Deteksi Konten YouTube Tidak Layak untuk Anak Batita (1 â€“ 3 Tahun) Berbahasa Indonesia Menggunakan
Ekstraksi Fitur TF-IDF dan IndoBERT dengan Classifier XGBoost dan SVM
Alif Rizqullah Ma'ruf, 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 YOUTUBE TIDAK LAYAK UNTUK ANAK BATITA (1–3 TAHUN) BERBAHASA INDONESIA MENGGUNAKAN EKSTRAKSI FITUR TF-IDF DAN INDOBERT DENGAN CLASSIFIER XGBOOST DAN SVM

Alif Rizqullah Maruf
22/503619SV/21522

YouTube merupakan platform video terbesar di Indonesia dengan 139 juta pengguna dan telah menjadi media hiburan utama bagi anak usia *toddler* (1–3 tahun). Algoritma rekomendasi YouTube tidak mempertimbangkan kesesuaian usia penonton, sehingga sesi menonton anak dapat berujung pada konten tidak layak hanya dalam beberapa klik. Penelitian ini membangun sistem klasifikasi konten video YouTube berbahasa Indonesia bernama *SafeWatch* menggunakan desain eksperimen faktorial 2×2 yang membandingkan empat kombinasi model: *TF-IDF + XGBoost*, *TF-IDF + SVM*, *IndoBERT + XGBoost*, dan *IndoBERT + SVM*. Dataset terdiri dari 40.044 video YouTube berbahasa Indonesia yang dilabeli secara manual ke dalam dua kelas yaitu Layak dan Tidak Layak. Hasil penelitian menunjukkan kombinasi *IndoBERT + XGBoost* mencapai performa tertinggi dengan *F1-Score* 88,23%, diikuti *TF-IDF + SVM* (87,96%), *IndoBERT + SVM* (86,80%), dan *TF-IDF + XGBoost* (65,15%). Temuan ini mengungkap bahwa kecocokan struktural antara representasi fitur dan algoritma *classifier* lebih menentukan dibandingkan keunggulan salah satu komponen secara terpisah. Uji penerimaan pengguna terhadap 15 responden orang tua menghasilkan skor 4,32/5,00 (86,3%) dengan kategori Sangat Diterima. Sistem *SafeWatch* berhasil diintegrasikan sebagai aplikasi web yang membantu pengguna menganalisis kelayakan konten video YouTube berdasarkan URL, dilengkapi visualisasi kata paling berpengaruh dan *confidence score* untuk transparansi keputusan model.

Kata kunci: *IndoBERT*, klasifikasi teks, konten tidak layak, *SHAP*, *SVM*, *XGBoost*, YouTube

UNIVERSITAS GADJAH MADA
Deteksi Konten YouTube Tidak Layak untuk Anak Batita (1 â€“ 3 Tahun) Berbahasa Indonesia Menggunakan Ekstraksi Fitur TF-IDF dan IndoBERT dengan Classifier XGBoost dan SVM
Alif Rizqullah Ma'ruf, 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 YOUTUBE CONTENT FOR INDONESIAN TODDLERS (AGED 1–3 YEARS) USING TF-IDF AND INDOBERT FEATURE EXTRACTION WITH XGBOOST AND SVM CLASSIFIERS

Alif Rizqullah Maruf
22/503619SV/21522

YouTube is the largest video platform in Indonesia with 139 million users and has become a primary entertainment medium for toddler-aged children (1–3 years old). YouTube's recommendation algorithm does not consider viewer age appropriateness, allowing children's viewing sessions to lead to inappropriate content within just a few clicks. This study develops an Indonesian-language YouTube video content classification system named SafeWatch using a 2×2 factorial experimental design comparing four model combinations: TF-IDF + XGBoost, TF-IDF + SVM, IndoBERT + XGBoost, and IndoBERT + SVM. The dataset consists of 40,044 Indonesian-language YouTube videos manually labeled into two classes: Appropriate and Inappropriate. The results show that IndoBERT + XGBoost achieved the highest performance with an F1-Score of 88.23%, followed by TF-IDF + SVM (87.96%), IndoBERT + SVM (86.80%), and TF-IDF + XGBoost (65.15%). These findings reveal that structural compatibility between feature representation and classifier algorithm is more decisive than the superiority of either component individually. The SafeWatch system was successfully integrated as a web application that helps users analyze YouTube video content appropriateness based solely on a URL, equipped with feature importance visualization and confidence scores to enhance decision transparency.

Keyword: *IndoBERT, inappropriate content, SHAP, SVM, text classification, XGBoost, YouTube.*