ANALISIS SENTIMEN DAN PEMODELAN TOPIK OPINI PUBLIK TERHADAP KEBIJAKAN PEMBATASAN USIA PENGGUNA MEDIA SOSIAL DI INDONESIA MENGGUNAKAN INDOBERT DAN BERTOPIC

Tugas Akhir

Diajukan untuk memenuhi persyaratan mencapai derajat

Sarjana Sistem Informasi

Bagus Handoko Suryo Saputro

NPM: 221712056

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNOLOGI INDUSTRI

UNIVERSITAS ATMA JAYA YOGYAKARTA

2026

HALAMAN PENGESAHAN

Tugas Akhir Berjudul

ANALISIS SENTIMEN DAN PEMODELAN TOPIK OPINI PUBLIK TERHADAP KEBIJAKAN PEMBATASAN USIA PENGGUNA MEDIA SOSIAL DI INDONESIA MENGGUNAKAN INDOBERT DAN BERTOPIC

yang disusun oleh

Bagus Handoko Suryo Saputro

221712056

dinyatakan telah memenuhi syarat pada tanggal 22 Juli 2026

			Keterangan
Dosen Pembimbing 1	:	Citra Yayu' Palangan, S.T., M.Sc.	Telah Menyetujui
Dosen Pembimbing 2	:	Generosa Lukhayu Pritalia, S.T., M.Eng.	Telah Menyetujui
Tim Penguji			
Penguji 1	:	Citra Yayu' Palangan, S.T., M.Sc.	Telah Menyetujui
Penguji 2	:	Hendro Gunawan, S.Si., M.T.	Telah Menyetujui
Penguji 3	:	Elisabeth Marsella, S.S., M.Li.	Telah Menyetujui

Yogyakarta, 22 Juli 2026

Universitas Atma Jaya Yogyakarta

Teknologi Industri

Dekan

ttd.

Prof. Dr. Ir. Parama Kartika Dewa SP., S.T., M.T., IPU, ASEAN Eng.

Dokumen ini merupakan dokumen resmi UAJY yang tidak memerlukan tanda tangan karena dihasilkan secara elektronik oleh Sistem Bimbingan UAJY. UAJY bertanggung jawab penuh atas informasi yang tertera di dalam dokumen ini

LEMBAR PERNYATAAN KEASLIAN PENELITIAN

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Bagus Handoko Suryo Saputro
NPM : 221712056
Program Studi : Sistem Informasi
Fakultas : Fakultas Teknologi Industri
Judul Penelitian : ANALISIS SENTIMEN DAN PEMODELAN TOPIK OPINI PUBLIK TERHADAP KEBIJAKAN PEMBATASAN USIA PENGGUNA MEDIA SOSIAL DI INDONESIA MENGGUNAKAN INDOBERT DAN BERTOPIC

Menyatakan dengan ini:

1. Skripsi ini adalah benar merupakan hasil karya sendiri dan tidak merupakan salinan sebagian atau keseluruhan dari karya orang lain.
2. Memberikan kepada Universitas Atma Jaya Yogyakarta, berupa Hak Bebas Royalti noneksklusif (*Non-Exclusive-Royalty-Free Right*) atas Penelitian ini, dan berhak menyimpan, mengelola dalam pangkalan data, mendistribusikan, serta menampilkan untuk kepentingan akademis, tanpa perlu meminta izin selama tetap mencantumkan nama penulis.
3. Bersedia menanggung secara pribadi segala bentuk tuntutan hukum yang mengikuti atas pelanggaran Hak Cipta dalam pembuatan skripsi ini.

Demikianlah pernyataan ini dibuat dan dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 10 Juli 2026
Yang menyatakan,

Bagus Handoko Suryo Saputro
221712056

PRAKATA

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas berkat dan penyertaan-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul "Analisis Sentimen dan Pemodelan Topik Opini Publik terhadap Kebijakan Pembatasan Usia Pengguna Media Sosial di Indonesia Menggunakan IndoBERT dan BERTopic" dengan baik. Tugas Akhir ini disusun untuk memenuhi salah satu syarat mencapai derajat Sarjana Sistem Informasi di Program Studi Sistem Informasi, Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta.

Penulis menyadari bahwa penyusunan Tugas Akhir ini tidak lepas dari bantuan, bimbingan, dan dukungan berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih kepada:

1. Ibu Citra Yayu' Palangan, S.T., M.Sc., selaku Dosen Pembimbing I, yang telah meluangkan waktu, memberikan arahan, masukan, dan bimbingan dengan penuh kesabaran selama proses penyusunan Tugas Akhir ini.
2. Ibu Generosa Lukhayu Pritalia, S.T., M.Eng., selaku Dosen Pembimbing II, yang telah memberikan bimbingan, saran, dan koreksi yang sangat berarti bagi penyempurnaan Tugas Akhir ini.
3. Seluruh dosen dan staf Program Studi Sistem Informasi, Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta, yang telah memberikan ilmu dan dukungan selama masa perkuliahan.
4. Almarhumah Jumi, yang senantiasa penulis panggil Bobo, atas segala doa, kasih sayang, dan semangat yang telah diberikan sejak awal perkuliahan hingga terselesaikannya Tugas Akhir ini.
5. Lina Susanti, yang akrab disapa YiLin, beserta seluruh keluarga besar, yang selalu memberikan dukungan, perhatian, dan motivasi kepada penulis dalam menyelesaikan studi ini.
6. Seluruh teman penulis yang tidak dapat disebutkan satu per satu, atas kebersamaan, dukungan, semangat, serta bantuan yang telah diberikan selama masa perkuliahan hingga proses penyusunan Tugas Akhir ini.
7. Seluruh pihak yang tidak dapat disebutkan satu per satu, yang telah membantu baik secara langsung maupun tidak langsung dalam penyusunan Tugas Akhir ini.

Penulis menyadari bahwa Tugas Akhir ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi perbaikan di masa mendatang. Akhir kata, penulis berharap Tugas Akhir ini dapat memberikan manfaat bagi pembaca serta menjadi referensi bagi penelitian selanjutnya.

Yogyakarta, 10 Juli 2026
Penulis,

Bagus Handoko Suryo Saputro
221712056

INTISARI

Pemberlakuan kebijakan pembatasan usia minimal 16 tahun bagi pengguna media sosial di Indonesia melalui PP Tunas yang berlaku sejak 28 Maret 2026 memicu volume komentar publik yang tidak terstruktur, sehingga pola sentimen serta topik dominan sulit dipahami secara sistematis. Penelitian ini bertujuan menganalisis distribusi sentimen, mengidentifikasi topik dominan, dan membandingkan sentimen serta topik dominan antara YouTube dan Instagram. Data sebanyak 33.910 komentar hasil prapemrosesan dikumpulkan dari YouTube dan Instagram pada berita resmi dan akun pemerintah. Analisis sentimen dilakukan menggunakan model IndoBERT hasil *fine-tuning* dengan 1.597 data berlabel yang mencapai akurasi 0,9155 dan *Macro F1* 0,8744 pada data validasi, serta akurasi 0,8500 dan *Macro F1* 0,8479 pada data test, sedangkan pemodelan topik menggunakan BERTopic yang menghasilkan 26 topik. Hasil penelitian menunjukkan sentimen positif dominan sebesar 52,5 persen, negatif 26,7 persen, dan netral 20,7 persen. Pemodelan topik menunjukkan bahwa pembahasan masyarakat mencakup mekanisme pembuatan akun dan verifikasi usia, perlindungan anak di ruang digital, penggunaan platform seperti YouTube dan Roblox, penanganan judi *online*, peran orang tua, serta perlindungan generasi muda. Secara lintas platform, respons di Instagram didominasi oleh tanggapan positif sebesar 62,3 persen, sedangkan di YouTube didominasi oleh tanggapan negatif sebesar 56,6 persen.

Kata Kunci: analisis sentimen; pemodelan topik; IndoBERT; BERTopic; media sosial

ABSTRACT

The implementation of Indonesia's minimum age policy of 16 years for social media users through PP Tunas, effective 28 March 2026, triggered unstructured public comments, making sentiment patterns and dominant topics difficult to understand systematically. This study aims to analyze sentiment distribution, identify dominant topics, and compare sentiment and dominant topics between YouTube and Instagram. A total of 33,910 preprocessed comments were collected from YouTube and Instagram on official news and government accounts. Sentiment analysis was conducted using a fine-tuned IndoBERT model with 1,597 labeled data, achieving an accuracy of 0.9155 and a Macro F1 of 0.8744 on the validation set, and an accuracy of 0.8500 and a Macro F1 of 0.8479 on the test set, while topic modeling employed BERTopic, producing 26 topics. The results show that positive sentiment dominated at 52.5 percent, negative at 26.7 percent, and neutral at 20.7 percent. Topic modeling showed that public discussions covered account creation and age verification mechanisms, child protection in the digital space, platforms such as YouTube and Roblox, online gambling, parental roles, and protection of future generations. Across platforms, Instagram was predominantly positive at 62.3 percent, whereas YouTube was predominantly negative at 56.6 percent.

Keywords: *sentiment analysis; topic modeling; IndoBERT; BERTopic; social media*

DAFTAR ISI

LEMBAR PENGESAHAN 1
PRAKATA 3
INTISARI 5
ABSTRACT 6
DAFTAR ISI 7
DAFTAR GAMBAR 10
DAFTAR TABEL 11
BAB I PENDAHULUAN 12
1.1. Latar Belakang 12
1.2. Rumusan Masalah 15
1.3. Pertanyaan Penelitian 15
1.4. Tujuan Penelitian 16
1.5. Batasan Masalah 16
1.6. Manfaat Penelitian 17
1.7. Bagan Keterkaitan 17
BAB II TINJAUAN PUSTAKA 19
2.1. Studi Sebelumnya 19
2.2. Dasar Teori 23
2.2.1. Media Sosial 23
2.2.2. Kebijakan Pembatasan Usia Pengguna Media Sosial 23
2.2.3. Opini Publik 24
2.2.4. Analisis Sentimen 24
2.2.5. Prapemrosesan Teks 25
2.2.6. BERT 25
2.2.7. IndoBERT 26
2.2.8. Pemodelan Topik 27
2.2.9. BERTopic 27
BAB III METODOLOGI PENELITIAN 30
3.1. Tahapan Penelitian 30
3.2. Metode Penelitian 32
3.2.1. Studi Literatur 32
3.2.2. Pengumpulan Data 32
3.2.3. Prapemrosesan Data 34
3.2.3.1. Penyaringan Awal (*Filtering*) 34
3.2.3.2. Pelipatan Huruf dan Pembersihan Teks (*Case Folding* dan *Cleaning*) 34
3.2.3.3. Normalisasi *Slang* 35
3.2.3.4. Tokenisasi, Penghapusan *Stopword*, dan *Stemming* 35
3.2.3.5. Penyaringan Bahasa 36
3.2.3.6. Penyaringan Panjang Token 36
3.2.4. Pelabelan Data 37
3.2.5. Analisis Sentimen dengan IndoBERT 38
3.2.5.1. *Fine-Tuning* IndoBERT 38
3.2.5.2. Inferensi Massal (*Mass Inference*) Seluruh *Dataset* 39
3.2.5.3. Analisis Deskriptif Hasil Sentimen 40
3.2.6. Pemodelan Topik dengan BERTopic 40
3.2.6.1. Representasi Vektor Teks 40

3.2.6.2. Reduksi Dimensi dengan UMAP 40
3.2.6.3. Pengelompokan dengan HDBSCAN 41
3.2.6.4. Representasi Topik dengan c-TF-IDF 42
3.3. Alat yang Digunakan 42
3.3.1. Perangkat Keras 42
3.3.2. Perangkat Lunak 42
3.3.2.1. Bahasa Pemrograman dan Perangkat Pengembangan 42
3.3.2.2. Pustaka Pengumpulan dan Prapemrosesan Data 43
3.3.2.3. Pustaka Analisis Sentimen dan Pemodelan Topik 43
3.3.2.4. Pustaka Visualisasi 43
3.3.2.5. Perangkat Pelabelan Data 43
BAB IV HASIL DAN PEMBAHASAN 44
4.1. Hasil 44
4.1.1. Pengumpulan Data 44
4.1.1.1. Pengumpulan Data YouTube 44
4.1.1.2. Pengumpulan Data Instagram 45
4.1.1.3. Penggabungan Data 47
4.1.2. Prapemrosesan Data 47
4.1.2.1. Penyaringan Awal (*Filtering*) 48
4.1.2.2. Pelipatan Huruf dan Pembersihan Teks 50
4.1.2.3. Normalisasi *Slang* 51
4.1.2.4. Tokenisasi, Penghapusan *Stopword*, dan *Stemming* 53
4.1.2.5. Penyaringan Bahasa 55
4.1.2.6. Penyaringan Panjang Token 57
4.1.3. Pelabelan Data 58
4.1.3.1. Pengambilan Sampel 58
4.1.3.2. Panduan Anotasi 58
4.1.3.3. Proses Anotasi 59
4.1.3.4. Distribusi Hasil Anotasi 60
4.1.3.5. Pembagian Data Latih dan Validasi 61
4.1.4. Analisis Sentimen dengan IndoBERT 62
4.1.4.1. Hasil *Fine-Tuning* dan Evaluasi Model 62
4.1.4.2. Pengaruh Jumlah Data Latih terhadap Kinerja Model 67
4.1.4.3. Distribusi Sentimen Keseluruhan 68
4.1.4.4. Kualitas Prediksi (*Confidence Score*) 70
4.1.4.5. Distribusi Sentimen per Platform 71
4.1.4.6. Distribusi Komentar dan Sentimen per Sumber Konten 72
4.1.4.7. Analisis Leksikal per Kelas Sentimen 73
4.1.5. Pemodelan Topik dengan BERTopic 74
4.1.5.1. Konfigurasi dan Statistik Pemodelan 74
4.1.5.2. Hasil Pemodelan Topik Keseluruhan 78
4.1.5.3. Distribusi Sentimen per Topik 84
4.1.5.4. Perbandingan Sentimen dan Topik Dominan Antarplatform 85
4.2. Pembahasan 87
4.2.1. Pola dan Distribusi Sentimen 87
4.2.2. Topik Dominan 88
4.2.3. Perbandingan Sentimen dan Topik Antarplatform 89

4.2.4. Implikasi 91
4.2.4.1. Implikasi bagi Pembuat Kebijakan 91
4.2.4.2. Implikasi bagi Masyarakat dan Orang Tua 92
4.2.4.3. Implikasi bagi Akademisi 92
BAB V KESIMPULAN DAN SARAN 94
5.1. Kesimpulan 94
5.2. Saran 95
DAFTAR PUSTAKA 97
LAMPIRAN 101

DAFTAR GAMBAR

Gambar 1.1. Bagan Keterkaitan 18
Gambar 3.1. Bagan Tahapan Penelitian 30
Gambar 4.1. Fungsi Pemuatan Data 49
Gambar 4.2. Fungsi Penyaringan Awal 50
Gambar 4.3. Fungsi Pembersihan Teks 51
Gambar 4.4. Fungsi Normalisasi *Slang* 52
Gambar 4.5. Cuplikan Kamus Normalisasi *Slang* 52
Gambar 4.6. Fungsi Tokenisasi, Penghapusan *Stopword*, dan *Stemming* 53
Gambar 4.7. Fungsi Perangkaian Pemrosesan Teks 54
Gambar 4.8. Fungsi Penyaringan Bahasa 56
Gambar 4.9. Penyaringan Panjang Token 57
Gambar 4.10. Antarmuka Anotasi pada Label Studio 60
Gambar 4.11. Konfigurasi *Fine-Tuning* IndoBERT 63
Gambar 4.12. Kurva Pelatihan per *Epoch* 65
Gambar 4.13. Kurva Kinerja Model terhadap Jumlah Data Latih 68
Gambar 4.14. Distribusi Sentimen Keseluruhan 69
Gambar 4.15. Distribusi *Confidence Score* Model 71
Gambar 4.16. Distribusi Sentimen per Platform 72
Gambar 4.17. Sepuluh Sumber dengan Komentar Terbanyak 73
Gambar 4.18. Distribusi Sentimen per Sumber Konten 73
Gambar 4.19. Sepuluh Kata Teratas per Kelas Sentimen 74
Gambar 4.20. *Word Cloud* per Kelas Sentimen 74
Gambar 4.21. Konfigurasi BERTopic 75
Gambar 4.22. Sepuluh Topik dengan Komentar Terbanyak 79
Gambar 4.23. Distribusi Sentimen per Topik 85
Gambar 4.24. Perbandingan Distribusi Sentimen Antarplatform 86
Gambar 4.25. Topik Dominan pada Setiap Platform 87

DAFTAR TABEL

Tabel 2.1. Penelitian Sebelumnya 21
Tabel 3.1. Spesifikasi Perangkat Keras 42
Tabel 4.1. Rincian Pengumpulan Data Komentar YouTube per Akun 45
Tabel 4.2. Rincian Pengumpulan Data Komentar Instagram per Akun 46
Tabel 4.3. Rekapitulasi Jumlah Komentar Berdasarkan Platform dan Kategori Sumber 47
Tabel 4.4. Jumlah Komentar pada Setiap Tahap Prapemrosesan 48
Tabel 4.5. Contoh Hasil Setiap Tahap Pemrosesan Teks 54
Tabel 4.6. Hasil Pengambilan Sampel Komentar untuk Pelabelan 58
Tabel 4.7. Distribusi Label Sentimen Hasil Anotasi 60
Tabel 4.8. Distribusi Label Sentimen per Platform 61
Tabel 4.9. Contoh Komentar untuk Setiap Label Sentimen 61
Tabel 4.10. Pembagian Data Latih dan Validasi 62
Tabel 4.11. Konfigurasi *Fine-Tuning* IndoBERT 63
Tabel 4.12. Perkembangan Pelatihan per *Epoch* 64
Tabel 4.13. Hasil Evaluasi Model per Kelas pada Data Validasi 65
Tabel 4.14. *Confusion Matrix* Hasil Evaluasi pada Data Validasi 65
Tabel 4.15. Hasil Evaluasi Model per Kelas pada *Test Set* 66
Tabel 4.16. *Confusion Matrix* Hasil Evaluasi pada *Test Set* 66
Tabel 4.17. Perbandingan Kinerja Model Tahap Awal dan Tahap Akhir 67
Tabel 4.18. Kinerja Model pada Berbagai Jumlah Data Latih 67
Tabel 4.19. Distribusi Sentimen Keseluruhan 68
Tabel 4.20. Contoh Komentar pada Setiap Kelas Sentimen 69
Tabel 4.21. Rata-rata *Confidence Score* per Kelas Sentimen 70
Tabel 4.22. Distribusi Sentimen per Platform 71
Tabel 4.23. Sepuluh Sumber dengan Komentar Terbanyak 72
Tabel 4.24. Konfigurasi BERTopic 76
Tabel 4.25. Contoh Komentar dengan Token Kurang dari Tiga 76
Tabel 4.26. Contoh Pemeriksaan Manual Komentar *Outlier* 77
Tabel 4.27. Sepuluh Topik dengan Komentar Terbanyak 78
Tabel 4.28. Contoh Komentar pada Topik 0 79
Tabel 4.29. Contoh Komentar pada Topik 1 80
Tabel 4.30. Contoh Komentar pada Topik 2 80
Tabel 4.31. Contoh Komentar pada Topik 3 81
Tabel 4.32. Contoh Komentar pada Topik 4 81
Tabel 4.33. Contoh Komentar pada Topik 5 82
Tabel 4.34. Contoh Komentar pada Topik 6 82
Tabel 4.35. Contoh Komentar pada Topik 7 83
Tabel 4.36. Contoh Komentar pada Topik 8 83
Tabel 4.37. Contoh Komentar pada Topik 9 84
Tabel 4.38. Perbandingan Distribusi Sentimen Antarplatform 85
Tabel 4.39. Lima Topik Terbesar pada Setiap Platform 86