

DAFTAR PAPIRAN PENDAHULUAN

1.1 Latar Belakang

Cyberbullying merupakan tindakan perundungan yang memanfaatkan teknologi digital, yang biasanya terjadi di platform media sosial. Perilaku merugikan ini menjadi semakin mengkhawatirkan pada era digital saat ini karena dapat memberikan dampak signifikan terhadap kondisi emosional dan psikologis korban (Rahayu, 2021). Saat ini kasus *cyberbullying* di Indonesia juga semakin meningkat. Berdasarkan survei yang dilakukan oleh Badan Nasional Penanggulangan Terorisme (BNPT) pada tahun 2023, sekitar 40% remaja di Indonesia tercatat pernah menjadi korban *cyberbullying*. Data tersebut menunjukkan tingginya kasus intimidasi online di kalangan anak-anak dan remaja khususnya di Indonesia.

Dengan kemajuan teknologi, terutama dalam bidang kecerdasan buatan (*artificial intelligence*), *chatbot* telah menjadi salah satu solusi menangani interaksi digital dan masalah yang muncul, termasuk *cyberbullying*. Chatbot bekerja dengan merespons pesan secara otomatis menggunakan *Natural Language Processing* (NLP), membuat interaksi lebih efisien dan responsif (N. Albayrak, 2018).

Chatbot memerlukan pemahaman konteks yang lebih mendalam agar dapat berfungsi dengan baik di berbagai platform seperti Discord. NLP adalah cabang dari kecerdasan buatan yang berfokus pada interaksi antara manusia dan komputer melalui bahasa manusia alami (Indarta, 2022). Dalam beberapa tahun terakhir, teknologi NLP telah mengalami perkembangan pesat. Kemajuan ini terutama didorong oleh peningkatan kemampuan dalam pengenalan otomatis entitas, analisis sentimen, serta peningkatan kecanggihan model-model NLP. Hal ini memungkinkan chatbot untuk memberikan tanggapan yang lebih akurat dan relevan, sehingga secara signifikan meningkatkan pengalaman pengguna (Aji, 2022).

Dalam pengembangan chatbot, salah satu teknologi NLP yang digunakan adalah Bidirectional Encoder Representations from Transformers (BERT). BERT adalah model transformasi berbasis deep learning yang memungkinkan pemrosesan bahasa alami dengan lebih akurat. Tidak

seperti model NLP tradisional, BERT membaca konteks teks secara dua arah (bidirectional), membuatnya lebih efektif dalam memahami maksud pengguna, termasuk ketika percakapan bersifat kompleks atau ambigu (Jacob Devlin, 2019). IndoBERT, yang merupakan versi BERT yang dilatih khusus pada data bahasa Indonesia (Fajri Koto, 2020). IndoBERT menambah kemampuan chatbot untuk memahami kata, frasa, dan struktur kalimat yang umum digunakan dalam bahasa Indonesia, termasuk bahasa gaul atau singkatan yang sering muncul. Dengan kemampuan analisis sentiment dan pengenalan entitas otomatis, chatbot yang menggunakan BERT dengan model IndoBERT dapat mengidentifikasi pesan berisi ujaran kebencian atau intimidasi secara lebih tepat dalam bahasa Indonesia di platform media sosial seperti Discord.

Di platform media sosial Discord, cyberbullying menjadi isu serius karena sifatnya yang memungkinkan obrolan grup besar, komunikasi real-time, dan anonimitas tinggi. Awalnya populer di kalangan komunitas pecinta game, Discord telah berkembang menjadi ruang obrolan bagi berbagai kelompok minat, menjadikan interaksi antar pengguna lebih kompleks dan sulit dipantau secara manual. Beberapa penelitian menunjukkan bahwa platform dengan fitur obrolan grup besar seperti Discord memiliki peningkatan kasus intimidasi atau ujaran kebencian, di mana anonimitas dan komunikasi cepat meningkatkan risiko perilaku berbahaya. Discord berupaya mengurangi risiko ini melalui aturan ketat dan moderasi otomatis, namun kasus cyberbullying tetap tinggi, terutama di server besar yang sulit dimoderasi dengan efektif (Discord, 2024).

Berdasarkan penjelasan diatas penelitian ini dibuat untuk mengembangkan chatbot dengan kemampuan mendeteksi dan menanggulangi cyberbullying menggunakan teknologi Natural Language Processing (NLP) yang didasari oleh meningkatnya angka kasus cyber bullying yang berdampak negatif pada Kesehatan mental korban, serta terbatasnya solusi otomatis dalam mengatasi masalah ini.

1.2 Rumusan Masalah

Bagaimana mengembangkan aplikasi chatbot dengan menggunakan teknologi Natural Language Processing (NLP) untuk mendeteksi dan menanggulangi cyberbullying.

1.3 Tujuan

Tujuan dari penelitian ini adalah untuk mengembangkan chatbot yang mampu mendeteksi, menanggulangi tindakan, serta membantu aksi moderasi terhadap tindakan cyberbullying di platform media sosial.

1.4 Manfaat

Apabila tujuan penelitian ini tercapai, manfaat yang dapat diperoleh adalah sebagai berikut:

1. Manfaat bagi Pengembangan Ilmu Pengetahuan:
Penelitian ini memperkaya pengetahuan dalam bidang Natural Language Processing (NLP), khususnya terkait penerapannya dalam mendeteksi dan menanggulangi cyberbullying.
2. Manfaat bagi Obyek yang Diteliti:
Aplikasi chatbot yang dikembangkan memberikan solusi bagi platform digital untuk mendeteksi dan menanggulangi cyberbullying secara otomatis. Membantu menciptakan lingkungan digital yang lebih aman bagi pengguna, terutama mereka yang rentan terhadap tindakan bullying.
3. Manfaat bagi Universitas:
Penelitian ini dapat meningkatkan reputasi universitas dalam pengembangan teknologi terbaru, khususnya dalam bidang NLP dan teknologi sosial.
4. Manfaat Sosial
Implementasi chatbot anti-cyberbullying ini berpotensi mengurangi kasus cyberbullying dan meningkatkan kesadaran masyarakat tentang pentingnya menjaga etika dalam berkomunikasi di dunia digital.

1.5 Batasan Masalah

1. Chatbot hanya mendeteksi tindakan cyberbullying berdasarkan pesan teks, tidak mencakup multimedia seperti gambar, video, atau suara
 2. Deteksi cyberbullying pada chatbot ini terbatas pada bahasa Indonesia yang didukung oleh model IndoBERT.
 3. Chatbot diimplementasikan di platform sosial media discord.
 4. Chatbot hanya mendeteksi pesan yang sangat memungkinkan mengandung cyberbullying, dan kata kasar, tidak mendukung pendekstrian sarkasme dan kalimat yang mengandung unsur humoris.

1.6 Metodologi Penelitian

Metodologi membahas mengenai berbagai metode ilmiah, langkah-langkah pelaksanaannya, jenis-jenisnya, hingga batasan-batasan dalam penerapan metode ilmiah tersebut. Sementara itu, penelitian adalah upaya untuk memperoleh pengetahuan dengan mengandalkan bukti-bukti faktual melalui prosedur kerja ilmiah yang kritis dan terkontrol.

1.6.1. Tempat dan Waktu Penelitian

Tempat: Lab UBHINUS, Malang

Waktu: Oktober 2024 – Juli 2025

Kegiatan	Tahun											
	Oktober				Februari				Maret			
	1	2	3	4	1	2	3	4	1	2	3	4
Pengujian Hasil									✓		✓	
Projek												

Tabel 1.6.1 Waktu Penelitian

1.6.2. Bahan dan Alat Penelitian

Bahan:

1. Data teks yang mengandung dan tidak mengandung unsur cyberbullying, yang dikumpulkan dari platform digital.
2. Dokumentasi literatur terkait NLP dan cyberbullying.

Alat:

1. Software pemrograman Python, dan library NLP.
2. Alat untuk pengujian chatbot, termasuk perangkat keras komputer dan akses ke platform digital Discord

1.6.3. Pengumpulan Data dan Informasi

Pengumpulan Data:

1. Data teks yang dikumpulkan dari HuggingFace dan Kaggle. Data ini mencakup pesan yang relevan untuk mendeteksi cyberbullying dan non-cyberbullying.
2. Pelabelan data dilakukan dengan menandai pesan yang mengandung unsur bullying dan yang tidak, untuk melatih model NLP.

Pengumpulan Informasi:

1. Studi literatur dilakukan untuk mengidentifikasi teori dan teknologi terkini. Informasi diperoleh dari jurnal ilmiah, artikel, dan studi kasus.

1.6.4. Analisis Data

Analisis data dalam penelitian ini dilakukan dengan metode-metode berikut untuk memastikan bahwa hasilnya akurat dan relevan dengan tujuan pengembangan chatbot anti-cyberbullying:

1. Analisis Deskriptif:

1. Tujuan: Untuk memberikan gambaran umum tentang data yang dikumpulkan, termasuk karakteristik umum pesan yang mengandung cyberbullying dan yang tidak.
2. Metode: Menggunakan statistik deskriptif untuk menganalisis frekuensi, distribusi, dan pola dalam data teks. Contohnya termasuk penghitungan jumlah pesan yang terdeteksi sebagai bullying, perbandingan antara pesan yang mengandung dan tidak mengandung unsur bullying, serta analisis distribusi kata dan frasa yang sering muncul.

2. Analisis Gap:

1. Tujuan: Untuk mengidentifikasi kesenjangan antara kemampuan deteksi chatbot saat ini dan kebutuhan atau ekspektasi yang diinginkan.
2. Metode: Membandingkan kinerja model NLP dalam mendeteksi cyber bullying dengan standar atau benchmark yang ada. Gap analisis dilakukan dengan mengukur akurasi deteksi, false positives (hasil yang salah mengidentifikasi pesan non-bullying sebagai bullying), dan false negatives (hasil yang gagal mendeteksi pesan bullying).

3. Analisis Sebab-Akibat:

1. Tujuan: Untuk memahami hubungan sebab-akibat dalam data teks, khususnya bagaimana pola tertentu dalam pesan berkontribusi pada terjadinya cyberbullying.
2. Metode: Menggunakan teknik analisis regresi atau model kausal untuk mengidentifikasi faktor-faktor yang mempengaruhi terjadinya bullying. Ini melibatkan analisis hubungan antara berbagai fitur teks (seperti penggunaan kata-kata ofensif, panjang pesan, dil.) dan probabilitas terjadinya bullying.

4. Analisis Kinerja Model:

1. Tujuan: Untuk mengevaluasi efektivitas model NLP dalam mendeteksi dan menanggulangi cyberbullying.
2. Metode: Melakukan evaluasi kinerja model dengan metrik seperti akurasi dan F1 score. Evaluasi ini mencakup analisis hasil deteksi pada data uji untuk menilai seberapa baik model mampu mengidentifikasi pesan yang mengandung cyberbullying serta memberikan respons yang sesuai.

1.6.5. Prosedur Penelitian

Prosedur penelitian ini dijalankan secara sistematis melalui beberapa tahapan yang dirancang untuk mengembangkan dan menguji chatbot anti-cyberbullying berbasis NLP.



Gambar 1.1 Diagram Alir Penelitian

Langkah-langkah ini dijelaskan secara terperinci sebagai berikut:

1. Tahap Persiapan:

1. Studi Literatur: Mengkaji teori-teori terkait cyber bullying, teknologi chatbot, dan NLP dari jurnal, buku, serta penelitian sebelumnya. Studi ini memberikan dasar teoritis dan referensi metode yang digunakan.
 2. Identifikasi Kebutuhan Sistem: Berdasarkan kajian literatur, dilakukan identifikasi kebutuhan untuk pengembangan chatbot yang mampu mendeteksi dan menanggulangi cyber bullying. Ini mencakup pemahaman tentang skenario penggunaan, platform target, dan bahasa yang didukung.
 3. Pengumpulan Data: Mengumpulkan data teks dari berbagai platform digital, seperti Discord. Data ini berisi contoh pesan yang mengandung unsur bullying dan non-bullying.
2. Tahap Pengembangan Model:
1. Pelabelan Data: Melakukan pelabelan data teks secara manual, menandai pesan yang mengandung unsur bullying dan pesan yang tidak, untuk digunakan dalam melatih model NLP.
 2. Pengembangan Model NLP: Menggunakan data berlabel untuk melatih model NLP yang mampu mendeteksi cyber bullying. Proses ini mencakup pemilihan algoritma pembelajaran mesin dan optimasi model.
 3. Pengembangan Chatbot: Mengintegrasikan model NLP ke dalam chatbot. Chatbot dikembangkan untuk dapat berinteraksi dengan pengguna secara otomatis, memberikan peringatan, saran, atau melaporkan pesan yang terdeteksi sebagai bullying.
3. Tahap Implementasi dan Pengujian:
1. Implementasi Chatbot: Mengimplementasikan chatbot pada platform sosial media Discord.
 2. Pengujian Chatbot: Melakukan uji coba chatbot pada platform tersebut dengan skenario yang telah dirancang. Pengujian difokuskan pada akurasi deteksi pesan bullying, respons chatbot, serta kemampuan interaksi dengan pengguna.
 3. Evaluasi Kinerja: Mengumpulkan data hasil uji coba untuk dievaluasi, termasuk metrik performa seperti akurasi, dan F1 score.

4. Tahap Analisis dan Evaluasi:
 1. Analisis Data Hasil Pengujian: Menganalisis hasil pengujian chatbot untuk mengetahui efektivitas model dalam mendeteksi bullying dan respons yang diberikan.
 2. Evaluasi Umpulan Balik Pengguna: Mengumpulkan umpan balik dari pengguna atau kelompok uji untuk mengevaluasi pengalaman pengguna dan area yang perlu diperbaiki dalam pengembangan chatbot.
 3. Penyesuaian Model dan Chatbot: Berdasarkan hasil analisis dan umpan balik, dilakukan penyesuaian dan perbaikan pada model NLP dan chatbot jika diperlukan.
5. Tahap Dokumentasi dan Pelaporan:
 1. Dokumentasi Penelitian: Menyusun seluruh hasil penelitian, termasuk metode, hasil uji coba, analisis, dan evaluasi, dalam bentuk laporan penelitian.
 2. Penyusunan Laporan Akhir: Laporan akhir penelitian mencakup kesimpulan, saran, dan rekomendasi untuk pengembangan lebih lanjut di masa depan.

1.7 Sistematika Penulisan

Sistematika penulisan dalam tugas akhir ini disusun untuk memberikan gambaran umum mengenai pembahasan pada setiap bab yang dijelaskan secara lebih detail. Struktur ini terdiri dari lima bab yang diatur sebagai berikut:

1. Pendahuluan: Bab ini menguraikan latar belakang penelitian yang menjelaskan alasan pentingnya pengembangan chatbot anti-cyber bullying berbasis Natural Language Processing (NLP). Kemudian, disajikan rumusan masalah yang dirumuskan untuk memfokuskan penelitian, tujuan yang ingin dicapai, manfaat penelitian bagi masyarakat, universitas, dan penulis. Selain itu, batasan masalah dan metodologi penelitian yang digunakan juga dijelaskan. Bab ini diakhiri dengan penjelasan sistematika penulisan tugas akhir.
2. Tinjauan Pustaka: Bab ini berisi ulasan mengenai landasan teori yang mendukung penelitian, termasuk konsep dasar cyber bullying, teknologi

chatbot, dan Natural Language Processing (NLP). Penelitian-penelitian terdahulu yang relevan dengan topik penelitian juga diulas untuk memberikan gambaran umum tentang perkembangan penelitian sebelumnya. Selain itu, kerangka pemikiran yang digunakan dalam penelitian ini juga dijelaskan.

3. Metodologi Penelitian: Bab ini menjelaskan secara rinci metode yang digunakan dalam penelitian. Dimulai dengan lokasi dan waktu penelitian, kemudian dilanjutkan dengan bahan dan alat yang digunakan. Metode pengumpulan data dan informasi dijelaskan, termasuk proses analisis data menggunakan metode yang tepat. Selanjutnya, prosedur penelitian disajikan secara terperinci, mulai dari perencanaan hingga implementasi dan pengujian.
4. Hasil dan Pembahasan: Bab ini menyajikan hasil dari pengembangan dan pengujian chatbot anti-cyber bullying. Hasil analisis data dari pengujian chatbot dijelaskan secara rinci, termasuk performa deteksi pesan bullying, respons chatbot, serta evaluasi hasil uji coba. Pembahasan mengenai hasil penelitian dilakukan dengan membandingkan hasil tersebut dengan penelitian terdahulu serta mengevaluasi ketercapaian tujuan penelitian.
5. Kesimpulan dan Saran: Bab ini menguraikan kesimpulan dari penelitian, yang didasarkan pada hasil dan pembahasan yang telah dilakukan. Kesimpulan ini mencerminkan pencapaian tujuan penelitian dan menjawab rumusan masalah yang diajukan. Saran-saran untuk pengembangan lebih lanjut chatbot dan penelitian di masa depan juga disampaikan sebagai rekomendasi.
6. Daftar Pustaka: Berisi daftar referensi dan sumber-sumber yang digunakan selama proses penelitian, baik berupa buku, jurnal, artikel, maupun sumber online lainnya.
7. Lampiran: Menyediakan dokumen tambahan seperti data hasil pengujian, kode program, dan materi pendukung lainnya yang relevan dengan penelitian.