Jurnal Publikasi Sistem Informasi dan Manajemen Bisnis
Volume 5, Nomor 1 Januari 2026
e-ISSN: 2808-8980; p-ISSN: 2808-9383, Hal 709-730
DOI: https://doi.org/10.55606/jupsim.v5i1.6293
Tersedia: https://journalcenter.org/index.php/jupsim

Klasifikasi Sentimen Komentar Kanal Youtube @Rhenald Kasali Tentang Degradasi Lingkungan Berbasis Algoritma Decision Tree

Ign.F.Bayu Andoro.S
Institut Widya Pratama, Indonesia

**Penulis Korespondensi: bayu.andoro@iwima.ac.id*

Abstract*. Environmental degradation in Sumatra, driven by mining activities and land clearing, has sparked significant public discourse across various digital platforms. However, manual monitoring of massive public sentiment has proven inefficient for policy evaluation purposes. This study aims to classify public sentiment regarding environmental issues based on YouTube comments from a high-level interview involving a Ministry of Forestry advisor on the #IntrigueRK channel. A total of 708 comments were collected via the YouTube API and subjected to comprehensive preprocessing stages, including slang handling and Indonesian language stemming. Classification was performed using the Decision Tree algorithm, selected for its high interpretability in rule-based sentiment mapping. The model achieved an accuracy of 84.4%, with a Precision of 87.5%, a Recall of 89.4%, and an F1-score of 0.843. The research findings reveal a predominance of negative sentiment, reaching 472 comments (66.95%), while positive sentiment accounted for 233 comments (33.05%), primarily focusing on themes of government accountability and the impacts of flash floods. This study demonstrates that the Decision Tree algorithm is a robust instrument for social listening within the environmental sector. These results provide actionable insights for stakeholders to bridge the clinical gap between forestry policy and public perception.*

Keywords*: Decision Tree; Environmental Degradation; Sentiment Analysis; Social Listening; YouTube Comments.*

Abstrak. Degradasi lingkungan di Sumatra yang dipicu oleh aktivitas pertambangan dan pembukaan lahan (land clearing) telah memicu diskursus publik yang signifikan di berbagai platform digital. Namun, pemantauan manual terhadap sentimen publik yang masif terbukti tidak efisien untuk keperluan evaluasi kebijakan. Penelitian ini bertujuan untuk mengklasifikasikan sentimen publik mengenai isu lingkungan berdasarkan komentar YouTube pada wawancara tingkat tinggi yang melibatkan penasehat Kementerian Kehutanan di kanal #IntrigueRK. Sebanyak 708 komentar dikumpulkan melalui YouTube API dan melalui tahapan preprocessing, termasuk penanganan bahasa gaul (slang) serta stemming bahasa Indonesia. Klasifikasi dilakukan menggunakan algoritma *Decision Tree*, yang dipilih karena tingkat interpretabilitasnya yang tinggi dalam pemetaan sentimen berbasis aturan *(rule-based).* Model ini menghasilkan akurasi sebesar 84,4%, dengan nilai *precision* sebesar 87,5%, *recall* 89,4% dan *F1-Score* sebesar 0,843. Temuan penelitian mengungkapkan bahwa sentimen negatif mencapai 472 (66,95%) sedangkan sentimen positif mencapai 233 (33,05%) yang terutama berpusat pada tema akuntabilitas pemerintah dan dampak banjir bandang. Penelitian ini menunjukkan bahwa algoritma Decision Tree merupakan instrumen yang kuat untuk *social listening* di sektor lingkungan. Hasil penelitian ini memberikan wawasan aplikatif bagi para pemangku kepentingan untuk menjembatani kesenjangan antara kebijakan kehutanan dan persepsi publik.

Kata Kunci: Analisis Sentimen; Degradasi Lingkungan; Komentar YouTube; Mendengarkan Media Sosial; Pohon Keputusan.

1. LATAR BELAKANG

Perkembangan media sosial, khususnya YouTube, telah mengubah pola komunikasi publik dalam menyampaikan opini, kritik, dan persepsi terhadap berbagai isu strategis, termasuk permasalahan lingkungan (Suhendra & Selly Pratiwi, 2024). Video yang membahas degradasi lingkungan sering kali memicu diskusi intens di kolom komentar, yang merepresentasikan respons emosional, sikap, dan penilaian masyarakat secara spontan. Komentar-komentar tersebut merupakan sumber data tekstual yang kaya dan bernilai strategis

Naskah Masuk: 23 November 2025; Revisi: 18 Desember 2025; Diterima: 28 Januari 2026; Terbit: 30 Januari 2026

untuk memahami persepsi publik terhadap isu lingkungan, terutama dalam konteks perubahan penggunaan lahan, penggundulan vegetasi, dan gangguan sistem ekologis yang berkontribusi terhadap bencana lingkungan (Pratama & Mustofa, 2025). Namun, volume data komentar yang besar dan tidak terstruktur menjadikan analisis manual tidak lagi efektif, sehingga diperlukan pendekatan komputasional berbasis data mining dan pemrosesan bahasa alami (Mulyatun et al., 2021).

Sejumlah penelitian terdahulu telah mengkaji analisis sentimen pada media sosial untuk berbagai topik, seperti kebijakan publik, layanan digital, dan isu sosial, dengan memanfaatkan algoritma klasifikasi seperti *Naïve Bayes, Support Vector Machine, dan Deep Learning* (Novianti et al., 2024). Studi-studi tersebut menunjukkan bahwa analisis sentimen mampu mengungkap kecenderungan opini publik secara kuantitatif dan sistematis. Dalam konteks isu lingkungan, beberapa penelitian juga telah memanfaatkan data media sosial untuk mengidentifikasi sikap masyarakat terhadap perubahan iklim, deforestasi, dan pencemaran (Bakti, 2024). Namun demikian, sebagian besar penelitian masih berfokus pada performa model secara umum tanpa menekankan interpretabilitas hasil klasifikasi, padahal aspek keterjelasan model menjadi penting ketika hasil penelitian akan digunakan sebagai dasar pengambilan kebijakan atau rekomendasi strategis.

Di sisi lain, algoritma *Decision Tree* memiliki keunggulan dalam hal interpretabilitas dan kemudahan visualisasi aturan keputusan dibandingkan dengan model klasifikasi yang lebih kompleks (Yusuf et al., 2025). *Decision Tree* memungkinkan peneliti untuk menelusuri proses pengambilan keputusan model secara transparan, sehingga pola-pola linguistik yang memengaruhi klasifikasi sentimen dapat dipahami dengan lebih baik (Fatah & Atreji, 2025). Meskipun demikian, penerapan algoritma *Decision Tree* dalam analisis sentimen komentar YouTube yang secara spesifik membahas isu degradasi lingkungan masih relatif terbatas. Selain itu, kajian yang mengaitkan performa klasifikasi sentimen dengan konteks narasi lingkungan dalam satu video tertentu, sebagai representasi diskursus publik yang terfokus, masih jarang ditemukan dalam literatur.

Berdasarkan kondisi tersebut, terdapat celah penelitian yang perlu dijembatani, yaitu perlunya kajian analisis sentimen yang tidak hanya berorientasi pada akurasi model, tetapi juga menekankan keterbacaan dan pemaknaan hasil klasifikasi dalam konteks isu lingkungan. Urgensi penelitian ini terletak pada meningkatnya peran media sosial sebagai arena pembentukan opini publik terkait degradasi lingkungan, yang berpotensi memengaruhi kesadaran, sikap, dan dukungan masyarakat terhadap upaya mitigasi dan kebijakan lingkungan. Kebaruan penelitian ini terletak pada penerapan algoritma *Decision Tree* untuk

mengklasifikasikan sentimen komentar YouTube dengan URL https://www.youtube.com/watch?v=3iOl7xTv7ds dengan judul "Penasehat MENHUT BICARA: Dibalik Banjir Sumatra, Land Clearing Tambang & Sawit #IntrigueRK", pada satu konten video yang secara spesifik mengangkat isu degradasi lingkungan, sehingga memberikan pemahaman yang lebih kontekstual dan interpretatif terhadap respons publik.

Dengan demikian, tujuan penelitian ini adalah untuk mengklasifikasikan sentimen komentar pengguna YouTube terhadap isu degradasi lingkungan menggunakan algoritma *Decision Tree*, serta mengevaluasi kinerja model klasifikasi yang dihasilkan. Selain itu, penelitian ini bertujuan mengidentifikasi kecenderungan sentimen publik dan memberikan gambaran awal mengenai bagaimana masyarakat merespons isu degradasi lingkungan yang disampaikan melalui media sosial, sehingga hasil penelitian diharapkan dapat menjadi kontribusi ilmiah bagi pengembangan kajian analisis sentimen berbasis Decision Tree sekaligus menjadi masukan bagi pemangku kepentingan di bidang lingkungan dan komunikasi publik.

Berdasarkan permasalahan dan celah penelitian yang telah diuraikan, penelitian ini bertujuan untuk menganalisis sentimen publik terhadap isu degradasi lingkungan berdasarkan komentar pengguna YouTube menggunakan pendekatan kuantitatif berbasis pembelajaran mesin. Secara khusus, penelitian ini bertujuan untuk membangun model klasifikasi sentimen komentar YouTube menggunakan algoritma *Decision Tree* berbasis representasi fitur TF-IDF, mengevaluasi kinerja model klasifikasi menggunakan metrik akurasi, precision, recall, dan F1-score, serta mengidentifikasi kecenderungan sentimen publik terhadap isu degradasi lingkungan yang disampaikan melalui konten video diskusi lingkungan di media sosial.

Berdasarkan kerangka teoritis analisis sentimen, konsep komunikasi lingkungan digital, serta karakteristik algoritma *Decision Tree* yang memiliki tingkat interpretabilitas tinggi, maka hipotesis penelitian ini dirumuskan sebagai berikut:

H1: Sentimen negatif mendominasi komentar publik terhadap isu degradasi lingkungan yang dibahas dalam konten video YouTube.

H2 : Algoritma *Decision Tree* mampu mengklasifikasikan sentimen komentar YouTube terkait isu degradasi lingkungan dengan tingkat akurasi yang tinggi.

H3: Model klasifikasi sentimen berbasis *Decision Tree* memiliki nilai *precision*, *recall*, dan F1-score yang baik dalam membedakan sentimen positif dan negatif.

2. KAJIAN TEORITIS

Tinjauan Pustaka

Transformasi digital telah mengubah secara fundamental cara isu-isu lingkungan dikomunikasikan, beralih dari model diseminasi satu arah oleh media tradisional menuju dialog multi-arah yang sangat partisipatif di berbagai platform media sosial (Dudy et al., 2022). Perubahan paradigma ini meruntuhkan batasan antara otoritas ilmiah dan masyarakat awam, menciptakan ruang publik digital yang inklusif bagi pertukaran gagasan ekologi secara *real-time* (Zainu Ridlo et al., 2024). Dalam konteks ini, komunikasi lingkungan di era digital tidak lagi dipandang sekadar sebagai penyebaran informasi ilmiah atau data teknis, melainkan partisipasi masyarakat luas memiliki agensi untuk memberikan respon langsung, kritik, maupun aspirasi terhadap kebijakan ekologi (Hermanto, 2023). Platform seperti YouTube kini berfungsi sebagai arena dialektika di mana narasi kebijakan dan praktik eksploitasi sumber daya alam diuji oleh persepsi kolektif penonton digital (Armandha & Fauziah, 2020).

Era digital memungkinkan munculnya fenomena *digital environmentalism*, di mana platform seperti YouTube bertindak sebagai "ruang publik digital" *(digital public sphere)* (De Putri & Toni, 2024). Dalam konteks video diskusi mengenai banjir Sumatra dan *land clearing* tambang, komunikasi lingkungan tidak lagi didominasi oleh pernyataan resmi pemerintah (Annisarizki & Surahman, 2022), tetapi juga diuji oleh pengalaman langsung masyarakat (Sitorus, 2021). Hal ini menciptakan fungsi pengawasan *(watchdog)* yang terdesentralisasi (Aulia & Ridwan Maksun, 2022), di mana setiap komentar penonton merupakan bentuk partisipasi masyarakat (Rifqi & Gusti Aji, 2023).

Implementasi analisis sentimen memungkinkan skalabilitas pengolahan data yang masif dengan intervensi manual minimal, di mana pemanfaatan kemajuan algoritma dan perangkat lunak khusus memastikan transformasi data tekstual menjadi informasi terstruktur secara otomatis dan presisi (Made et al., 2023). Sebagai teknik klasifikasi berbasis pembelajaran mesin, *Decision Tree* mengadopsi struktur hierarkis yang menyerupai pohon, di mana setiap simpul *(node)* internal merepresentasikan pengujian pada atribut, setiap cabang melambangkan nilai hasil pengujian tersebut, dan simpul daun *(leaf)* berfungsi sebagai penentu kategori atau label kelas akhir (Khatib Sulaiman et al., 2024). Struktur hierarki algoritma ini dimulai dari *root node* atau simpul akar yang menempati posisi puncak, bertindak sebagai titik pusat pembagi yang memproses input tunggal dan menghasilkan dua cabang keluaran berdasarkan hasil pengujian fitur tertentu (Wiratama Putra & Triayudi, 2022). Struktur pohon keputusan merepresentasikan kriteria evaluasi atau kondisi logis pada setiap percabangannya, sementara bagian ujung atau simpul daun *(leaf nodes)* berfungsi sebagai representasi nilai kelas final dari

data yang diklasifikasikan (Fatah & Maghfiro, 2025). Arsitektur pohon keputusan mengintegrasikan simpul internal *(internal nodes)* untuk mengevaluasi variabel input atau atribut melalui mekanisme partisi data ke dalam subset yang lebih homogen, sementara simpul daun *(leaf nodes)* berperan dalam menetapkan label kelas spesifik bagi setiap observasi pada segmen yang terbentuk (Nazifah et al., 2023).

Pelabelan Sentimen *(Lexicon-Based)*

Pelabelan dilakukan menggunakan metode *Lexicon-Based*, yaitu pendekatan yang menentukan polaritas sentimen berdasarkan kecocokan token dengan kamus kata positif dan negatif. Secara umum, setiap dokumen ddd diberi skor sentimen berdasarkan akumulasi bobot kata-kata polar:

$$Score(d) = \sum_{w \epsilon d} S\,(w)$$

Dengan s(w) adalah bobot polaritas kata w (positif bernilai >0, negatif bernilai <0, dan 0 bila tidak ada di kamus). Prinsip ini sejalan dengan praktik leksikon berbobot pada analisis sentimen. Untuk konteks Bahasa Indonesia, leksikon dapat merujuk pada sumber leksikon sentimen Indonesia (mis. InSet) bila digunakan dalam penelitian. Aturan penetapan label ditetapkan sebagai berikut:

$$label\ (d) = \begin{cases} Positif, & Score\ (d) > 0 \\ Negatif, & Score\ (d) < 0 \\ Netral, & Score\ (d) = 0 \end{cases}$$

Selain itu, aturan negasi dapat diterapkan untuk meningkatkan akurasi pelabelan, misalnya jika token negasi (“tidak”, “bukan”, “jangan”, “tanpa”) muncul sebelum kata polar dalam jendela tertentu, maka polaritas kata tersebut dibalik. Hasil tahap ini adalah dataset yang telah memiliki label kelas sentimen sebagai target untuk pelatihan model klasifikasi.

Pembentukan Fitur (Representasi Teks)

Agar teks dapat diproses oleh algoritma klasifikasi, dokumen hasil prapemrosesan direpresentasikan menjadi vektor numerik menggunakan pendekatan pembobotan berbasis term seperti TF-IDF. Bobot TF-IDF untuk term ttt pada dokumen ddd dirumuskan sebagai:

$$tfidf(t,d) = tf(t,d)\ \mathrm{x}\ \log\left(\frac{N}{df(t)}\right)$$

Dengan tf(t,d) adalah frekuensi term pada dokumen, df(t) adalah jumlah dokumen yang memuat term, dan N adalah total dokumen. Representasi ini membantu menurunkan pengaruh kata yang terlalu umum dan meningkatkan kontribusi term yang lebih diskriminatif terhadap kelas.

Pembagian Dataset *(Train–Test Split)*

Dataset berlabel dibagi menjadi data latih dan data uji untuk mengukur kemampuan generalisasi model. Dalam implementasi umum, pembagian dilakukan menggunakan rasio 80%:20% (atau sesuai rancangan eksperimen), serta disarankan menggunakan pembagian terstratifikasi bila distribusi kelas tidak seimbang agar proporsi kelas pada train dan test tetap representatif.

Klasifikasi Menggunakan *Decision Tree*

Tahap klasifikasi menggunakan *Decision Tree*, yaitu model yang membentuk aturan keputusan berbasis pemilihan atribut paling informatif melalui proses pemisahan rekursif. Secara praktik, pemilihan split dapat menggunakan ukuran impuritas seperti *entropy (Information Gain)* atau Gini. Untuk *entropy*, ukuran ketidakpastian pada himpunan data SSS dapat dituliskan sebagai:

$$H(S) = -\sum_{i=1}^{k} pi\ log2(pi)$$

dengan pip_ipi adalah proporsi kelas ke-i. Model dilatih pada data latih untuk memetakan fitur TF-IDF ke label sentimen yang dihasilkan dari pelabelan *lexicon-based.*

Evaluasi Model

Kinerja model dievaluasi pada data uji menggunakan *confusion matrix* dan metrik klasifikasi umum: *Accuracy, Precision, Recall*, dan *F1-Score*. Definisi metrik biner (dapat diperluas ke multi-kelas dengan *macro/micro averaging*):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2.\frac{Precision\ Recall}{Precision + Recall}$$

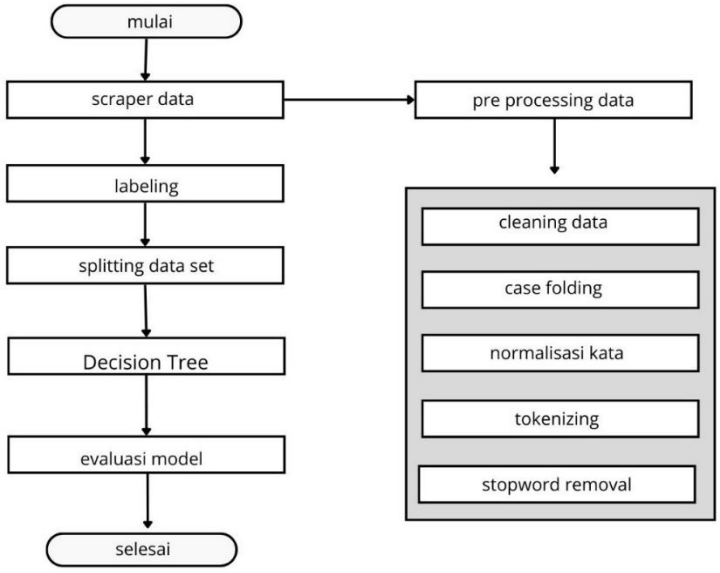
Metrik-metrik ini digunakan untuk menilai ketepatan model secara keseluruhan serta keseimbangan kemampuan model dalam mendeteksi kelas positif maupun negatif.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis sentimen berbasis pembelajaran mesin karena tujuan penelitian berfokus pada pengukuran, pengklasifikasian, dan evaluasi sentimen publik secara objektif dan terukur. Pendekatan kuantitatif dipilih untuk memungkinkan pengolahan data komentar dalam jumlah besar secara sistematis serta menghasilkan indikator kinerja model yang dapat dievaluasi secara empiris melalui metrik statistik. Algoritma *Decision Tree* dipilih sebagai metode klasifikasi utama karena memiliki tingkat interpretabilitas yang tinggi dibandingkan dengan algoritma klasifikasi lain seperti *Support Vector Machine* atau model berbasis *deep learning*. Interpretabilitas ini menjadi aspek krusial dalam konteks analisis sentimen isu degradasi lingkungan, di mana hasil klasifikasi tidak hanya dituntut akurat, tetapi juga dapat dijelaskan secara logis dan transparan kepada pemangku kepentingan. Struktur pohon keputusan memungkinkan peneliti untuk menelusuri aturan keputusan yang terbentuk dari fitur linguistik, sehingga pola-pola kata yang memengaruhi sentimen positif maupun negatif dapat dianalisis secara lebih mendalam.

Metode pelabelan sentimen berbasis leksikon *(lexicon-based)* digunakan sebagai tahap awal untuk membentuk data berlabel *(ground truth)* karena pendekatan ini relatif efisien dan sesuai untuk dataset berbahasa Indonesia yang belum sepenuhnya memiliki label manual. Pendekatan ini juga memungkinkan konsistensi pelabelan pada skala data yang besar dengan intervensi manual minimal. Selanjutnya, representasi teks menggunakan pembobotan TF-IDF dipilih karena kemampuannya dalam menekan pengaruh kata-kata umum serta menonjolkan term yang memiliki daya diskriminatif tinggi terhadap kelas sentimen. Kombinasi antara pelabelan *lexicon-based*, representasi fitur TF-IDF, dan algoritma *Decision Tree* membentuk kerangka metodologis yang seimbang antara efisiensi komputasi, keterbacaan model, dan akurasi klasifikasi. Oleh karena itu, pemilihan metode dalam penelitian ini dinilai relevan dan tepat untuk menjawab tujuan penelitian serta menguji hipotesis yang telah dirumuskan.

Penelitian ini menerapkan alur pipeline analisis sentimen yang mencakup: pengumpulan data *(scraping)*, prapemrosesan teks, pelabelan otomatis berbasis leksikon *(lexicon-based)*, pembagian dataset, klasifikasi menggunakan *Decision Tree*, dan evaluasi kinerja model. Pendekatan ini memadukan pelabelan berbasis kamus untuk membentuk *ground truth* awal, kemudian memanfaatkan *supervised learning* untuk mempelajari pola sentimen dari fitur teks.



Gambar 1. Alur Penelitian.

Akuisisi Data (Scraping)

Penelitian ini dilaksanakan melalui beberapa tahapan yang tersusun secara sistematis sebagaimana ditunjukkan pada diagram alur penelitian. Setiap tahapan dirancang untuk memastikan proses analisis sentimen berjalan secara terstruktur, transparan, dan dapat direplikasi.

Pengambilan data (Scraping Data)

Pada tahap ini, data berupa komentar pengguna dikumpulkan dari platform YouTube pada video yang menjadi objek kajian. Proses scraping dilakukan secara otomatis menggunakan pustaka pemrograman Python untuk memperoleh data teks mentah *(raw data)* dalam jumlah besar. Data yang diperoleh pada tahap ini masih bersifat tidak terstruktur dan mengandung berbagai *noise* seperti simbol, emoji, variasi ejaan, serta karakter non-alfabet.

Prapemrosesan Data Teks

Tahap prapemrosesan bertujuan menstandarkan dan meningkatkan kualitas data teks agar lebih representatif untuk proses pelabelan dan pembentukan fitur. Sesuai diagram, prapemrosesan meliputi:

a. *Cleaning*: menghapus elemen non-informatif (URL, mention, angka, tanda baca berlebih, karakter khusus, dan spasi ganda).
b. *Case folding*: mengubah seluruh karakter menjadi huruf kecil untuk menghindari perbedaan token akibat kapitalisasi.
c. Normalisasi kata: mengonversi kata tidak baku/slang menjadi bentuk baku dan menyederhanakan repetisi karakter (mis. “baaguuusss” → “bagus”).

d. *Tokenizing*: memecah teks menjadi token kata untuk memudahkan perhitungan skor leksikon dan pembobotan fitur.
e. *Stopword removal*: menghapus kata umum berfrekuensi tinggi yang kontribusi semantiknya rendah (mis. “yang”, “dan”, “di”), dengan catatan kata negasi seperti “tidak/bukan” sebaiknya dipertahankan bila digunakan dalam aturan pembalikan polaritas.

Keluaran tahap ini adalah korpus teks yang lebih bersih, konsisten, dan siap diproses pada tahap pelabelan.

Pelabelan data *(Labeling)*

Pada tahap ini, data teks yang telah melalui prapemrosesan diberi label sentimen menggunakan pendekatan *lexicon-based*. Pelabelan dilakukan dengan menghitung skor polaritas berdasarkan kecocokan kata-kata dalam komentar dengan kamus kata positif dan negatif. Hasil dari tahap ini adalah dataset berlabel yang merepresentasikan sentimen positif dan negatif, yang selanjutnya digunakan sebagai data latih dan data uji dalam proses klasifikasi.

Pembagian dataset (Splitting Data Set)

Pembagian ini dilakukan untuk memisahkan data menjadi data latih dan data uji dengan proporsi tertentu. Data latih digunakan untuk membangun model klasifikasi, sedangkan data uji digunakan untuk mengevaluasi kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.

Tahap selanjutnya adalah proses klasifikasi menggunakan algoritma *Decision Tree*. Pada tahap ini, model *Decision Tree* dibangun berdasarkan data latih dengan memanfaatkan fitur teks yang telah direpresentasikan dalam bentuk numerik. Algoritma ini membentuk struktur pohon keputusan yang terdiri atas simpul dan cabang, di mana setiap simpul merepresentasikan atribut atau fitur yang digunakan untuk memisahkan data berdasarkan kriteria tertentu hingga mencapai simpul daun yang menentukan kelas sentimen akhir.

Evaluasi model

Evaluasi dilakukan menggunakan data uji untuk mengukur kinerja model klasifikasi yang telah dibangun. Metrik evaluasi yang digunakan meliputi *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil evaluasi ini digunakan untuk menilai sejauh mana algoritma *Decision Tree* mampu mengklasifikasikan sentimen komentar secara akurat dan seimbang.

Seluruh rangkaian tahapan tersebut membentuk alur penelitian yang berkesinambungan, dimulai dari pengumpulan data hingga evaluasi hasil, sehingga penelitian dapat diselesaikan secara sistematis dan menghasilkan temuan yang dapat dipertanggungjawabkan secara ilmiah.

Data teks dikumpulkan menggunakan teknik *web scraping* dari sumber yang relevan dengan objek kajian (misalnya komentar, ulasan, atau opini pada platform daring). Tahap ini menghasilkan *raw dataset* yang umumnya masih mengandung *noise* seperti URL, emoji, variasi ejaan, dan simbol, sehingga memerlukan prapemrosesan sebelum dianalisis lebih lanjut.

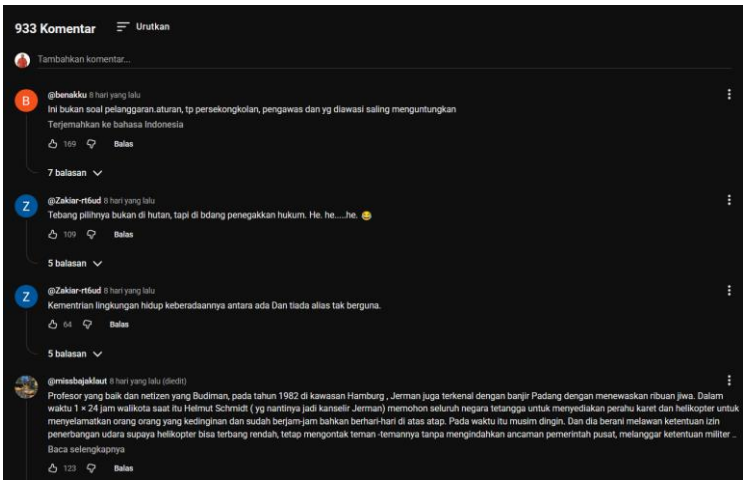
4. HASIL DAN PEMBAHASAN

Hasil dan Pembahasan pada penelitian dibuat dalam beberapa tahana yakni

Pengumpulan Data

Dataset yang dipakai dalam penelitian ini adalah komentar di video kanal Youtube @Rhenald Kasali

Gambar 2. Video Youtube .



Gambar 3. Komentar di Video Youtube.

Teknik yang digunakan untuk menggambil data yaitu *scraping* data memakai library Python yaitu *google-play-scraper*.

Gambar 4. Alur Proses Scraping Data.

Berikut kode program proses scraping data menggunakan google colab ditunjukan pada gambar 5.

```
youtube = build(
    "youtube",
    "v3",
    developerKey=API_KEY
)

comments = []

request = youtube.commentThreads().list(
    part="snippet",
    videoId=VIDEO_ID,
    maxResults=100,
    textFormat="plainText"
)

while request:
    response = request.execute()

    for item in response["items"]:
        comment = item["snippet"]["topLevelComment"]["snippet"]
        comments.append({
            "author": comment["authorDisplayName"],
            "comment": comment["textDisplay"],
            "like_count": comment["likeCount"],
            "published_at": comment["publishedAt"]
        })

    request = youtube.commentThreads().list_next(request, response)
```

Gambar 5. Program *Scraping* Data.

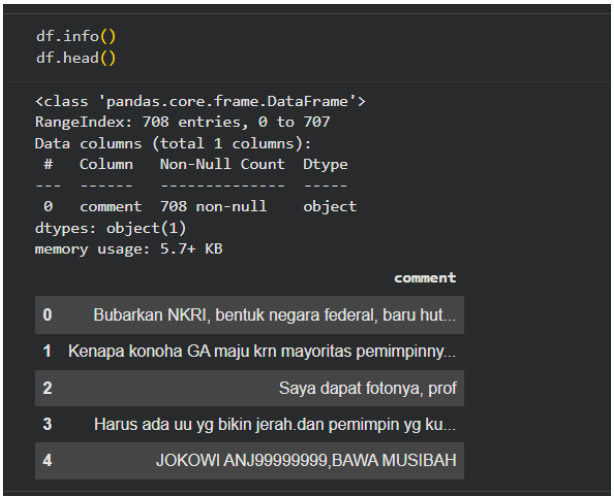
Preprocessing Data

Pada tahapan preprocessing dataset yang diperoleh kemudian diproses agar dataset dibersihkan dari kalimat atau kata yang tidak diperlukan pada penelitian ini, data yang kurang lengkap, data yang tidak valid, serta atribut yang kurang ditampilkan pada gambar 6.

```
df=pd.DataFrame(df[["comment"]])
df.head(5)
```

	comment
0	Bubarkan NKRI, bentuk negara federal, baru hut...
1	Kenapa konoha GA maju krn mayoritas pemimpinny...
2	Saya dapat fotonya, prof
3	Harus ada uu yg bikin jerah.dan pemimpin yg ku...
4	JOKOWI ANJ99999999,BAWA MUSIBAH

Gambar 6. *Preprocessing* Data.

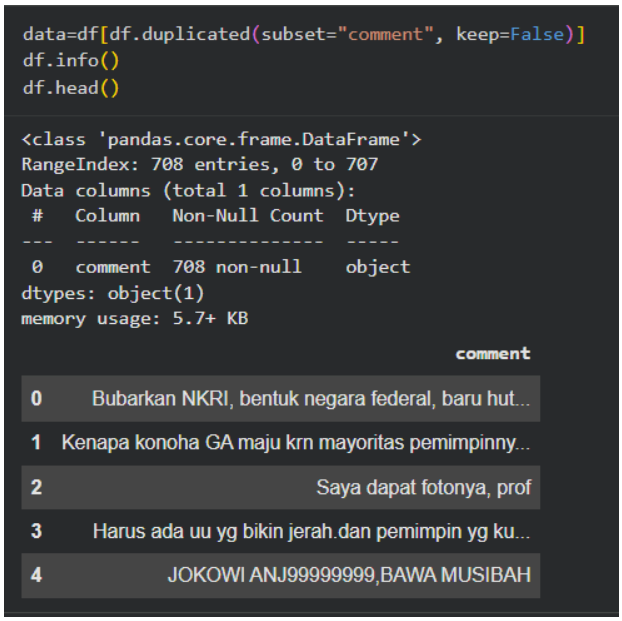


```
df.info()
df.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 708 entries, 0 to 707
Data columns (total 1 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   comment  708 non-null    object
dtypes: object(1)
memory usage: 5.7+ KB
```

	comment
0	Bubarkan NKRI, bentuk negara federal, baru hut...
1	Kenapa konoha GA maju krn mayoritas pemimpinny...
2	Saya dapat fotonya, prof
3	Harus ada uu yg bikin jerah.dan pemimpin yg ku...
4	JOKOWI ANJ99999999,BAWA MUSIBAH

Gambar 7. Proses Menghapus Data Duplikat.

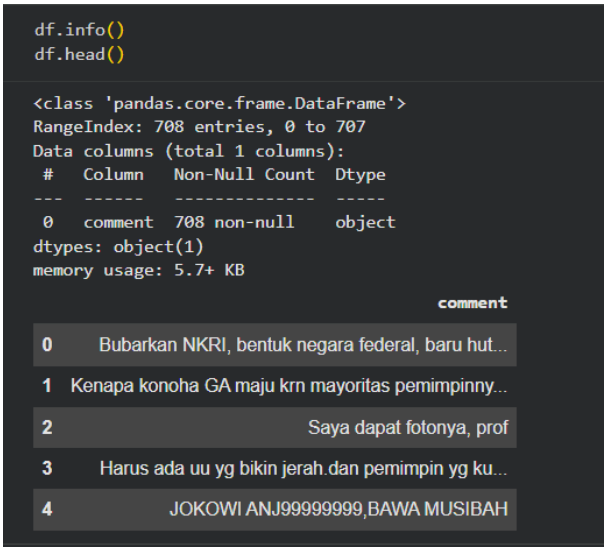


```
data=df[df.duplicated(subset="comment", keep=False)]
df.info()
df.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 708 entries, 0 to 707
Data columns (total 1 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   comment  708 non-null    object
dtypes: object(1)
memory usage: 5.7+ KB
```

	comment
0	Bubarkan NKRI, bentuk negara federal, baru hut...
1	Kenapa konoha GA maju krn mayoritas pemimpinny...
2	Saya dapat fotonya, prof
3	Harus ada uu yg bikin jerah.dan pemimpin yg ku...
4	JOKOWI ANJ99999999,BAWA MUSIBAH

Gambar 8. Proses Menghapus Data.



```
df.info()
df.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 708 entries, 0 to 707
Data columns (total 1 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   comment  708 non-null    object
dtypes: object(1)
memory usage: 5.7+ KB
```

	comment
0	Bubarkan NKRI, bentuk negara federal, baru hut...
1	Kenapa konoha GA maju krn mayoritas pemimpinny...
2	Saya dapat fotonya, prof
3	Harus ada uu yg bikin jerah.dan pemimpin yg ku...
4	JOKOWI ANJ99999999,BAWA MUSIBAH

Gambar 9. Proses Menghapus Data.

a. Tahap cleaning yaitu data yang dibersihkan dari elemen-elemen yang kurang relevan yaitu menghapus simbol, tanda baca, emoji, angka, dan elemen lain yang tidak diperlukan ditampilkan pada gambar 10.

```
import re
import string
import nltk

# Fungsi untuk menghapus URL
def remove_URL(tweet):
    if tweet is not None and isinstance(tweet, str):
        url = re.compile(r'https?://\S+|www\.\S+')
        return url.sub(r'', tweet)
    else:
        return tweet

# Fungsi untuk menghapus HTML
def remove_html(tweet):
    if tweet is not None and isinstance(tweet, str):
        html = re.compile(r'<.*?>')
        return html.sub(r'', tweet)
    else:
        return tweet

# Fungsi untuk menghapus emoji
def remove_emoji(tweet):
    if tweet is not None and isinstance(tweet, str):
        emoji_pattern = re.compile("["
            u"\U0001F600-\U0001F64F"  # emoticons
            u"\U0001F300-\U0001F5FF"  # symbols & pictographs
            u"\U0001F680-\U0001F6FF"  # transport & map symbols
            u"\U0001F700-\U0001F77F"  # alchemical symbols
            u"\U0001F780-\U0001F7FF"  # Geometric Shapes Extended
            u"\U0001F800-\U0001F8FF"  # Supplemental Arrows-C
            u"\U0001F900-\U0001F9FF"  # Supplemental Symbols and Pictographs
            u"\U0001FA00-\U0001FA6F"  # Chess Symbols
            u"\U0001FA70-\U0001FAFF"  # Symbols and Pictographs Extended-A
            u"\U0001F004-\U0001F0CF"  # Additional emoticons
```

Gambar 10. Proses *Cleaning.*

	comment	cleaning
0	Bubarkan NKRI, bentuk negara federal, baru hut...	Bubarkan NKRI bentuk negara federal baru hutan...
1	Kenapa konoha GA maju krn mayoritas pemimpinny...	Kenapa konoha GA maju krn mayoritas pemimpinny...
2	Saya dapat fotonya, prof	Saya dapat fotonya prof
3	Harus ada uu yg bikin jerah.dan pemimpin yg ku...	Harus ada uu yg bikin jerahdan pemimpin yg kua...
4	JOKOWI ANJ99999999,BAWA MUSIBAH	JOKOWI ANJBAWA MUSIBAH

Gambar 11. Hasil *Cleaning.*

b. *Case Folding*

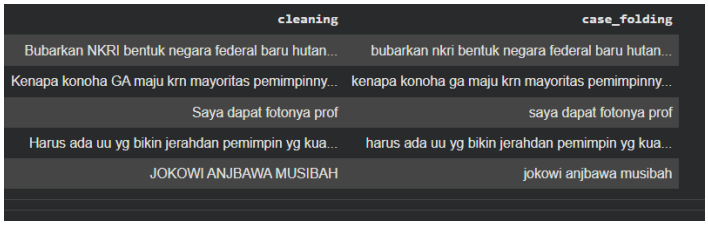
Pada tahap case folding semua huruf kapital diubah menjadi huruf kecil, tahap case folding ditunjukan pada gambar 12.

cleaning	case_folding
Bubarkan NKRI bentuk negara federal baru hutan...	bubarkan nkri bentuk negara federal baru hutan...
Kenapa konoha GA maju krn mayoritas pemimpinny...	kenapa konoha ga maju krn mayoritas pemimpinny...
Saya dapat fotonya prof	saya dapat fotonya prof
Harus ada uu yg bikin jerahdan pemimpin yg kua...	harus ada uu yg bikin jerahdan pemimpin yg kua...
JOKOWI ANJBAWA MUSIBAH	jokowi anjbawa musibah

Gambar 12. Hasil *Case Folding.*

c. Normalisasi

Pada tahap normalisasi yaitu mengubah kata tidak baku menjadi kata baku berdasarkan kamus baku yang telah disediakan, tahap normalisasi ditunjukan pada gambar 12.

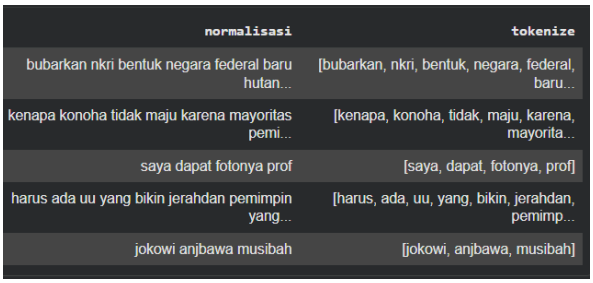


	cleaning	case_folding
	Bubarkan NKRI bentuk negara federal baru hutan...	bubarkan nkri bentuk negara federal baru hutan...
	Kenapa konoha GA maju krn mayoritas pemimpinny...	kenapa konoha ga maju krn mayoritas pemimpinny...
	Saya dapat fotonya prof	saya dapat fotonya prof
	Harus ada uu yg bikin jerahdan pemimpin yg kua...	harus ada uu yg bikin jerahdan pemimpin yg kua...
	JOKOWI ANJBAWA MUSIBAH	jokowi anjbawa musibah

Gambar 13. Hasil Normalisasi.

d. Tokenization

Pada tahap tokenization memecah kalimat menjadi kata, tahap tokenization ditunjukan pada gambar 13.

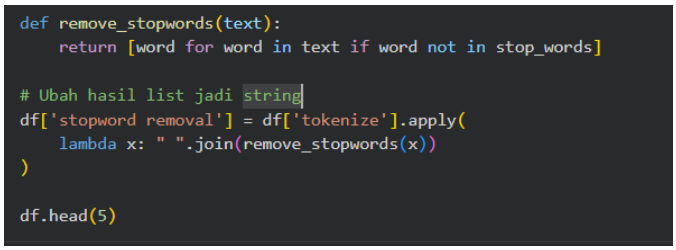


	normalisasi	tokenize
	bubarkan nkri bentuk negara federal baru hutan...	[bubarkan, nkri, bentuk, negara, federal, baru...
	kenapa konoha tidak maju karena mayoritas pemi...	[kenapa, konoha, tidak, maju, karena, mayorita...
	saya dapat fotonya prof	[saya, dapat, fotonya, prof]
	harus ada uu yang bikin jerahdan pemimpin yang...	[harus, ada, uu, yang, bikin, jerahdan, pemimp...
	jokowi anjbawa musibah	[jokowi, anjbawa, musibah]

Gambar 14. Hasil Tokenization.

e. *Stopword Removal*

Pada tahap *Stopword Removal* yaitu menghapus kata-kata yang tidak diperlukan dalam proses *preprocessing*, contoh kata yaitu yang, di, ke dan masih banyak lagi kata terdapat dalam kamus sastrawi yang disediakan. Tahap stopword removal ditunjukan pada gambar 14 dan 15.

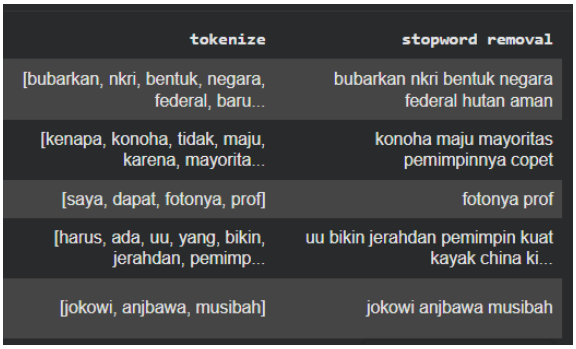


```
def remove_stopwords(text):
    return [word for word in text if word not in stop_words]

# Ubah hasil list jadi string
df['stopword removal'] = df['tokenize'].apply(
    lambda x: " ".join(remove_stopwords(x))
)

df.head(5)
```

Gambar 15. Proses *Stopword Removal*.



	tokenize	stopword removal
	[bubarkan, nkri, bentuk, negara, federal, baru...	bubarkan nkri bentuk negara federal hutan aman
	[kenapa, konoha, tidak, maju, karena, mayorita...	konoha maju mayoritas pemimpinnya copet
	[saya, dapat, fotonya, prof]	fotonya prof
	[harus, ada, uu, yang, bikin, jerahdan, pemimp...	uu bikin jerahdan pemimpin kuat kayak china ki...
	[jokowi, anjbawa, musibah]	jokowi anjbawa musibah

Gambar 16. Hasil *Stopword Removal*.

f. Pelabelan Data

Pelabelan kelas sentimen dilakukan menggunakan metode *lexicon-based*, yaitu pendekatan yang menentukan polaritas sentimen berdasarkan kecocokan kata pada teks terhadap kamus kata positif dan negatif.

```
import pandas as pd

data = pd.read_csv("Hasil_Preprocessing_Data_Renald.csv")
data.info()
data.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 708 entries, 0 to 707
Data columns (total 6 columns):
 #   Column           Non-Null Count  Dtype 
---  ------           --------------  ----- 
 0   comment          708 non-null    object
 1   cleaning         705 non-null    object
 2   case_folding     705 non-null    object
 3   normalisasi      705 non-null    object
 4   tokenize         708 non-null    object
 5   stopword removal  705 non-null    object
dtypes: object(6)
memory usage: 33.3+ KB
```

Gambar 17. Proses Pelabelan.

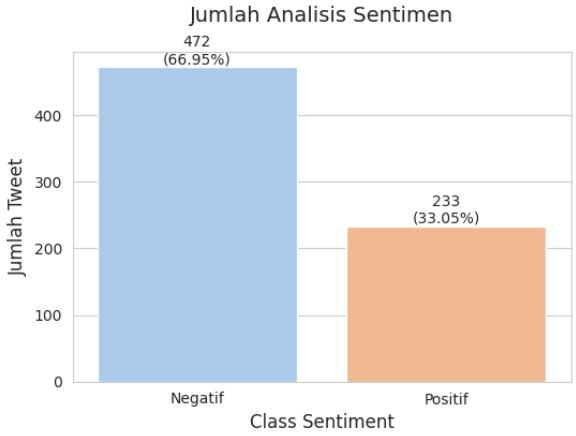
	stopword removal	Score	Sentiment
0	bubarkan nkri bentuk negara federal hutan aman	2	Positif
1	konoha maju mayoritas pemimpinnya copet	0	Negatif
2	fotonya prof	0	Negatif
3	uu bikin jerahdan pemimpin kuat kayak china ki...	-2	Negatif
4	jokowi anjbawa musibah	-1	Negatif

Gambar 18. Hasil *Spiling Data.*

Dari cuplikan dataset pasca-pemrosesan, terlihat mekanisme pelabelan sentimen yang berbasis pada skor numerik:

a. Kategori Positif: Diatribusikan pada komentar dengan skor 2 (contoh: narasi tentang keamanan hutan dan bentuk negara).
b. Kategori Negatif: Diatribusikan pada komentar dengan skor ≤ 0 (skor 0, -1, dan -2).
c. Observasi Teknis: Terlihat kata kunci spesifik seperti "bubarkan", "copet", dan "musibah" yang telah melewati tahap stopword removal. Kata-kata ini akan menjadi atribut kunci dalam pembentukan simpul keputusan *(decision nodes)*.

Berikut merupakan grafik jumlah anasisi sentiment yang telah terlabeli, dari hasil pelabelan tersebut didapatkan jumlah kalimat negatif 472 (66,95%) dan positif 233 (33,05%). Grafik jumlah kata terlabeli ditunjukan pada gambar berikut.



Gambar 19. Hasil Pelabelan Data.

Distribusi sentimen hasil pelabelan dan klasifikasi menunjukkan dominasi sentimen negatif sebesar 66,95% dibandingkan sentimen positif sebesar 33,05%. Dominasi ini mencerminkan tingginya tingkat ketidakpuasan, kekhawatiran, serta kritik publik terhadap isu degradasi lingkungan yang dibahas dalam konten video. Kemunculan kata-kata bermuatan negatif seperti *"musibah"*, *"bubarkan"*, dan *"copet"* mempertegas adanya tekanan moral dan tuntutan akuntabilitas terhadap aktor kebijakan. Temuan ini secara empiris mendukung hipotesis H1 dan menegaskan peran media sosial sebagai ruang ekspresi publik terhadap isu lingkungan yang bersifat krusial.

a. Klasifikasi

Proses klasifikasi dalam penelitian ini diimplementasikan menggunakan algoritma *Decision Tree*. Struktur algoritma ini terdiri atas simpul *(node)* dan cabang, di mana setiap simpul merepresentasikan fitur prediktif, sementara cabang-cabangnya melambangkan nilai diskrit atau kategori yang dihasilkan dari fitur tersebut.

$$H(S) = -\sum_{i=1}^{k} pi\ log2(pi)$$

Penjelasan:

S = Himpunan Kasus

A = Atribut

N = Jumlah dari patisi S

Pi = Proporsi dari Si terhadap S

Berikut kode program klasifikasi menggunakan *Decission Tree* program ditunjukan pada gambar 18.

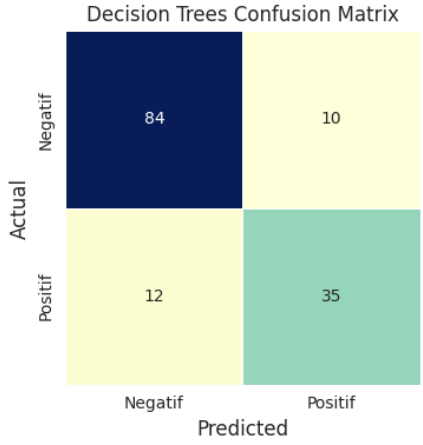
```python
from sklearn.svm import SVC
from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import MultinomialNB
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from sklearn.neural_network import MLPClassifier
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense, Embedding, SpatialDropout1D
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.tree import DecisionTreeClassifier

# Initialize models
models = {
    "SVM": SVC(kernel='linear', random_state=42),
    "KNN": KNeighborsClassifier(n_neighbors=5),
    "Naive Bayes": MultinomialNB(),
    "Random Forest": RandomForestClassifier(random_state=42, n_estimators=100),
    "Decision Trees": DecisionTreeClassifier(random_state=42),
    "Neural Network": MLPClassifier(random_state=42, hidden_layer_sizes=(100,), max_iter=500),
}
```

Gambar 20. Kode Program *Decission Tree.*

b. Evaluasi

Tahap evaluasi dilakukan pada data uji menggunakan metrik berbasis *confusion matrix*. Metrik yang digunakan meliputi *Accuracy, Precision*, *Recall*, dan *F1-Score*. Untuk kasus multi-kelas (positif/negatif/netral), evaluasi umumnya dilaporkan menggunakan *macro-average* agar setiap kelas mendapat bobot yang sama, sehingga performa pada kelas minoritas tetap terlihat. Hasil evaluasi ini menjadi dasar untuk menilai efektivitas model *Decision Tree* dalam mengklasifikasikan sentimen serta mengidentifikasi titik lemah, misalnya kesalahan pada kalimat ironi, negasi kompleks, atau istilah domain yang tidak tercakup pada leksikon.



Gambar 21. Hasil *Confusion Matrix.*

Nilai akurasi yang didapat adalah 0.844

	precision	recall	f1-score	support
Negatif	0.875	0.894	0.884	94.000
Positif	0.778	0.745	0.761	47.000
accuracy	0.844	0.844	0.844	0.844
macro avg	0.826	0.819	0.823	141.000
weighted avg	0.843	0.844	0.843	141.000

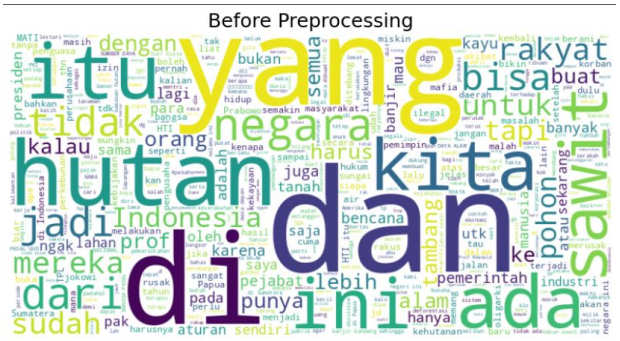
Gambar 22. Hasil Klasifikasi *Decision Tree.*

Hasil pengujian model menunjukkan bahwa algoritma *Decision Tree* mampu melakukan klasifikasi sentimen komentar YouTube dengan tingkat akurasi sebesar 84,4%. Nilai ini mengindikasikan bahwa struktur pohon keputusan yang dibangun dari fitur TF-IDF mampu menangkap pola linguistik yang membedakan sentimen positif dan negatif secara efektif. Temuan ini mendukung hipotesis H2 yang menyatakan bahwa *Decision Tree* memiliki kemampuan klasifikasi yang tinggi dalam konteks analisis sentimen isu degradasi lingkungan. Tingginya akurasi tersebut juga memperkuat argumen bahwa model berbasis aturan *(rule-based interpretability)* tetap relevan untuk analisis opini publik digital yang bersifat kontekstual dan sensitif terhadap isu kebijakan lingkungan.

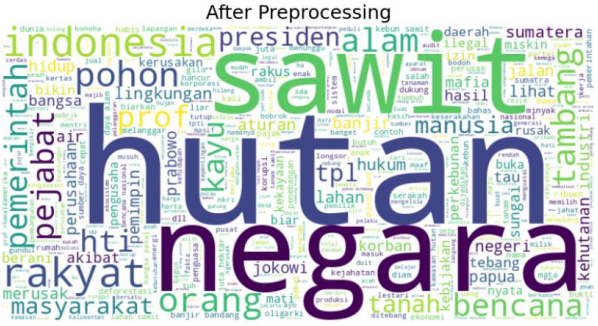
Evaluasi kinerja model menggunakan metrik *precision*, *recall*, dan *F1-score* menunjukkan performa yang konsisten dan relatif seimbang, khususnya pada kelas sentimen negatif. Nilai precision sebesar 87,5% dan recall sebesar 89,4% mengindikasikan bahwa model tidak hanya akurat dalam memprediksi sentimen negatif, tetapi juga memiliki tingkat sensitivitas yang tinggi dalam mendeteksi komentar bermuatan kritik terhadap isu degradasi lingkungan. Nilai *F1-score* sebesar 0,843 memperkuat temuan bahwa model memiliki keseimbangan antara ketepatan dan kelengkapan prediksi. Hasil ini mendukung hipotesis H3 dan menunjukkan bahwa *Decision Tree* efektif digunakan untuk analisis sentimen dengan distribusi kelas yang tidak seimbang.

c. Visualisasi *Word Cloud*

Visualisasi word cloud setelah proses labeling ditunjukkan dengan gambar berikut:

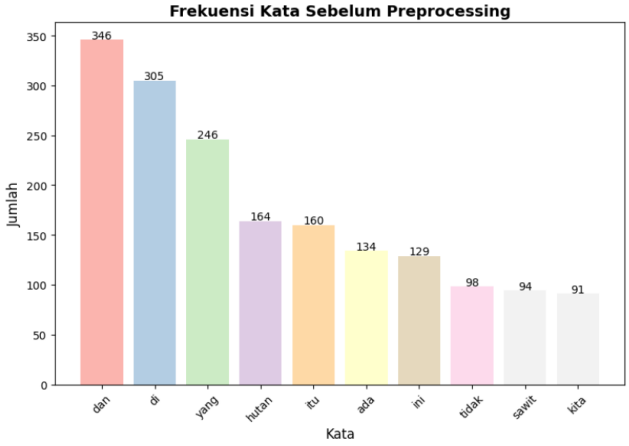


Gambar 23. Visualisasi Sebelum *Prepocessing*.

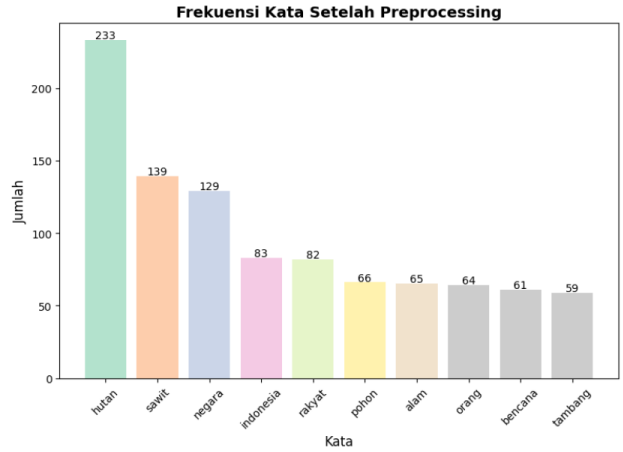


Gambar 24. Visualisasi Setelah *Prepocessing*.

Berikut frekuensi kata yang sering muncul dari hasil analisis sintimen pada penelitian ini.



Gambar 25. Frekuensi Kata Sebelum *Prepocessing*.



Gambar 26. Frekuensi Kata Setelah *Prepocessing*.

5. KESIMPULAN DAN SARAN

Penelitian ini menyimpulkan bahwa implementasi algoritma *Decision Tree* terbukti efektif dalam memetakan sentimen publik terhadap isu degradasi lingkungan melalui serangkaian tahapan teknis yang dimulai dari *scraping* data hingga evaluasi model. Keberhasilan klasifikasi ini didukung oleh proses pra-pemrosesan data yang mendalam, meliputi *cleaning*, *case folding*, normalisasi, *tokenizing*, serta *stopword removal*. Berdasarkan hasil pengujian menggunakan 708 komentar (baris data), model ini berhasil mencapai tingkat akurasi sebesar 84,4%. Secara spesifik, model menunjukkan performa unggul dalam mengidentifikasi sentimen negatif dengan nilai precision sebesar 87,5%, recall 89,4%, dan F1-score 0,884, yang jauh lebih tinggi dibandingkan performa pada kelas positif. Keandalan model ini dikonfirmasi lebih lanjut melalui nilai *weighted average F1-score* yang mencapai 0,843, membuktikan validitas algoritma ini dalam menangkap dinamika opini publik digital secara otomatis.

Analisis terhadap kecenderungan respons publik mengungkapkan dominasi sentimen negatif yang signifikan, di mana jumlah sampel negatif pada data uji mencapai 472 (66,95%) entitas, berbanding terbalik dengan sampel positif yang hanya berjumlah 233 (33,05%) entitas. Munculnya diksi bermuatan kritis seperti "musibah", "copet", dan "bubarkan" dalam kategori sentimen negatif memberikan gambaran nyata mengenai kegelisahan ekologis dan tuntutan akuntabilitas masyarakat terhadap kebijakan kehutanan yang disampaikan di media sosial.

Meskipun model telah mencapai hasil yang memuaskan, penelitian mendatang disarankan untuk mengeksplorasi penggunaan algoritma ensemble learning seperti Random Forest atau XGBoost guna mengejar potensi peningkatan akurasi melampaui angka 84,4%. Selain itu, optimasi pada tahap normalisasi kata sangat direkomendasikan untuk menangani variasi bahasa tidak baku atau kesalahan pengetikan seperti "jerahdan" dan "anjbawa" yang ditemukan dalam dataset. Peneliti juga menyarankan pengembangan skema pelabelan dengan menyertakan kategori 'Netral' untuk komentar dengan skor nol guna mereduksi potensi bias klasifikasi, serta memperluas cakupan data dari berbagai platform media sosial lainnya untuk mendapatkan perspektif publik yang lebih heterogen dan komprehensif terkait isu lingkungan.

DAFTAR REFERENSI

Annisarizki, A. A., & Surahman, S. (2022). Upaya komunikasi publik pemerintah Kota Cilegon dalam mengedukasi masyarakat di masa pandemi COVID-19. *Jurnal Ilmu Komunikasi, 20*(2), 170. https://doi.org/10.31315/jik.v20i2.4344

Armandha, S. T., & Fauziah, N. (2020). *Ekonomi politik media*.

Aulia, S., & Ridwan Maksun, I. (2022). *Jurnal Ilmiah Administrasi Publik (JIAP) reformasi kelembagaan unit pengawas internal: Mengatasi desentralisasi korupsi*. In Sandra Aulia dan Irfan Ridwan Maksun/ JIAP, *8*(1). https://doi.org/10.21776/ub.jiap.2022.008.01.1

Bakti, I. (2024). Peran media sosial dalam meningkatkan kepedulian terhadap isu lingkungan. In *Indonesian Research Journal on Education Web Jurnal Indonesian Research Journal on Education* (Vol. 4). https://doi.org/10.31004/irje.v4i4.1561

De Putri, S., & Toni, A. (2024). Analisis percakapan netizen pada channel YouTube @TotalPolitik sebagai media ruang publik komunikasi politik. *SOSIOLOGI: Jurnal Ilmiah Kajian Ilmu Sosial dan Budaya, 26*(2), 133-150. https://doi.org/10.23960/sosiologi.v26i2.1389

Dudy, I., Effendi, M., Ag, D., Lukman, M., Ag, R., Rustandi, M., Sos, P., Sarbini, A., & Ag, M. (2022). For Millennial Generation *DAKWAH DIGITAL BERBASIS MODERASI BERAGAMA*.

Fatah, Z., & Atreji, R. (2025). Klasifikasi penyakit gagal jantung menggunakan metode decision tree. *Jurnal Ilmiah Sistem Informasi, 4*(3), 844-854. https://doi.org/10.51903/ana03n66

Fatah, Z., & Maghfiro, M. (2025). Klasifikasi jenis transaksi terbanyak pada layanan agen Brilink di Numart dengan algoritma decision tree. *RIGGS: Journal of Artificial Intelligence and Digital Business, 4*(4), 4663-4669. https://doi.org/10.31004/riggs.v4i4.4407

Hermanto, B. (2023). Public participation dynamics toward participatory legislation. https://doi.org/10.29123/jy/v16i2.668

Khatib Sulaiman, J., Lembaga Pemberi Pinjaman Untuk Menentukan Faktor Yang Mempengaruhi Kelayakan Individu Memperoleh Pinjaman Muhammad Riza Wirasena, B., Reinaldi, Mr., Ihsan Jambak, M., & Sriwijaya, U. (2024). Algoritma Decision Tree ID3. *Indonesian Journal of Computer Science*.

Made, I., Adhitya, Y., Ratnawati, D. E., & Arwani, I. (2023). Analisis sentimen feedback mahasiswa terhadap dosen Prodi Teknologi Informasi Departemen Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya (Vol. 7, Issue 6).

Mulyatun, S., Utama, H., & Mustopa, A. (2021). Pendekatan natural language processing pada aplikasi chatbot sebagai alat bantu customer service (Vol. 2, Issue 2). https://doi.org/10.24076/JOISM.2021v3i1.404

Nazifah, N., Prianto, C., & Id, C. A. (2023). Terbit online pada laman web jurnal: https://ejurnalunsam.id/index.php/jicom/ Analisis perbandingan decision tree algoritma C4.5 dengan algoritma lainnya: Sistematic Literature Review.

Novianti, D. N., Shiddieq, D. F., Roji, F. F., & Susilawati, W. (2024). Komparasi algoritma support vector machine dan naïve bayes untuk analisis sentimen pada metaverse.

MALCOM: Indonesian Journal of Machine Learning and Computer Science, *4*(1), 231-239. https://doi.org/10.57152/malcom.v4i1.1061

Pratama, H. Y., & Mustofa, M. U. (2025). Narasi Selat Muria dalam membentuk persepsi publik atas banjir besar Demak Maret 2024: Sebuah analisis wacana kritis. *Journal of Political Issues*, *7*(1), 17-33. https://doi.org/10.33019/jpi.v7i1.282

Rifqi, M., & Gusti Aji, G. (2023). Partisipasi publik di media sosial (Analisis isi pada kolom komentar laman Facebook E100ss dalam isu transportasi publik). In *The Commercium* (Vol. 7). https://doi.org/10.26740/tc.v7i3.56594

Sitorus, E. (2021). Pengaruh sistem informasi manajemen terhadap kinerja pegawai. *Riset Dan E-Jurnal Manajemen Informatika Komputer*, *5*(2). https://doi.org/10.33395/remik.v5i2.11568

Suhendra, S., & Selly Pratiwi, F. (2024). Peran komunikasi digital dalam pembentukan opini publik: Studi kasus media sosial. *Iapa Proceedings Conference*, *293*. https://doi.org/10.30589/proceedings.2024.1059

Wiratama Putra, T., & Triayudi, A. (2022). Analisis sentimen pembelajaran daring menggunakan metode naïve bayes, KNN, dan decision tree. *Jurnal Teknologi Informasi Dan Komunikasi*, *6*(1), 2022. https://doi.org/10.35870/jtik.v6i1.368

Yusuf, D., Razi, F., Arman, S. A., Terisia, V., & Nurjayanti, R. (2025). Prediksi risiko stunting pada balita menggunakan algoritma logistic regression dan decision tree berbasis data terbuka.

Zainu Ridlo, A., Ahmad Rievai, R., Rahayu Yuningsih, S., Rysni Mahalani, E., Khoirutun Nisa, P., Pengembangan Masyarakat Islam, J., & Dakwah dan Ilmu Komunikasi, F. (2024). Ruang publik baru melalui desain di media sosial. *01*(04).