HOAQ: JURNAL TEKNOLOGI INFORMASI, Volume 16 Nomor 2, Desmeber 2025, p193-201
p-ISSN: 2337-5280, e-ISSN: 26207427
DOI: https://doi.org/10.52972/hoaq.vol16no2

ANALISIS SENTIMEN KOMENTAR BERPOTENSI *TOXIC* PADA MEDIA SOSIAL TIKTOK MENGGUNAKAN METODE *DECISION TREE*

Bobbin Ariyadi Jasno[1], Ahmad Ariful Fathoni[2], David[3], Dwiki Dharma Putra[4], Mohammad Zidane Hasan[5], Fachri Amsury[6], Hendra Supendar[7]

[1,2,3,4,5,6,7] Program Studi Teknologi & Informasi, Universitas Bina Sarana Informatika
Jl. Kramat Raya No. 98, Senen – Jakarta Pusat, Indonesia
Email: [1]17230808@bsi.ac.id, [2]17230785@bsi.ac.id, [3]17230682@bsi.ac.id, [4]17230545@bsi.ac.id, [5]17230570@bsi.ac.id, [6] fachri.fcy@bsi.id, [7] hendra.hds@bsi.ac.id

ABSTRAK

Penelitian ini bertujuan untuk mengatasi tantangan dalam mendeteksi dan mengklasifikasikan komentar berpotensi *toxic* secara otomatis pada media sosial TikTok, yang dikenal padat dengan bahasa informal, *slang*, dan *cyber-aggression*, menggunakan Analisis Sentimen dengan algoritma *Decision Tree*. *Dataset* yang digunakan terdiri dari 271 komentar primer yang dikumpulkan langsung dari *feed* video TikTok dan diklasifikasikan secara seimbang ke dalam kategori Toxic (Label = 1) dan Non-Toxic (Label = 0). Tahapan metodologi mencakup normalisasi bahasa slang TikTok, *preprocessing* teks, dan pembobotan fitur menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk menonjolkan fitur linguistik yang berkaitan dengan toksisitas. Hasil pengujian menunjukkan bahwa model menghasilkan akurasi sebesar 0,75, *precision* untuk kelas *toxic* sebesar 0,787, dan *recall* sebesar 0,765, sehingga mencerminkan performa aktual model dalam mendeteksi komentar *toxic* setelah proses *preprocessing* dan TF-IDF. Teknik *post-pruning* turut membantu mengurangi *overfitting* dan meningkatkan kemampuan generalisasi model terhadap data baru, meskipun penelitian ini tidak melakukan pengujian formal terhadap efisiensi komputasi maupun keandalan sistem. Secara keseluruhan, kombinasi normalisasi *slang*, TF-IDF, dan *Decision Tree* dengan *post-pruning* mampu menghasilkan performa klasifikasi yang stabil dalam identifikasi komentar *toxic* pada TikTok berbasis data primer.

Kata kunci: *Decision Tree*, Analisis Sentimen, Komentar *Toxic*, TikTok, TF-IDF.

ABSTRACT

This study aims to address the challenges of automatically detecting and classifying potentially toxic comments on the TikTok social media platform, which is characterized by heavy use of informal language, slang, and cyber-aggression, by applying Sentiment Analysis using the Decision Tree algorithm. The dataset consists of 271 primary comments collected directly from TikTok video feeds and evenly categorized into Toxic (Label = 1) and Non-Toxic (Label = 0). The methodological stages include TikTok-specific slang normalization, text preprocessing, and feature weighting using Term Frequency–Inverse Document Frequency (TF-IDF) to highlight linguistic features associated with toxicity. Experimental results show that the model achieves an accuracy of 0.75, a precision of 0.787, and a recall of 0.765 for the toxic class, reflecting the model's actual performance after preprocessing and TF-IDF optimization. The application of post-pruning also helps reduce overfitting and improves the model's generalization ability toward new data, although the study does not conduct formal evaluations of computational efficiency or system reliability. Overall, the combination of slang normalization, TF-IDF, and a pruned Decision Tree demonstrates stable classification performance in identifying toxic comments on TikTok based on the primary data used.

Keywords: Decision Tree, Sentiment Analysis, Toxic Comments, TikTok, TF-IDF.

1. PENDAHULUAN

Media Sosial TikTok telah menjadi *platform* revolusioner yang memfasilitasi interaksi masif melalui konten video pendek. Popularitasnya yang meluas menciptakan volume data dan interaksi pengguna yang sangat besar dari berbagai usia dan latar belakang. Namun, peningkatan interaksi ini secara otomatis juga menghasilkan lonjakan konten dan komentar yang cenderung negatif, agresif, atau berpotensi *toxic*. Komentar *toxic* ini mencakup berbagai bentuk, mulai dari ucapan kebencian *(hate speech)* hingga perundungan siber *(cyberbullying)*, yang dapat menimbulkan dampak psikologis buruk pada pengguna dan menciptakan lingkungan diskusi yang tidak sehat. Berbagai penelitian telah menyoroti dampak serius dari toksisitas ini terhadap kesehatan mental pengguna [1]. Mengingat volume komentar digunakan pada penelitian ini sebanyak 255 komentar primer yang diambil secara

langsung dari *feed video* TikTok sebagai sample analisis. Diperlukan sebuah pendekatan otomatis berbasis kecerdasan buatan (AI), seperti Analisis Sentimen, untuk mendeteksi dan memoderasi komentar negatif tersebut secara efisien [2].

Dalam upaya mengatasi tantangan deteksi komentar *toxic*, terdapat beberapa kendala utama pada data komentar TikTok, seperti penggunaan bahasa tidak baku, kata slang yang berubah-ubah, serta variasi bentuk penghinaan yang sering kali dituliskan dalam bentuk singkatan modifikasi ejaan. Tantangan ini menyebabkan model klasifikasi kesulitan mengenali pola *linguistik toxic* secara konsisten dan berdampak pada rendahnya performa identifikasi komentar negatif. Bedasarkan kondisi tersebut, penelitian sebelunya menunjukkan bahwa algoritma *Decision Tree* merupakan salah satu pendekatan klasifikasi yang efektif karena memiliki kemampuan membangun aturan keputusan secara eksplisit mengenai ciri-ciri linguistik *toxic* pada data komentar. Algoritma ini dinilai efektif karena kemampuannya menghasilkan model yang relatif mudah diinterpretasikan [3]. Referensi terdahulu memang menunjukkan bahwa algoritma *Decision Tree* mampu mencapai akurasi tinggi dalam analisis sentimen komentar TikTok maupun deteksi kalimat *toxic* berbahasa Indonesia. Namun, penelitian-penelitian tersebut umumnya menggunakan *dataset* yang telah terstruktur atau bersifat sekunder, bukan komentar primer yang diambil secara langsung dari *feed* TikTok. Selain itu belum banyak penelitian yang secara spesifik menangani keragaman bentuk bahasa *slang* TikTok, sehingga tidak dijelaskan bagaiman model mengatasi variasi ejaan, singkatan, atau kata kasar yang dimodifikasi. Keterbatasan ini menjadi celah penting untuk mengevaluasi kinerja *Decision Tree* pada data komentar *real-time* yang memiliki sifat mengganggu dan tidak baku. Selain itu, efektivitas penggunaan *Term Frequency-Inverse Document Frequency* (TF-IDF) sebagai fitur pembobotan telah terbukti meningkatkan kinerja klasifikasi teks di berbagai konteks [4]. Studi lain menekankan pentingnya optimasi ekstraksi fitur TF-IDF dalam klasifikasi sentimen pada komentar *online* yang mengandung kata-kata informal dan slang untuk meningkatkan akurasi model [5]. Permasalahan utama yang kami temukan adalah bahwa belum ada studi yang secara spesifik menggabungkan *Decision Tree*, Analisis Sentimen, dengan data komentar berpotensi *toxic* langsung dari *feed* video TikTok. Meskipun akurasi model tunggal mungkin tinggi, *Decision Tree* sering menghadapi masalah dalam mengidentifikasi tingkat toksisitas yang beragam pada data *real-time* TikTok yang padat dengan bahasa informal, slang, dan *cyber-aggression* [6]. Data jurnal mengindikasikan bahwa tanpa penyesuaian fitur yang tepat, *Decision Tree* dapat memiliki nilai *recall* yang rendah untuk kategori minoritas (misalnya sentimen negatif), sehingga gagal mengidentifikasi semua kasus *toxic* yang ada [7].

Untuk mengatasi permasalahan akurasi klasifikasi yang beragam dan meningkatkan *recall* model, kami mengusulkan solusi melalui implementasi mendalam dari algoritma *Decision Tree* dengan fokus pada tahap pra-pemrosesan data dan rekayasa fitur *(feature engineering)* yang canggih. Solusi ini akan melibatkan *text preprocessing* yang disesuaikan untuk mengatasi tantangan bahasa TikTok, seperti normalisasi bahasa slang dan penghilangan *stop words* yang tidak relevan. Kunci utama metode kami adalah optimasi teknik pembobotan kata *Term Frequency-Inverse Document Frequency* (TF-IDF), yang diharapkan mampu menonjolkan fitur-fitur linguistik unik yang mengindikasikan toksisitas, sehingga model *Decision Tree* dapat membangun aturan klasifikasi (cabang-cabang) yang lebih akurat dan terfokus [8]. Selain itu, untuk mengatasi kecenderungan *Decision Tree* terhadap *overfitting*, kami akan menerapkan teknik *pruning*, khususnya *post-pruning*, untuk menghilangkan cabang yang minim dampak dan meningkatkan kemampuan generalisasi model saat menghadapi data komentar baru [9], [10].

Penelitian ini bertujuan untuk menghasilkan sebuah model *Decision Tree* yang akurat dalam mengklasifikasikan komentar berpotensi *toxic* pada platform TikTok, sesuai dengan ruang lingkup dan karakteristik data yang digunakan. Fokus penelitian tidak ditujukan untuk membuktikan generalisasi model terhadap domain lain di luar TikTok, karena setiap *platform* memiliki pola bahasa dan karakteristik kometar yang berbeda. Oleh sebab itu, aspek generalisasi lintas platform tidak dibahas dalam penelitian ini dan dapat menjadi topik eksplorasi pada penelitian sebelumnya. Secara spesifik, kontribusi penelitian adalah: pertama, menyediakan implementasi *Decision Tree* berbasis data komentar primer TikTok yang dipadukan dengan tahapan pra-pemrosesan dan rekayasa fitur TF-IDF yang disesuaikan dengan karakteristik bahasa slang TikTok. Kedua, memberikan pembuktian nyata mengenai perngaruh optimasi *preprocessing* dan TF-IDF terhadap peningkatan nilai *recall* dan akurasi model dalam mengidentifikasi komentar *toxic*, serta menunjukkan efektivitas teknik pruning dalam menghasilkan struktur pohon keputusan yang lebih sederhana namun tetap akurat, sehingga dapat dijadikan acuan untuk pengembangan metode klasifikasi komentar *toxic* pada *platform* media sosial lainnya [11]. Dengan demikian, penelitian ini diharapkan mampu memberikan kontribusi terhadap pengembangan metode analisis sentimen pada komentar berpotensi *toxic* dengan pendekatan yang lebih sesuai dengan karakteristik linguistik TikTok, sekaligus menjadi dasar bagi penelitian lanjutan dalam pengembangan sistem moderasi komentar secara otomatis pada *platform* media sosial, penelitian ini juga tidak hanya berfokus pada penerapan algortima, tetapi juga memberikan pemahaman yang lebih mendalam mengenai pola linguistik toksisitas yang muncul dalam interaksi pengguna Tiktok.

2. METODE PENELITIAN

Data Komentar TikTok

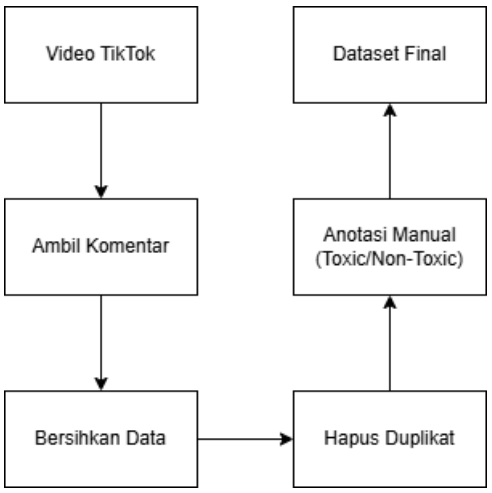
Penelitian ini menggunakan data primer berupa komentar yang diambil langsung dari *feed video* media sosial TikTok. *Dataset* yang digunakan mencakup total 271 komentar , yang telah diklasifikasikan ke dalam dua

kategori label: *Toxic* (*Label* = 1) dan *Non-Toxic* (*Label* = 0). Distribusi data menunjukkan bahwa komentar *toxic* berjumlah 136 dengan persentase 50,18% komentar per kategori dan komentar *non-toxic* berjumlah 135 dengan persentase 49,81% komentar per kategori.

Sumber *Dataset Primer*

Pengumpulan dataset pada penelitian ini dilakukan secara langsung dari komentar video TikTok dengan mengikuti alur proses sistematis seperti terlihat pada Gambar 1.

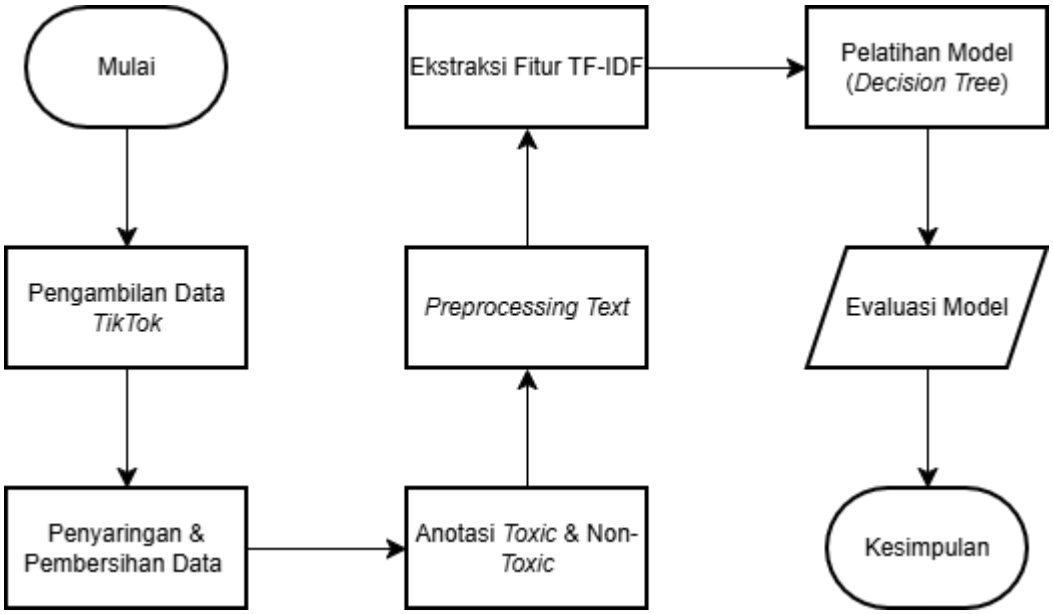
1. Pemilihan Video
   a. *Platform*: TikTok.
   b. Jenis Data: Komentar teks.
   c. Alamat Video:
      1) https://www.tiktok.com/@mahdi_abang/video/7563280041698741525?is_from_webapp=1&sender_device=pc&web_id=7570694531445376529
      2) https://www.tiktok.com/@feedgramindo/video/7560894770197875979?is_from_webapp=1&sender_device=pc&web_id=7570694531445376529
      3) https://www.tiktok.com/@sixtytwomedia/video/7248870892715953414?is_from_webapp=1&sender_device=pc&web_id=7570694531445376529
      4) https://www.tiktok.com/@puteriraniaa/video/7524607485861268754?is_from_webapp=1&sender_device=pc&web_id=7570694531445376529
      5) https://www.tiktok.com/@kompastv.indonesia/video/7564022073102765333?is_from_webapp=1&sender_device=pc&web_id=7570694531445376529
2. Pengambilan Komentar
   a. Tanggal: 10 – 11 – 2025 sampai 15 – 11 – 2025.
   b. Kondisi Data: Komentar yang tersedia pada saat pengambilan tanpa manipulasi atau perubahan konten peneliti.
3. Pembersihan Awal *(Initial Cleaning)*
   a. Menghapus komentar duplikat.
   b. Menghapus komentar bot atau promosi.
   c. Menghapus komentar non-teks yang tidak relevan.
4. Penggabungan dan Penyimpanan *Dataset*
   a. Semua komentar digabungkan dalam satu *file* excel
   b. Melakukan pengecekan ulang untuk memastikan data valid dan tidak rusak
5. Proses Anotasi Data (*Labeling Toxic* dan *Non-Toxic*)
   a. Kriteria Klasifikasi
      Komentar dikelompokkan menjadi dua kategori:
      1) Toxic
         a) Mengandung penghinaan, cacian, atau ujaran kebencian.
         b) Memicu provokasi atau menyerang individu atau kelompok.
         c) Mengandung *body shamming*, *hate speech*, merendahkan, atau *bullying.*
         d) Mengandung makian, kata kasar, atau konteks negatif secara terbuka.
      2) Non-Toxic
         a) Komentar netral, positif, atau tidak mengandung unsur urgensi.
         b) Berisi opini yang tidak menyerang pihak lain.
         c) Tidak mengandung kata kasar maupun negatif.
   b. Prosedur Anotasi
      1) Setiap komentar diekspor ke *file spreadsheet* (Excel).
      2) Memberi label *toxic* = 1 dan *non-toxic* = 0.
      3) Jika terdapat perbedaan label, diskusi akan dilakukan hingga mencapai hasil akhir.
      4) *Dataset final* disimpan dengan kolom:
         a) *Username* Komentar.
         b) Teks Komentar.
         c) Label (*Toxic*/*Non-Toxic*).



Gambar 1. Alur Proses Pengumpulan *Dataset Primer* TikTok

Tahapan Metodologi

Pada penelitian ini, proses analisis dilakukan melalui beberapa tahap utama yang meliputi pengumpulan data komentar TikTok, *preprocessing text*, ekstraksi fitur menggunakan TF-IDF, pelatihan model *Decision Tree*, serta evaluasi performa model. Secara keseluruhan, alur orkestrasi penelitian ditunjukkan pada Gambar 2 berikut.



Gambar 2. Alur Penelitian

Untuk mengatasi tantangan deteksi komentar *toxic* yang beragam, penelitian ini mengimplementasikan algoritma *Decision Tree* melalui serangkaian tahapan pemrosesan data dan rekayasa fitur sebagai berikut.

A. Pra-pemrosesan data (*text preprocessing*)

Tahap pra-pemrosesan data dilakukan untuk menormalisasi, membersihkan, dan menyederhanakan teks komentar TikTok sebelum masuk ke proses ekstraksi fitur dan klasifikasi [12]. Setiap langkah dijelaskan secara operasional agar proses dapat direplikasi oleh peneliti lain.

1. Normalisasi Bahasa *Slang TikTok*

   Normalisasi dilakukan untuk mengubah bahasa informal, singkatan, atau *slang TikTok* ke dalam bentuk bahasa Indonesia yang baku. Proses normalisasi dilakukan dengan membangun kamus *slang* baku yang disusun secara manual bedasarkan istilah yang sering muncul dalam *dataset,* seperti “anjir”, “gblk”, “wkwk”, “ngab”, “cringe” dan mengubahnya menjadi bentuk lebih sesuai [13]. Kamus ini dipadukan dengan referensi dari kamus bahasa gaul daring dan daftar *slang* populer TikTok.

2. Tokenisasi

   Tokenisasi bertujuan untuk memecah teks menjadi unit kata *(token)*. Tokenisasi dilakukan menggunakan fungsi *word_tokenize* dari *library Natural Languange Toolkit*, yang mampu memisahkan

kata bedasarkan spasi maupun tanda baca secara konsisten [14]. Tahap ini memungkinkan setiap kata diproses sebagai fitur tersendiri dalam tahap pembobotan TF-IDF.

3. Penyaringan *Stop Words*

   Stop Words adalah kata-kata umum yang tidak memiliki kontribusi besar terhadap proses klasifikasi, seperti kata sambung, kata ganti, atau ekspresi umum [15]. Penyaringan *stop words* dilakukan menggunakan daftar *stop words* bahasa Indonesia dari *Natural Languange Toolkit,* kemudian ditambah dengan *stop words* tambahan khas percakapan TikTok.

4. *Stemming*

   Stemming dilakukan untuk mengubah setiap kata menjadi bentuk dasarnya. Proses ini menggunakan *library* Sastrawi *Stemmer* [16], yang merupakan implementasi algoritma Nazief & Adriani untuk bahasa Indonesia. Sastrawi dipilih karena mampu mengurai imbuhan dengan baik pada kata kerja maupun kata sifat. Contoh *stemming*: "menghujat" → "hujat", "dimainkannya" → "main", "berteriak" → "teriak". *Stemming* penting karena kata-kata *toxic* sering muncul dalam berbagai bentuk turunan kata sehingga perlu disederhanakan sebelum diberi bobot TF-IDF.

   Pada penelitian ini, proses *lemmatization* tidak digunakan karena bahasa Indonesia tidak memiliki *lemmatizer* yang stabil untuk kasus *slang*, kata kasar, dan morfologi tidak baku. Sebagai gantinya, penelitian menggunakan *stemming* Sastrawi yang lebih efektif dalam mengubah bentuk kata berimbuhan menjadi akar kata.

B. Rekayasa fitur (*feature engineering*)

Setelah data bersih, teknik *Term Frequency-Inverse Document Frequency* (TF-IDF) akan diterapkan untuk melakukan pembobotan kata.

TF-IDF

Metode ini bekerja dengan mengutamakan kata-kata yang memiliki frekuensi kemunculan tinggi pada komentar *toxic*, tetapi jarang muncul pada keseluruhan dokumen. Pada tahap *Term Frequency* (TF), kata-kata *toxic* seperti "tolol", "dongo", "anjir" atau "jelek" sering muncul berulang pada komentar berlabel *toxic* sehingga nilai TF-nya tinggi. Sementara itu, pada tahap *Inverse Document Frequency* (IDF), kata-kata tersebut memiliki nilai IDF besar karena hanya muncul pada sebagian kecil dokumen dan hampir tidak pernah muncul pada komentar *non-toxic*. Ketika kedua komponen tersebut dikalikan (TF x IDF), diperoleh bobot TF-IDF yang sangat tinggi pada kata-kata *toxic*, sehingga kata-kata tersebut secara otomatis dianggap sebagai fitur yang paling informatif untuk proses klasifikasi oleh algoritma *Decision Tree*. Dengan demikian, TF-IDF secara efektif menonjolkan pola linguistik berbahaya yang mengindikasikan toksisitas pada komentar TikTok.

C. Klasifikasi menggunakan algoritma *decision tree*

Pada penelitian ini, algoritma *Decision Tree* yang digunakan adalah CART *(Classification and Regression Tree)* yang diimplementasikan melalui *library Scikit-Learn.* Pemilihan metode CART didasarkan pada penggunaan metrik *Gini Index* sebagai kriteria pemisahan, sesuai dengan standar *default Scikit-Learn* dan efektivitasnya dalam analisis sentimen berbasis fitur TF-IDF. CART dipilih karena mampu menangani fitur numerik yaitu hasil dari pembobotan TF-IDF, menghasilkan aturan klasifikasi yang mudah diintepretasikan, serta memiliki performa tinggi pada *dataset* berukuran kecil hingga sedang seperti komentar TikTok.

1. Pembentukan Pohon:

   Prosesnya melibatkan pemecahan *dataset* menjadi subset yang lebih kecil, dengan *Node* Internal menguji fitur kata tertentu, dan *Node* Daun menghasilkan kelas sentimen (*toxic* atau *non-toxic*). Pemisahan cabang (atribut/fitur) ditentukan berdasarkan perhitungan metrik kemurnian (*purity*) seperti *Gini Index* atau *Information Gain*.

2. *Hyperparameter Decision Tree*

   Penelitian ini menggunakan beberapa *hyperparameter* utama yang ditampilkan berserta nilai dan alasan pemilihannya sebagai berikut.

Tabel 1. Ringkasan *Hyperparameter* dan Kriteria Pemisahan Algoritma *Decision Tree*

No	*Hyperparameter*	Nilai	Alasan/Rasionalisasi
1	Criterion	Gini	Sesuai metode CART: cepat, stabil, dan efektif untuk *dataset* kecil
2	Max-depth	None (sebelumnya) diatur menjadi 10 setelah *tunning*	Pembatasan kedalaman mencegah overfitting pada dataset kecil
3	Min_samples_split	2	Nilai default: *dataset* relatif kecil sehingga pemisahan awal tetap diperlukan
4	Min_samples_leaf	1 diubah menjadi 2	Untuk mencegah pembentukan leaf yang terlalu kecil dan rentan *overfitting*

No	*Hyperparameter*	Nilai	Alasan/Rasionalisasi
5	Max_features	Sqrt	Mengurangi kompleksitas model dan meningkatkan generalisasi untuk data TF-IDF berdimensi besar
6	Random_state	42	Menjamin hasil yang konsisten dan dapat direplikasi

3. *Pruning* (Pemangkasan):
   Untuk mengatasi kecenderungan *overfitting* yang umum pada model *Decision Tree* tunggal, teknik *post-pruning* akan diterapkan. *Pruning* akan menghilangkan cabang-cabang yang memiliki dampak minimal pada akurasi, bertujuan untuk menyederhanakan pohon, meningkatkan generalisasi, dan memastikan identifikasi komentar *toxic* dilakukan secara andal.

D. Pengujian Komparatif Menggunakan *Dataset Iris*

Selain menggunakan dataset primer komentar TikTok, penelitian ini juga melakukan eksperimen tambahan menggunakan *Dataset Iris* sebagai pembanding. *Dataset Iris* merupakan *dataset* klasik dalam *machine learning* yang terdiri dari bunga *Iris* dengan tiga kelas target, yaitu *Iris Setosa*, *Iris Versicolor*, dan *Iris Virginica*. *Dataset Iris* digunakan untuk:

1. Menguji validitas dan konsistensi model *Decision Tree* dalam kondisi yang sudah idel.
2. Menunjukkan performa *Decision Tree* ketika digunakan pada *dataset* terstruktur dan bersih.
3. Menjadi pembanding terhadap hasil klasifikasi biner pada komentar TikTok yang bersifat tidak terstruktur dan penuh *noise*.

Tabel 2. Perbedaan Dengan Klasifikasi Biner Komentar TikTok

No	Aspek	Dataset Iris	Klasifikasi *Toxic/Non-Toxic*
1	Jenis Data	Numerik, terstruktur	Teks, tidak terstruktur
2	Jumlah Kelas	3 kelas	2 kelas
3	Kualitas Data	Bersih dan stabil	Banyak *noise*, *slang*, singkatan
4	Tantangan Utama	Separasi antar kelas	*Slang*, singkatan, konteks *toxic*
5	Alasan Penggunaan	Uji *baseline* performa model	Uji performa pada data *real world*

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan temuan-temuan utama dari penelitian analisis sentimen komentar *toxic* pada media sosial TikTok. Pembahasan diawali dengan analisis deskriptif terhadap karakteristik *dataset* dan pola bahasa yang teridentifikasi, dilanjutkan dengan evaluasi kinerja serta efektivitas algoritma *Decision Tree* yang telah dioptimalkan dengan pembobotan TF-IDF.

A. Analisis deskriptif *dataset* komentar tiktok

Dataset yang digunakan dalam penelitian ini terdiri dari 271 komentar dari media sosial TikTok, yang diklasifikasikan menjadi dua kategori utama.

Tabel 3. Distribusi Data Komentar *Toxic* dan *Non-Toxic*

No	Kategori Komentar	Label	Jumlah	Persentase
1	*Toxic*	1	136	50,18%
2	*Non Toxic*	0	135	49,81%

Hasil analisis menunjukkan bahwa komentar *toxic* dan *non-toxic* memiliki distribusi yang seimbang dalam *dataset*. Kategori *toxic* yang paling dominan ditemukan adalah Hinaan dan *Body Shaming*.

1. Pola Bahasa Komentar *Toxic*
   Setelah data komentar terkumpul, proses klasifikasi dilakukan secara manual dengan mengelompokkan komentar ke dalam dua kategori, yaitu komentar positif dan komentar negatif, bedasarkan konteks dan muatan bahasa yang digunakan.

Tabel 4. Pola dan Kata Kunci Utama Komentar *Toxic*

No	Jenis Toksisitas	Contoh Kata Kunci	Contoh Komentar
1	Hinaan Intelektual	Dongo, Tolol, Bodoh	“Kenyataan nya tolol ya😂”
2	Kata Kasar/Umpatan	Anjir,Anjg	“ikhlas g si anjg”
3	Body Shaming	Jelek, tua, norak	“rambut nya jelek bgt”

No	Jenis Toksisitas	Contoh Kata Kunci	Contoh Komentar
4	Merendahkan Niat	Pansos, haus validasi	“si haus validasi”
5	Diskriminasi	(Menyebut etnis tertentu)	“biasalah jawa banyak gaya”

2. Pola *Username Toxic*

Karakteristik pengguna yang cenderung melontarkan komentar *toxic* memiliki kecenderungan pola penggunaan username tertentu:

a. Menggunakan kata-kata negatif: *hater, toxic, burn, dark, negative.*
b. Menunjukkan niat buruk atau merendahkan: *troll*, anti, julid.

Sebaliknya, komentar Non-*Toxic* didominasi oleh Pujian, Apresiasi, dan Dukungan Moral. Kata kunci yang sering muncul adalah keren, bagus, semangat, kreatif, dan salut, disertai penggunaan emoji positif (misalnya: 😍, 🔥, 👏, ✨).

B. Implementasi *decision tree* dan rekayasa fitur

Dalam penelitian ini, Algoritma *Decision Tree* digunakan untuk klasifikasi, karena kemampuannya dalam memodelkan hasil dan menghasilkan model yang mudah diinterpretasikan. Namun, data komentar *real-time* TikTok yang padat dengan bahasa informal *(slang)* dan agresi siber (*cyber-aggression*) merupakan tantangan dalam mencapai akurasi optimal.

1. Tahap Pra-pemrosesan Data

Tahap pra-pemrosesan data difokuskan pada normalisasi bahasa slang TikTok dan penghilangan stop words yang tidak relevan. Normalisasi ini krusial karena kata-kata toxic sering muncul dalam bentuk tidak baku (misalnya, dongo, anjg, boros bnget), yang bila tidak dinormalisasi akan mengurangi efektivitas *Decision Tree* dalam membangun aturan klasifikasi yang akurat.

2. Pembobotan Fitur Menggunakan TF-IDF

Untuk mengatasi masalah akurasi yang rendah pada kategori minoritas (seperti sentimen negatif/*toxic*) yang sering terjadi pada Decision Tree tanpa penyesuaian fitur yang tepat , teknik *Term Frequency-Inverse Document Frequency* (TF-IDF) dioptimalkan untuk pembobotan kata.

a. Tujuan TF-IDF: Memberikan bobot lebih tinggi kepada kata-kata yang sering muncul dalam dokumen *toxic* tetapi jarang muncul di keseluruhan korpus (misalnya, tolol, anjir, *cringe*), sehingga menonjolkan fitur linguistik unik indikator toksisitas.

3. Pengujian Model dan Pencegahan *Overfitting*

Model *Decision Tree* rentan terhadap *overfitting*, di mana model bekerja sangat baik pada data pelatihan tetapi gagal pada data baru. Untuk memitigasi risiko ini, penelitian menerapkan teknik *post-pruning*.

a. *Post-pruning*: Memungkinkan pohon menjadi lebih sederhana dan kokoh dengan menghilangkan cabang-cabang yang memiliki dampak minimal terhadap akurasi, sehingga menyeimbangkan antara akurasi tinggi dan generalisasi yang baik untuk mengidentifikasi komentar *toxic* baru di TikTok secara efisien.

C. Hasil kinerja model (disediakan setelah pengujian)

Bagian ini menyajikan hasil dari dua eksperimen klasifikasi yang dilakukan menggunakan Algoritma *Decision Tree*. Kinerja model dievaluasi berdasarkan *Accuracy, Precision, Recall, F1-Score*, dan Matriks Kebingungan pada data pengujian.

1. Kinerja pada *Dataset* Iris

Eksperimen pertama menggunakan *Dataset* Iris, di mana 20% data digunakan untuk pengujian.

Tabel 5. Metrik Kinerja Model *Decision Tree* pada *Dataset* Iris

No	Metrik Evaluasi	Nilai	Keterangan
1	*Accuracy*	1.00	Tingkat prediksi benar secara keseluruhan.
2	*Precision*	1.00	Proporsi prediksi positif yang benar.
3	*Recall*	1.00	Kemampuan model menemukan semua kasus positif.
4	*F1-Score*	1.00	Ukuran keseimbangan *Precision* dan *Recall*

2. Kinerja pada Klasifikasi Biner (*Non-Toxic* dan *Toxic*)

Bedasarkan tabel 5, model mampu mengklasifikasikan komentar *toxic* dengan cukup baik, ditunjukkan oleh jumlah *True Positive* (TP) sebanyak 26 data. Nilai *accuracy* sebesar 0,75 menunjukkan bahwa sebagian besar data berhasil diprediksi dengan benar, meskipun masih terdapat kesalahan klasifikasi berupa *False Positive* (FP) dan *False Negative* (FN).

Kelebihan model terletak pada kemampuannya mendeteksi komentar *toxic* secara langsung. Namun, kelemahannya masih terdapat komentar *non-toxic* yang diprediksi sebagai toxic, serta

komentar *toxic* yang tidak terdeteksi, yang menunjukkan keterbatasan model dalam memahami konteks bahasa tertentu.

Hasil ini mengindikasikan bahwa model klasifikasi biner dapat digunakan sebagai penyaringan awal komentar *toxic* dan *non-toxic*.

Tabel 6. Matriks Kebingunan Klasifikasi Biner

No	Aktual/Prediksi	Prediksi *Non-Toxic*	Prediksi *Toxic*
1	Aktual *Non-Toxic*	19 *(True Negative)*	7 *(False Positive)*
2	Aktual *Toxic*	8 *(False Negative)*	26 *(True Positive)*

Dari matriks tersebut, dapat dilakukan perhitungan metrik evaluasi sebagai berikut:

a. *Accuracy*: Mengukur proporsi total prediksi yang benar.
   Accuracy = TP + TN / TP + TN + FP + FN = 26 + 19 / 26 + 19 + 7 + 8 = 45/60 = 0.75
b. *Precision* untuk kelas *Toxic*: Mengukur keakuratan model saat memprediksi positif.
   Precision = TP / TP + FP = 26 / 26 + 7 = 26/33 ≈ 0,787
c. *Recall* untuk kelas *Toxic*: Mengukur kemampuan model menemukan semua kasus positif yang sebenarnya.
   Recall = TP / TP + FN = 26 / 26 + 8 = 26/34 ≈ 0,765

4. SIMPULAN

Penelitian ini berhasil mengimplementasikan analisis sentimen untuk mengklasifikasikan komentar berpotensi *toxic* pada media sosial TikTok menggunakan algoritma *Decision Tree* yang diperkuat dengan *preprocessing* teks dan pembobotan fitur TF-IDF. *Dataset primer* yang terdiri dari 271 komentar berhasil dinormalisasi melalui proses pembersihan teks, normalisasi *slang*, dan penanganan variasi kata kasar, sehingga memberikan representasi fitur yang lebih akurat bagi model. Hasil pengujian menunjukkan bahwa model memperoleh akurasi sebesar 0,75, *precision* sebesar 0,787, dan *recall* sebesar 0,765 untuk kelas *toxic*, yang mengindikasikan performa yang cukup stabil dalam mendeteksi komentar *toxic* pada lingkungan bahasa informal TikTok. Penerapan teknik *post-pruning* juga terbukti membantu mengurangi *overfitting* serta meningkatkan kemampuan generalisasi model terhadap komentar baru. Dengan demikian, kombinasi *preprocessing slang*, pembobotan TF-IDF, dan *Decision Tree* dengan *pruning* menunjukkan efektivitas dalam mengidentifikasi komentar *toxic* berbasis data *primer* TikTok, meskipun penelitian selanjutnya dapat mempertimbangkan evaluasi efisiensi komputasi maupun pengujian pada *dataset* yang lebih besar untuk meningkatkan kinerja model.

DAFTAR PUSTAKA

[1] Rizki Misbah Hidayat, Abdul Rifki, and Ichsan Fauzi Rachman, "Dampak Paparan Konten Negatif di TikTok dan Instagram terhadap Kesehatan Mental Siswa: Kajian Literatur," *RISOMA : Jurnal Riset Sosial Humaniora dan Pendidikan.*, vol. 3, no. 3, pp. 136–143, 2025. doi: 10.62383/risoma.v3i3.769.

[2] R. G. Nugraha, K. Rahmani, H. Kiswantomo, D. N. Aliifah, and A. Rahma, "Hubungan antara Self-Control dan Toxic Disinhibition Online Effect pada Mahasiswa yang Menggunakan Sosial Media Berdasarkan Survei yang dilakukan Asosiasi Penyelenggara Jasa Internet Indonesia," *Humanitas Jurnal Psikologi*, vol. 7, no. 2, pp. 259–272, 2023. doi: 10.28932/humanitas.v7i2.5661.

[3] A. Fatkhudin, F. A. Artanto, N. A. Safli, and D. Wibowo, "Decision Tree Berbasis SMOTE Dalam Analisis Sentimen Penggunaan Artificial Intelligence Untuk Skripsi," *REMIK: Riset dan E-Jurnal Manajemen Informatika*, vol. 8, no. April, pp. 494–505, 2024. doi: 10.33395/remik.v8i2.13531.

[4] H. Azalia, Nailah and Voutama. Apriade, "Penerapan Social Media Analytics Dalam Decision Support System di TikTok Shop," *Journal of Information Systems And Informatics Engineering*, vol. 9, no. 1, pp. 167–176, 2025. doi: 10.35145/joisie.v9i1.4912.

[5] A. Gerliandeva, Y. Chrisnanto, and H. Ashaury, "Optimasi Klasifikasi Sentimen pada Komentar Online menggunakan Multinomial Naïve Bayes dan Ekstraksi Fitur TF-IDF serta N-grams," *Jurnal Sistem Informasi*, vol. 9, no. 2, pp. 260–272, 2024. doi: 10.30656/jsii.v11i2.9161.

[6] Larasati, and W. S. J. Saputra., "Implementasi Deteksi Kekerasan Dengan Peringatan Visual Menggunakan," *Jurnal Riset Teknik Komputer.*, vol. 2, no. 2, pp. 7–17, 2025. doi: 10.69714/38wan661.

[7] M. H. A. Sam, "Sentiment Analysis of Data Security in Indonesia Using Naive Bayes," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 12, no. 3, pp. 244-254, 2025. doi: 10.35957/jatisi.v12i3.12899.

[8] D. Ciang, "Klasifikasi Toksisitas Komentar Dengan Algoritma Naive Bayes dan Decision Tree," *Jurnal Komputer dan Informatika.*, vol. 18, no. 1, pp. 31–37, 2023. doi: 10.24912/jki.v18i1.34444.

[9] N. H. Inda Arifin and W. J. Shudiq, "Algoritma Decision Tree Dengan Menggunakan Pruning dan Missing Value Untuk Prediksi Kredit Macet," *COREAI (Jurnal Kecerdasan Buatan, Komputasi dan Teknologi Informasi).*, vol. 3, no. 1, pp. 38–45, 2022. doi: 10.33650/coreai.v3i1.4124.

[10] C. N. Syahputri and M. S. Hasibuan, "Optimasi Klasifikasi Decision Tree Dengan Teknik Pruning Untuk Mengurangi Overfitting," *JSiI (Jurnal Sist. Informasi)*, vol. 11, no. 2, pp. 87–96, 2024. doi: 10.30656/jsii.v11i2.9161.

[11] S. Azhari, N. Rahaningsih, R. D. Dana, K. Cirebon, A. Sentimen, and G. Play, "Peningkatan Akurasi Analisis Sentimen Pada Aplikasi Loklok Dengan Metode Naïve Bayes," *Jurnal Informatika Dan Teknik Elektro Terapan.*, vol. 13, no. 1, 2025. doi: 10.23960/jitet.v12i3.5848.

[12] L. Rhomaningtias, A. Khairunisa, S. Shella, M. Wara, and K. M. Hindrayani, "Analisis Sentimen Ulasan Aplikasi Smile Indonesia Menggunakan Metode Naive Bayes dan Support Vector Machine (SVM)," *Jurnal Teknolologi Informasi.*, vol. 16, no. 1, pp. 79–91, 2025. doi: 10.52972/hoaq.vol16no1.

[13] P. M. S. Ardinata, A. A. J. Permana, and I. N. S. W. Wijaya, "Identifikasi dan Normalisasi Teks Slang Dengan," *Jurnal Pendidikan Teknologi dan Kejuruan.*, vol. 21, no. 1, 2024. doi: 10.23887/jptkundiksha.v21i1.66381.

[14] L. A. Fitriana, "Analisis Ulasan Konsumen sebagai Data Non-Keuangan dalam Sistem Informasi Akuntansi," *Jurnal Profitabilitas.*, vol. 5, no. 1, pp. 64–74, 2025. doi: 10.31294/profitabilitas.v5i1.8269.

[15] A. Nurian and T. N. Padilah, "Disdukcapil Karawang Menggunakan Naive," *Jurnal Informatika Dan Teknik Elektro Terapan.*, vol. 12, no. 2, 2024. doi: 10.23960/jitet.v12i2.4178.

[16] A. Nafi, A. Tri, J. Harjanta, B. A. Herlambang, and S. Fahmi, "Analisis Sentimen Review Pelanggan Lazada dengan Sastrawi Stemmer dan SVM-PSO untuk Memahami Respon Pengguna," *Journal Information Technology.*, vol. 12, no. 204, pp. 330–339, 2024. doi: 10.32664/j-intech.v12i02.1450.