UNIVERSITAS
GADJAH MADA
Studi Komparatif Arsitektur Deep Learning Berbasis BiLSTM untuk Klasifikasi Konten YouTube Tidak Pantas bagi Anak-Anak dan Implementasinya pada Sistem Berbasis Web
Saidil Halim, Divi Galih Prasetyo Putri, S.Kom., M.Kom., Ph.D
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

STUDI KOMPARATIF ARSITEKTUR DEEP LEARNING BERBASIS *BiLSTM* UNTUK KLASIFIKASI KONTEN YOUTUBE TIDAK PANTAS BAGI ANAK-ANAK DAN IMPLEMENTASINYA PADA SISTEM BERBASIS WEB

Saidil Halim
22/498691/SV/21241

Meningkatnya konsumsi konten YouTube oleh anak-anak menimbulkan kekhawatiran terhadap paparan konten yang tidak sesuai usia. Penelitian ini membandingkan empat arsitektur berbasis *BiLSTM*, yaitu *BiLSTM*, *CNN-BiLSTM*, *BiLSTM-Attention*, dan *CNN-BiLSTM-Attention*, untuk mendeteksi konten tidak pantas pada YouTube menggunakan metadata teks berbahasa Indonesia dari 40.044 video yang dilabeli ke dalam kategori *Appropriate* dan *Inappropriate*. Lima kolom metadata (judul, tag, deskripsi, komentar, dan transkrip) digabungkan menjadi satu representasi teks per video dengan strategi pemotongan head+tail sepanjang 512 token, kemudian keempat model dilatih dengan konfigurasi identik dan dievaluasi menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil evaluasi pada 6.007 video data uji menunjukkan *BiLSTM* sebagai model terbaik dengan accuracy 88,56% dan Macro F1-Score 0,8821, sementara penambahan *CNN* maupun *Attention* tidak meningkatkan performa secara keseluruhan. Namun, *BiLSTM-Attention* mencatat recall tertinggi pada kelas *Inappropriate* sebesar 89,25%, meskipun disertai peningkatan jumlah False Positive yang menunjukkan adanya trade-off antara kemampuan deteksi konten berbahaya dan risiko salah blokir konten aman. Model kemudian diimplementasikan ke dalam sistem berbasis web untuk memeriksa kelayakan konten YouTube melalui input URL, dengan hasil User Acceptance Testing terhadap 10 responden mencapai 88,57%.

Kata kunci: deteksi konten tidak pantas, YouTube, anak-anak, *BiLSTM*, deep learning, klasifikasi teks, metadata video

UNIVERSITAS
GADJAH MADA
Studi Komparatif Arsitektur Deep Learning Berbasis BiLSTM untuk Klasifikasi Konten YouTube Tidak Pantas bagi Anak-Anak dan Implementasinya pada Sistem Berbasis Web
Saidil Halim, Divi Galih Prasetyo Putri, S.Kom., M.Kom., Ph.D
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

A COMPARATIVE STUDY OF BILSTM-BASED DEEP LEARNING ARCHITECTURES FOR CLASSIFYING INAPPROPRIATE YOUTUBE CONTENT FOR CHILDREN AND ITS IMPLEMENTATION IN A WEB-BASED SYSTEM

Saidil Halim
22/498691/SV/21241

The growing consumption of YouTube content by children raises concerns about exposure to age-inappropriate material. This study compares four BiLSTM-based architectures, namely BiLSTM, CNN-BiLSTM, BiLSTM-Attention, and CNN-BiLSTM-Attention, for detecting inappropriate YouTube content using Indonesian-language text metadata from 40,044 videos labeled as Appropriate or Inappropriate. Five metadata columns (title, tags, description, comments, and transcript) were concatenated into a single text representation per video, truncated using a head+tail strategy with a maximum length of 512 tokens, then used to train all four models under identical configurations and evaluate them using accuracy, precision, recall, and F1-score. Evaluation on 6,007 test videos showed BiLSTM as the best-performing model, achieving 88.56% accuracy and a Macro F1-Score of 0.8821, while adding CNN or Attention did not improve overall performance. However, BiLSTM-Attention recorded the highest recall for the Inappropriate class at 89.25%, albeit with more False Positives, indicating a trade-off between detecting harmful content and the risk of blocking safe content. The best model was then implemented into a web-based system for checking YouTube content suitability via URL input, with User Acceptance Testing across 10 respondents reaching 88.57%.

Keywords: *Inappropriate content detection, YouTube, children, BiLSTM, deep learning, text classification, video metadata*