

ABSTRAK

Gunawan Adi Wijaya, 2025. **Pengembangan Chatbot Anti Cyberbullying dengan Kemampuan Deteksi dan Penanggulangan Menggunakan Teknologi NLP**. Tugas Akhir, Program Studi Informatika (S1), Universitas Bhinneka Nusantara, Pembimbing: Chaulina Alfianti Oktavia, Co. Pembimbing: Mukhlis Amien

Kata kunci: Chatbot, Cyberbullying, Natural Language Processing, BERT, Deteksi Teks, Discord

Penelitian ini bertujuan mengembangkan chatbot anti cyberbullying dengan kemampuan mendeteksi dan menanggulangi cyberbullying secara otomatis menggunakan teknologi Natural Language Processing (NLP). Latar belakang penelitian ini adalah tingginya kasus cyberbullying pada platform digital yang dapat berdampak pada kesehatan mental korban. Hipotesis awal penelitian ini adalah chatbot berbasis NLP dengan model BERT yang dilatih menggunakan dataset cyberbullying dan hate speech berbahasa Indonesia dapat secara akurat mendeteksi ujaran cyberbullying pada pesan teks di Discord. Penelitian dilakukan dengan mengumpulkan dataset, melakukan preprocessing, pelatihan model IndoBERT dengan evaluasi akurasi dan F1-score, serta integrasi ke Discord bot agar dapat mendeteksi dan memberikan peringatan secara otomatis. Hasil penelitian menunjukkan chatbot berhasil mencapai akurasi di atas 90% dalam mendeteksi pesan yang mengandung cyberbullying dengan latensi rendah pada pengiriman notifikasi. Kesimpulan penelitian ini adalah chatbot dapat membantu moderator dalam menjaga lingkungan komunikasi yang sehat pada platform Discord. Saran untuk penelitian selanjutnya adalah pengembangan fitur pelatihan dan pengambilan dataset lanjutan, seperti menyimpan log pesan dan pemahaman konteks bertingkat.

ABSTRACT

Gunawan Adi Wijaya, 2025. **Development of an Anti-Cyberbullying Chatbot with Detection and Mitigation Capabilities Using NLP Technology.**

Final Project, Study Program Informatics Bachelor Degree,
Universitas Bhinneka Nusantara, Advisor 1 : Chaulina Alfianti
Oktavia, Advisor 2 : Mukhlis Amien

Keyword: Chatbot, Cyberbullying, Natural Language Processing, BERT, Text
Detection, Discord

This research aims to develop an anti-cyberbullying chatbot capable of automatically detecting and mitigating cyberbullying using Natural Language Processing (NLP) technology. The background of this study is the high incidence of cyberbullying on digital platforms, which can negatively impact the mental health of victims. The initial hypothesis of this research is that an NLP-based chatbot utilizing a BERT model trained on Indonesian cyberbullying and hate speech datasets can accurately detect cyberbullying utterances in text messages on Discord. The research was conducted by collecting datasets, performing preprocessing, training the IndoBERT model while evaluating its accuracy and F1-score, and integrating it into a Discord bot to enable automatic detection and notification. The results showed that the chatbot successfully achieved an accuracy of over 90% in detecting messages containing cyberbullying with low latency in notification delivery. The conclusion of this study is that the chatbot can assist moderators in maintaining a healthy communication environment on the Discord platform. Suggestions for future research include the development of advanced dataset collection and training features, such as storing message logs and implementing layered contextual understanding.