UNIVERSITAS
GADJAH MADA
**Deteksi Konten Tidak Layak untuk Anak Usia Dini pada Video YouTube Berbahasa Indonesia Menggunakan
IndoBERT dengan Informasi Tekstual dan Metadata**
Dragan Abrisam Widijanto, Divi Galih Prasetyo Putri, S.Kom., M.Kom., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

Deteksi Konten Tidak Layak untuk Anak Usia Dini pada Video YouTube Berbahasa Indonesia Menggunakan IndoBERT dengan Informasi Tekstual dan Metadata

Dragan Abrisam Widijanto

22/505608/SV/21834

Perkembangan platform video seperti YouTube meningkatkan kebutuhan deteksi konten otomatis sebagai pendukung moderasi, khususnya untuk melindungi balita usia 1–3 tahun dari konten yang tidak layak. Tantangan tersebut lebih signifikan dalam konteks bahasa Indonesia karena karakteristik data teks yang tidak terstruktur dan mengandung bahasa tidak baku. IndoBERT merupakan model bahasa pralatih untuk bahasa Indonesia yang telah banyak digunakan dalam klasifikasi teks, analisis sentimen, dan deteksi ujaran kebencian. Kemampuan tersebut menjadikannya berpotensi untuk memahami karakteristik teks pada metadata video YouTube berbahasa Indonesia. Penelitian ini bertujuan mengembangkan model deteksi konten tidak layak pada video YouTube berbahasa Indonesia yang memanfaatkan informasi tekstual dan metadata numerik. Model yang dikembangkan menggunakan IndoBERT dengan arsitektur tambahan berupa *sliding window*, mekanisme *attention*, *Gated Recurrent Unit*, dan *gated fusion*. Dataset yang digunakan terdiri atas 40.044 video YouTube dengan label konten layak dan tidak layak, dengan kelas positif ditetapkan pada konten tidak layak. Hasil pengujian menunjukkan bahwa model memperoleh *recall* kelas tidak layak sebesar 91,41%, *recall* kelas layak sebesar 85,79%, presisi kelas tidak layak sebesar 81,55%, presisi kelas layak sebesar 93,57%, dan akurasi sebesar 88,08%. *Ablation study* mengonfirmasi bahwa mekanisme *attention* merupakan komponen paling kritis dalam arsitektur model. Penelitian ini juga mengembangkan sistem inferensi berbasis web dengan Next.js dan FastAPI yang memperoleh tingkat penerimaan pengguna sebesar 90% berdasarkan *User Acceptance Testing*. Sistem tersebut mampu membantu pengguna dalam mendeteksi konten tidak layak pada video yang akan ditonton oleh balita.

Kata kunci: *attention*, IndoBERT, klasifikasi teks, *metadata fusion*, moderasi konten, *sliding window*, YouTube

UNIVERSITAS
GADJAH MADA
**Deteksi Konten Tidak Layak untuk Anak Usia Dini pada Video YouTube Berbahasa Indonesia Menggunakan
IndoBERT dengan Informasi Tekstual dan Metadata**
Dragan Abrisam Widijanto, Divi Galih Prasetyo Putri, S.Kom., M.Kom., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

Detection of Inappropriate Content for Early Childhood on Indonesian-Language YouTube Videos Using IndoBERT with Textual Information and Metadata

Dragan Abrisam Widijanto
22/505608/SV/21834

The growth of video platforms such as YouTube has increased the need for automated content detection to support moderation, particularly to protect toddlers aged 1–3 years from inappropriate content. This challenge is pronounced in the Indonesian context due to unstructured text and informal language. IndoBERT is a pre-trained language model for Indonesian widely used in text classification, sentiment analysis, and hate speech detection, making it well suited for modeling text in Indonesian YouTube video metadata. This study aims to develop a detection model for inappropriate content on Indonesian YouTube videos using textual and numerical metadata. The proposed model employs IndoBERT with additional components: a sliding window, attention mechanism, Gated Recurrent Unit (GRU), and gated fusion. The dataset consists of 40,044 YouTube videos labeled as appropriate or inappropriate, with the positive class set to inappropriate content. Experimental results show that the model achieves a recall of 91.41% for the inappropriate class, a recall of 85.79% for the appropriate class, a precision of 81.55% for the inappropriate class, a precision of 93.57% for the appropriate class, and an overall accuracy of 88.08%. The ablation study confirms that the attention mechanism is the most critical component of the architecture. This study also develops a web-based inference system using Next.js and FastAPI, achieving a user acceptance rate of 90% based on User Acceptance Testing (UAT). The system assists users in detecting inappropriate video content before toddlers view it.

Keyword: attention, content moderation, IndoBERT, metadata fusion, sliding window, text classification, YouTube