J. H., & Ross, E. G. (2023). Evaluating prediction model performance. *Surgery (United States)*, 174(3), 723–726. <https://doi.org/10.1016/j.surg.2023.05.023>
- Chopra, S., Agarwal, P., Ahmed, J., Biswas, S. S., & Obaid, A. J. (2024). Roberta and BERT: Revolutionizing Mental Healthcare Through Natural. *SN Computer Science*, 5(7), 889.
- Dotse, S. Q., Larbi, I., Limantol, A. M., & De Silva, L. C. (2024). A review of the application of hybrid machine learning models to improve rainfall prediction. *Modeling Earth Systems and Environment*, 10(1), 19–44.
- Du, X. (2025). *Financial Text Analysis Using 1D-CNN: Risk Classification and Auditing Support*. 515–520. <https://doi.org/10.1145/3730436.3730521>
- Fati, S. M., Mahdi, M. A., Hazber, M. A. G., Ahamad, S., Saad, S. A., Ragab, M. G., & Al-Shalabi, M. (2025). Enhancing Multi-

- Class Cyberbullying Classification with Hybrid Feature Extraction and Transformer-Based Models. *CMES - Computer Modeling in Engineering and Sciences*, 143(2), 2109–2131. <https://doi.org/10.32604/cmes.2025.063092>
- Firdaus, D. R., & Soetanto, H. (2024). *MULTINOMIAL NAÏVE BAYES UNTUK PUNDIT*. 5(3), 1–12.
- Gardazi, N. M., Daud, A., Malik, M. K., Bukhari, A., Alsahfi, T., & Alshemaimri, B. (2025). BERT applications in natural language processing: a review. *Artificial Intelligence Review*, 58(6). <https://doi.org/10.1007/s10462-025-11162-5>
- Goldman, O., Caciularu, A., Eyal, M., Cao, K., Szpektor, I., & Tsarfaty, R. (2024). *Unpacking Tokenization: Evaluating Text Compression and its Correlation with Model Performance*. <http://arxiv.org/abs/2403.06265>
- Guo, L. (2025). BiLSTM based Temporal Modeling for Financial Risk Forecasting using Multivariate Time Series Data. *In 2025 3rd International Conference on Data Science and Information System (ICDSIS) IEEE*, 1–6.
- Hadi, Z., Suryadi, E., Akbar, A., Zaenudin, & Muslim, R. (2024). Cyber Bullying Sentiment Analysis Based On Social Categories Using The Chi-Square Test. *Journal Computer and Technology*, 2(1), 1–9.
- Ibrahim, Y. M., Essameldin, R., & Saad, S. M. (2024). Social

- Media Forensics: An Adaptive Cyberbullying-Related Hate Speech Detection Approach Based on Neural Networks With Uncertainty. *IEEE Access*, 12(March), 59474–59484. <https://doi.org/10.1109/ACCESS.2024.3393295>
- Irfani, M., & Khomsah, S. (2024). Analisis Sentimen Berbasis Aspek pada EDOM Pembelajaran Menggunakan Metode CNN dan Word2vec. *Jurnal Sistem Dan Teknologi Informasi (JustIN)*, 12(3), 413. <https://doi.org/10.26418/justin.v12i3.75610>
- Jamjoom, A. A., Karamti, H., Umer, M., Alsubai, S., Kim, T. H., & Ashraf, I. (2024). RoBERTaNET: Enhanced RoBERTa Transformer Based Model for Cyberbullying Detection With GloVe Features. *IEEE Access*, 12(April 2024), 58950–58959. <https://doi.org/10.1109/ACCESS.2024.3386637>
- Jaradat, S., Nayak, R., Paz, A., Ashqar, H. I., & Elhenawy, M. (2024). Multitask Learning for Crash Analysis: A Fine-Tuned LLM Framework Using Twitter Data. *Smart Cities*, 7(5), 2422–2465. <https://doi.org/10.3390/smartcities7050095>
- Kang, H., Wang, Y., Wang, M., Al Imran Yasin, M., Osman, M. N., & Ang, L. H. (2024). Navigating digital network: Mindfulness as a shield against cyberbullying in the knowledge economy era. *Journal of the Knowledge Economy*, 15(3), 13233–13271.

- Kumar, D. (2024). *CYBERBULLYING: TYPES AND THE EFFECTS OF VICTIMIZATION*. Mindful Media: Mental Health Challenges in the Digital Age.
- Kumar, D. (2025). *A Hybrid DeBERTa and Gated Broad Learning System for Cyberbullying Detection in English Text*. 1–17. <http://arxiv.org/abs/2506.16052>
- Kurniawan, K., Yosep, I., Khoirunnisa, K., Mardhiyah, A., & Fitriani, N. (2024). Dampak Cyberbullying terhadap Risiko Bunuh Diri pada Remaja: Narrative Review. *JPP (Jurnal Kesehatan Poltekkes Palembang)*, 19(1), 8–16. <https://doi.org/10.36086/jpp.v19i1.2012>
- Lesmana, T., Theresia, A., Daniela, P., & Rorong, E. (2025). *SIBER DIMEDIASI KECENDERUNGAN DEPRESI REMAJA*. 12, 157–176. <https://doi.org/10.24854/jpu1133>
- Liu, Y., Huang, L., Giunchiglia, F., Feng, X., & Guan, R. (2024). Improved Graph Contrastive Learning for Short Text Classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(17), 18716–18724. <https://doi.org/10.1609/aaai.v38i17.29835>
- Lubis, A. R., & Nasution, M. K. M. (2023). Twitter Data Analysis and Text Normalization in Collecting Standard Word. *Journal of Applied Engineering and Technological Science*, 4(2), 855–863. <https://doi.org/10.37385/jaets.v4i2.1991>

- Mielke, S. J., Alyafeai, Z., Salesky, E., Raffel, C., Dey, M., Gallé, M., Raja, A., Si, C., Lee, W. Y., Sagot, B., & Tan, S. (2021). *Between words and characters: A Brief History of Open-Vocabulary Modeling and Tokenization in NLP*. <http://arxiv.org/abs/2112.10508>
- Ningsih, R., Hardiyansyah, M. R., Nugraha, M. A., Nurasih, N., & Azis, A. (2023). Analisis Pelaksanaan Pendidikan Karakter Untuk Mengatasi Krisis Moral Mahasiswa Program Studi Pendidikan Sejarah Universitas Islam Sumatera Utara. *Education & Learning*, 3(2), 20–25. <https://doi.org/10.57251/el.v3i2.1033>
- Philipo, A. G., Sarwatt, D. S., Ding, J., Daneshmand, M., & ... (2024). Cyberbullying Detection: Exploring Datasets, Technologies, and Approaches on Social Media Platforms. *ArXiv Preprint ArXiv* <https://arxiv.org/abs/2407.12154> <https://arxiv.org/pdf/2407.12154>
- Prabhu, R., & Seethalakshmi, V. (2025). A comprehensive framework for multi-modal hate speech detection in social media using deep learning. *Scientific Reports*, 15(1), 1–21. <https://doi.org/10.1038/s41598-025-94069-z>
- Ramadhana, N., Purwanti, P., & Dahlan, D. (2024). Fenomena Sarkasme Penggunaan Nama Binatang Pada Politik Indonesia dalam Media Sosial X: Kajian Linguistik Forensik. *Ilmu*

Budaya: Jurnal Bahasa, Sastra, Seni, Dan Budaya, 8(3), 390–402.

Razzaq, K., & Shah, M. (2025). Machine Learning and Deep Learning Paradigms: From Techniques to Practical Applications and Research Frontiers. *Computers*, 14(3). <https://doi.org/10.3390/computers14030093>

Rejeb, A., Rejeb, K., Zrelli, I., & Süle, E. (2025). Tracing knowledge diffusion trajectories in the research field of cyberbullying. *Heliyon*, 11(1), e41141. <https://doi.org/10.1016/j.heliyon.2024.e41141>

Salloum, S. A., Basiouni, A., Alfaisal, R., Salloum, A., & Shaalan, K. (2024). Predicting Student Retention in Higher Education Using Machine Learning. *Communications in Computer and Information Science*, 2162 CCIS(July), 197–206. https://doi.org/10.1007/978-3-031-65996-6_17

Salsabila, Sarno, R., Ghozali, I., & Sungkono, K. R. (2024). Improving cyberbullying detection through multi-level machine learning. *International Journal of Electrical and Computer Engineering*, 14(2), 1779–1787. <https://doi.org/10.11591/ijece.v14i2.pp1779-1787>

Samarasinghe, K., Ahamed Sabani, M., & Shafana, M. (2022). *Cyberbullying Behaviors in Anonymous Social Network*. *Icst*, 77–82.

- Shetty, S. H., Shetty, S., Singh, C., & Rao, A. (2022). Supervised machine learning: Algorithms and applications. *Fundamentals and Methods of Machine and Deep Learning: Algorithms, Tools, and Applications*, 1–16. <https://doi.org/10.1002/9781119821908.ch1>
- Sofa, N., Utomo, F. S., & Saputro, R. E. (2025). Eksplorasi Model Hybrid Transformer-Latent Semantic Analysis (LSA) Untuk Pemahaman Konteks Teks Berita Berbahasa Indonesia. *Jurnal Pendidikan Dan Teknologi Indonesia*, 5(5), 1239–1252. <https://doi.org/10.52436/1.jpti.662>
- Tay, Y., Tran, V. Q., Ruder, S., Gupta, J., Chung, H. W., Bahri, D., Qin, Z., Baumgartner, S., Yu, C., & Metzler, D. (2022). Charformer: Fast Character Transformers Via Gradient-Based Subword Tokenization. *ICLR 2022 - 10th International Conference on Learning Representations*, 1–20.
- Tribuana, D., Maramis, L., Resky, A. M., & Hidayat, R. (2025). *Deep Learning*. Serasi Media Teknologi.
- Ullah, A., Khan, S. N., & Nawi, N. M. (2023). Review on sentiment analysis for text classification techniques from 2010 to 2021. *Multimedia Tools and Applications*, 82(6), 8137–8193. <https://doi.org/10.1007/s11042-022-14112-3>
- Uyun, S., Rosalin, R. P., Sari, L. V., & Sucinta, H. H. (2025). *A Hybrid Classification Model Based on BERT for Multi-Class*

Sentiment Analysis on Twitter. 11(2), 194–205.
<https://doi.org/10.26555/jiteki.v11i2.30665>

Vivekananth, P., & Sharma, N. (2025). Detecting Cyberbullying in Social Media: An NLP-Based Classification Framework. *Indian Journal Of Science And Technology*, 18(5), 380–389.
<https://doi.org/10.17485/ijst/v18i5.1491>

Wang, Z., Liu, K., Li, J., Zhu, Y., & Zhang, Y. (2024). Various frameworks and libraries of machine learning and deep learning: a survey. *Archives of Computational Methods in Engineering*, 31(1), 1–24.

Wang, J., Fu, K., & Lu, C. T. (2020). SOSNet: A Graph Convolutional Network Approach to Fine-Grained Cyberbullying Detection. *Proceedings - 2020 IEEE International Conference on Big Data, Big Data 2020*, 1699–1708. <https://doi.org/10.1109/BigData50022.2020.9378065>

Xiang, F., Zhao, Y., Zhang, M., Zuo, Y., Zou, X., & Tao, F. (2024). Ensemble learning-based stability improvement method for feature selection towards performance prediction. *Journal of Manufacturing Systems*, 74, 55–67.

Zalte, J., & Shah, H. (2024). (2024). Contextual classification of clinical records with bidirectional long short-term memory (Bi-LSTM) and bidirectional encoder representations from transformers (BERT) model. *Computational Intelligence*,

40(4), 12692.

Zhang, H., & Shang, J. (2025). Information Filtering. In Natural Language Processing and Applications. *Singapore: Springer Nature Singapore*, 303–329.

