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/

DAFTAR ISI

HALAMAN JUDUL i
LEMBAR PENGESAHAN ii
PERNYATAAN KEASLIAN PROYEK AKHIR iii
KATA PENGANTAR iv
DAFTAR ISI vi
DAFTAR GAMBAR x
DAFTAR TABEL xii
DAFTAR LAMPIRAN xv
DAFTAR SIMBOL xvi
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 7
1.3 Tujuan dan Manfaat Proyek Akhir 7
1.4 Batasan Penelitian 9
1.5 Sistematika Penulisan 11
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI 13
2.1 Tinjauan Pustaka 13
2.1.1 Analisis Video Tidak layak (*Inappropriate Content*) untuk Anak pada Platform YouTube 13
2.1.2 Klasifikasi Teks Menggunakan Model Bahasa Pra-latih BERT sebagai Feature Extractor *dengan* Classifier Machine Learning 20
2.2 Dasar Teori 28
2.2.1 YouTube sebagai Platform Video Digital 29
2.2.2 Anak Batita (1–3 Tahun) 29

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/

2.2.3 Konten Tidak Layak (*Inappropriate Content*) 30
2.2.4 Natural Language Processing (NLP) 31
2.2.5 Arsitektur *Transformer* 35
2.2.6 BERT (Bidirectional Encoder Representations from Transformers) 37
2.2.7 IndoBERT 37
2.2.8 Extreme Gradient Boosting (XGBoost) 39
2.2.9 Support Vector Machine (SVM) 41
2.2.10 Hyperparameter Tuning 42
2.2.11 SHAP (SHapley Additive exPlanations) 44
2.2.12 Pengambilan Keputusan dengan Opsi Penundaan (Three-Way Decisions) 46
2.2.13 REST API dan FastAPI 47
2.2.14 Notasi Diagram Perancangan Sistem 48
2.2.15 Evaluasi Model Klasifikasi 51
2.2.16 User Acceptance Testing (UAT) 53
BAB III METODE PENELITIAN 56
3.1 Alat dan Bahan 56
3.1.1 Daftar Alat dan Spesifikasi 57
3.2 Tahapan Proyek Akhir 62
3.3 Perancangan Alat/Purwarupa 66
3.3.1 Persiapan Platform Analisis (Arsitektur Sistem) 66
3.3.2 Analisis Data 68
3.3.3 Perancangan Platform Perangkat Lunak 71
3.4 Tahapan Analisis Data 82
3.4.1 Data Collection, Extraction, and Preprocessing 82
3.4.2 Modelling dan Hyperparameter 96

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/

3.4.3 Evaluasi Performa Model 108
3.4.4 Rancangan Uji Penerimaan Pengguna (User Acceptance Testing) 113
BAB IV HASIL DAN PEMBAHASAN 116
4.1 Dataset dan Implementasi Sistem 116
4.1.1 Distribusi Data per Kelas 117
4.1.2 Analisis Data Hilang (Missing Values) 117
4.1.3 Cohen's Kappa 119
4.1.4 Hasil Preprocessing Teks 120
4.1.5 Implementasi Model 122
4.1.6 Implementasi Platform SafeWatch 128
4.2 Hasil Pelatihan 4 Kombinasi Model 133
4.2.1 Hasil 4 Kombinasi Model 133
4.2.2 Kurva ROC 134
4.2.3 Confusion Matrix 136
4.2.4 Analisis Faktorial 2×2 137
4.3 Hasil Uji Penerimaan Pengguna (UAT) 139
4.3.1 Profil Responden 139
4.3.2 Hasil Penilaian Responden 141
4.3.3 Analisis Hasil UAT 142
4.3.4 Masukan dan Saran Pengguna 143
4.4 Pembahasan 143
4.4.1 Kinerja 4 KKombinasi Model 144
4.4.2 Representasi Teks TF-IDF vs IndoBERT Mana Lebih Unggul? 150
4.4.3 Fitur Tekstual Paling Diskriminatif 153
4.4.4 Hasil User Acceptance Testing SafeWatch 161

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/

BAB V PENUTUP 165

5.1 Kesimpulan.....165

5.2 Saran.....166

DAFTAR PUSTAKA 168

LAMPIRAN.....176