



Tersedia online di [www.jurnal.unipdu.ac.id](http://www.jurnal.unipdu.ac.id)  
**Unipdu**

Halaman jurnal di [www.jurnal.unipdu.ac.id/index.php/teknologi](http://www.jurnal.unipdu.ac.id/index.php/teknologi)



# Analisis Sentimen Perspektif Konsumsi Rokok Tembakau Dan Rokok Elektrik Di Media Sosial X Menggunakan Metode Decision Tree

**M Rizky Ramadhan, Abdul Halim Hasugian**

<sup>a,b</sup> Program Studi Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

email: <sup>a,\*</sup> [kymadh00@gmail.com](mailto:kymadh00@gmail.com)

\*Korespondensi

Dikirim 1 juli 2025; Direvisi 16 juli 2025; Diterima 22 Juli 2025; Diterbitkan 01 Agustus 2025

## Abstrak

Merokok, baik dalam bentuk rokok tradisional maupun elektrik, telah menjadi isu sosial yang menonjol di Indonesia. Topik ini semakin sering dibahas melalui berbagai percakapan publik di media sosial, terutama yang menyoroti dampak kesehatannya. Oleh karena itu, penelitian ini dilakukan untuk menelaah pandangan masyarakat terhadap kedua jenis rokok tersebut di platform media sosial X, dengan menerapkan algoritma klasifikasi Decision Tree sebagai metode analisis. Permasalahan yang dikaji meliputi perbedaan opini publik antara rokok tembakau dan rokok elektrik, serta tingkat efektivitas algoritma Decision Tree dalam mengelompokkan opini publik. Penelitian dilakukan dengan pendekatan kuantitatif, menggunakan lebih dari 2.000 tweet yang dikumpulkan melalui proses crawling dengan kata kunci tertentu. Data yang diperoleh diproses melalui tahapan preprocessing, pelabelan berdasarkan leksikon, pemberian bobot menggunakan metode TF-IDF, dan klasifikasi dengan algoritma Decision Tree CART. Temuan penelitian menunjukkan dominasi opini negatif terhadap kedua jenis rokok, dengan tingkat akurasi model sebesar 87,50% untuk rokok tembakau dan 80,60% untuk rokok elektrik. Penelitian ini menyimpulkan bahwa metode Decision Tree cukup efektif untuk menganalisis opini publik berbasis teks dan mampu menghasilkan klasifikasi yang mudah dipahami. Hasil studi ini dapat dimanfaatkan sebagai dasar dalam pembuatan kebijakan publik dan program kampanye kesehatan berbasis data, serta memberikan kontribusi terhadap kajian analisis sentimen dalam ranah isu sosial.

**Kata Kunci:** Analisis Sentimen, Decision Tree, Rokok Tembakau, Rokok Elektrik, Media Sosial

## Sentiment Analysis Of Tobacco And Electronic Cigarettes Consumption Perspectives On Social Media X Using The Decision Tree Method

### Abstract

Smoking, both traditional and e-cigarettes, has become a prominent social issue in Indonesia. This topic is increasingly discussed through various public conversations on social media, especially those highlighting its health impacts. Therefore, this study was conducted to examine public views on both types of cigarettes on social media platform X, by applying the Decision Tree classification algorithm as an analytical method. The issues studied include differences in public opinion between tobacco cigarettes and e-cigarettes, as well as the effectiveness of the Decision Tree algorithm in classifying public opinion. The study was conducted using a quantitative approach, using more than 2,000 tweets collected through a crawling process with specific keywords. The obtained data were processed through preprocessing, lexicon-based labeling, weighting using the TF-IDF method, and classification using the Decision Tree CART algorithm. The research findings show a predominance of negative opinions towards both types of cigarettes, with a model accuracy rate of 87.50% for tobacco cigarettes and 80.60% for e-cigarettes. This study concludes that the Decision Tree method is quite effective for analyzing text-based public opinion and is able to produce easy-to-understand classifications. The results of this study can be used as a basis for creating public policies and data-based health campaign programs, as well as contributing to sentiment analysis studies in the realm of social issues.

**Keywords:** : Sentiment Analysis, Decision Tree, Tobacco Cigarettes, Electronic Cigarettes, Social Media

Untuk mengutip artikel ini dengan APA Style:

Ramadhan, M. R., & Hasugian, A. H. (2025). Analisis Sentimen Perspektif Konsumsi Rokok Tembakau Dan Rokok Elektrik Di Media Sosial X Menggunakan Metode Decision Tree. TEKNOLOGI: Jurnal Ilmiah Sistem Informasi, 15(2), 87-100: <https://doi.org/10.26594/teknologi.v15i2.5750>



© 2022 Penulis. Diterbitkan oleh Program Studi Sistem Informasi, Universitas Pesantren Tinggi Darul Ulum. Ini adalah artikel open access di bawah lisensi CC BY-NC-NA (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

### 1. Pendahuluan

Kebiasaan merokok, baik rokok tembakau maupun rokok elektrik, telah menjadi fenomena sosial yang kompleks di berbagai negara, termasuk Indonesia (Bagus & Cahyaningrum, 2023). Rokok tembakau merupakan produk yang telah lama dikonsumsi masyarakat dan dikenal memiliki dampak buruk terhadap kesehatan, seperti meningkatkan risiko penyakit jantung, kanker paru, dan gangguan pernapasan (Aditya

et al., 2022). Dalam beberapa tahun terakhir, rokok elektrik atau vape mulai diperkenalkan sebagai alternatif dengan klaim risiko yang lebih rendah. Produk ini berfungsi dengan cara memanaskan cairan yang mengandung nikotin serta zat perasa, tanpa melalui proses pembakaran daun tembakau (Mahirah et al., 2024). Meski demikian, rokok elektrik tetap menimbulkan kekhawatiran terkait kandungan zat kimia dan potensi adiksi, terutama di kalangan remaja (Hutapea & Fasya, 2021).

Data dari World Health Organization (WHO) tahun 2022 menunjukkan bahwa sekitar 1,25 miliar orang berusia di atas 15 tahun di seluruh dunia merupakan perokok. Sementara itu, World Population Review pada tahun yang sama mencatat bahwa Indonesia berada di peringkat kedelapan sebagai negara dengan jumlah perokok terbanyak, yakni mencapai 38,2% dari total populasi. Tren merokok di Indonesia terus menunjukkan peningkatan. Berdasarkan Survei Kesehatan Indonesia (SKI) tahun 2023 yang dirilis oleh Kementerian Kesehatan, tercatat sekitar 70 juta orang di Indonesia merupakan perokok aktif, di mana 7,4% di antaranya adalah remaja berusia 10 hingga 18 tahun. Kondisi ini cukup memprihatinkan, mengingat kelompok usia 15–19 tahun tercatat sebagai penyumbang terbesar dalam angka perokok secara nasional. (Pratama et al., 2022). Di sisi lain, meningkatnya kesadaran masyarakat akan bahaya rokok tembakau turut mendorong transisi ke rokok elektrik, meskipun perdebatan tentang keamanan dan efektivitasnya masih berlangsung (Nurhamidah et al., 2024). Beberapa penelitian menunjukkan bahwa rokok elektrik juga mengandung zat-zat kimia yang berpotensi berbahaya bagi kesehatan, terutama bagi anak muda yang mulai mencoba vaping sebagai gaya hidup (Hutapea & Fasya, 2021).

Di era digital, media sosial menjadi sarana utama masyarakat dalam menyampaikan opini terkait isu kesehatan, termasuk kebiasaan merokok. Platform X (sebelumnya Twitter) banyak digunakan untuk membahas perbandingan antara rokok tembakau dan rokok elektrik. Tanya ChatGPT. Melalui opini yang dibagikan, masyarakat menyuarakan pandangan, pengalaman pribadi, dan reaksi terhadap tren atau kebijakan yang berlaku. Dalam konteks ini, analisis sentimen menjadi alat yang relevan untuk menggali persepsi publik secara sistematis.

Analisis sentimen, yang juga dikenal sebagai opinion mining, merupakan suatu teknik untuk secara otomatis mendeteksi dan mengklasifikasikan opini dalam bentuk teks berdasarkan arah sentimennya, apakah bersifat positif, negatif, atau netral (Petiwi et al., 2022). Teknik ini semakin banyak digunakan untuk memahami opini publik terhadap suatu isu karena mampu mengolah data tidak terstruktur dalam jumlah besar dari platform digital. Dalam studi ini, analisis sentimen digunakan untuk mengevaluasi respons masyarakat terhadap konsumsi rokok tembakau dan rokok elektrik di media sosial X.

Algoritma Decision Tree dipilih sebagai metode klasifikasi dalam penelitian ini karena keunggulannya dalam menghasilkan model yang mudah dipahami dan efektif dalam pengelompokan data tekstual. Berbeda dari pendekatan statistik seperti Naïve Bayes atau metode berbasis vektor seperti SVM, Decision Tree membentuk aturan klasifikasi dalam struktur pohon hierarkis, yang memudahkan dalam interpretasi hasil (Surbakti et al., 2021). Metode ini terbukti efektif dalam analisis sentimen, terutama pada isu kebijakan publik dan layanan digital.

Penelitian sebelumnya oleh Aditya et al. (2022) menganalisis opini masyarakat Indonesia terhadap rokok elektrik (vape) di Twitter menggunakan metode Naïve Bayes. Hasilnya menunjukkan bahwa mayoritas sentimen bersifat netral (77,3%) dengan akurasi model sebesar 77,5%, menandakan efektivitas metode tersebut dalam klasifikasi opini publik (Aditya et al., 2022). Sementara itu, Surbakti et al. (2021) meneliti tanggapan masyarakat terhadap kebijakan PSBB melalui 2.439 tweet menggunakan Decision Tree, Naïve Bayes, dan K-Nearest Neighbors. Algoritma Decision Tree menunjukkan performa terbaik dengan akurasi 84,78% dan recall 100%, menegaskan keunggulannya dalam klasifikasi sentimen berbasis media sosial (Surbakti et al., 2021).

Dengan mempertimbangkan hasil-hasil dari penelitian terdahulu, terlihat bahwa algoritma Decision Tree memiliki keandalan yang baik dalam mengklasifikasikan opini masyarakat berdasarkan data teks dari media sosial (Wijoyo A et al., 2024). Namun, sebagian besar studi tersebut masih terfokus pada isu tunggal dan belum secara komprehensif membandingkan dua jenis produk tembakau, yaitu rokok konvensional dan rokok elektrik, dari perspektif publik di media sosial. Dengan demikian, tujuan dari penelitian ini adalah untuk mengevaluasi sentimen publik terhadap penggunaan rokok tembakau dan rokok elektrik di platform media sosial X melalui penerapan metode Decision Tree. Penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam penyusunan kebijakan publik berbasis data, memperkaya literatur mengenai opini publik terhadap produk tembakau, serta mengembangkan model analisis sentimen yang akurat dan relevan dalam konteks isu kesehatan masyarakat di era digital.

## 2. State of the Art

Analisis sentimen di media sosial telah menjadi topik penting dalam bidang data mining dan pemrosesan bahasa alami, khususnya dalam memahami opini publik terhadap isu-isu sosial. Sejumlah penelitian sebelumnya telah dilakukan dengan pendekatan dan objek yang beragam. Studi-studi berikut menunjukkan perkembangan teknologi dan metode dalam konteks analisis sentimen terhadap konsumsi rokok, baik konvensional maupun elektrik, serta aplikasi algoritma pembelajaran mesin seperti Decision Tree.

### 2.1. Penerapan Algoritma Decision Tree dalam Analisis Sentimen

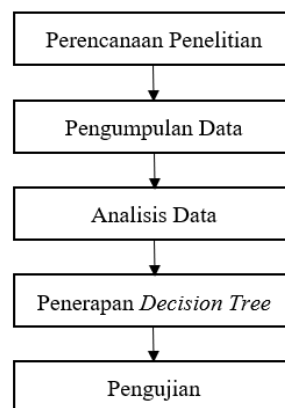
Sejumlah penelitian menunjukkan bahwa algoritma Decision Tree mampu menghasilkan klasifikasi yang akurat dalam analisis sentimen. Surbakti et al. (2021), misalnya, menganalisis opini publik terkait kebijakan PSBB dan mencatat akurasi 84,78% serta recall sebesar 100%. Keunggulan Decision Tree terletak pada kemampuannya membentuk model klasifikasi yang mudah diinterpretasikan. Studi lain oleh Pramesti dan Pratiwi (2023) menganalisis opini publik terhadap program MBKM menggunakan Decision Tree dan SVM. Meskipun SVM menunjukkan akurasi lebih tinggi (84,76%), Decision Tree tetap menunjukkan performa yang kompetitif (72,86%) serta keunggulan dalam interpretabilitas hasil klasifikasi. Di luar konteks kebijakan dan pendidikan, Petiwi et al. (2022) menerapkan metode Decision Tree dalam klasifikasi sentimen terhadap layanan digital seperti GoFood. Dalam studi ini, metode Naïve Bayes dan SVM dibandingkan dengan Decision Tree, dan SVM terbukti unggul. Namun, Decision Tree tetap menunjukkan akurasi di atas 70%, membuktikan bahwa metode ini tetap layak digunakan terutama ketika interpretabilitas menjadi pertimbangan utama.

### 2.2. Perbedaan dan Keunikan Penelitian

Berdasarkan studi-studi sebelumnya, mayoritas penelitian masih berfokus pada analisis sentimen terhadap satu jenis rokok (tembakau atau elektrik) secara terpisah. Penelitian ini berbeda karena secara eksplisit membandingkan perspektif masyarakat terhadap kedua jenis rokok secara bersamaan, dengan pendekatan klasifikasi dua kelas menggunakan algoritma Decision Tree. Melalui analisis data dari media sosial X, penelitian ini merepresentasikan persepsi publik secara waktu nyata sekaligus menyajikan hasil klasifikasi dalam bentuk visual yang mudah dipahami oleh pembuat kebijakan.

## 3. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif yang bertujuan untuk menguji teori dengan menganalisis keterkaitan antara variabel-variabel yang terlibat. Fokus utama dari studi ini adalah mengembangkan analisis sentimen terhadap pandangan masyarakat mengenai konsumsi rokok tembakau dan rokok elektrik di platform media sosial X dengan memanfaatkan metode Decision Tree (Tuasamu et al., 2023). Untuk melaksanakan penelitian ini, diperlukan serangkaian proses yang terdiri dari beberapa tahapan sebagai berikut.



Gambar 1. Alur Penelitian

### 3.1 Perencanaan Penelitian

Penelitian ini dimulai dengan tahap perencanaan sebagai landasan pelaksanaan seluruh proses penelitian. Metode Decision Tree digunakan untuk membandingkan sentimen masyarakat terhadap konsumsi rokok tembakau dan rokok elektrik, berdasarkan data yang diambil dari platform media sosial X.

### 3.2 Pengumpulan Data

Data dikumpulkan melalui crawling di media sosial X dengan kata kunci terkait rokok tembakau (#rokok, #tembakau, #bahaya, dll.) dan rokok elektrik (#vape, #rokokelektrik, #bahaya, dll.). Proses ini menghasilkan

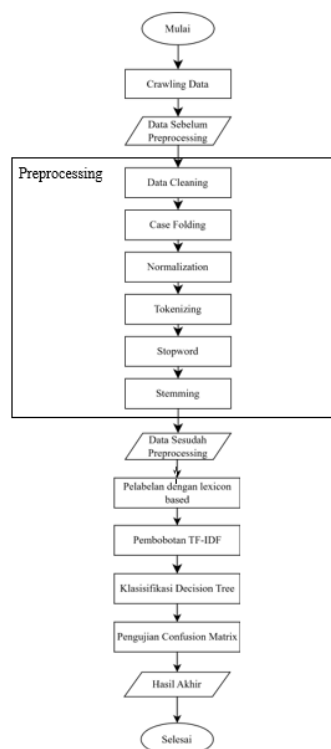
lebih dari 1.000 tweet untuk masing-masing kategori, yang digunakan dalam analisis sentimen menggunakan metode Decision Tree.

### 3.3 Analisis Data

Analisis data dimulai dengan proses preprocessing untuk membersihkan teks dari elemen yang tidak diperlukan, seperti tanda baca, angka, dan stopwords. Selanjutnya, dilakukan pembobotan kata menggunakan metode TF-IDF, kemudian dilanjutkan dengan klasifikasi sentimen menggunakan algoritma Decision Tree.

### 3.4 Penerapan Decision Tree

*Decision Tree* merupakan teknik analisis data yang menggunakan struktur bercabang seperti pohon untuk menentukan keputusan berdasarkan syarat dan aturan tertentu. Metode ini juga termasuk dalam kategori algoritma machine learning. Prosesnya melibatkan transformasi data dalam bentuk tabel menjadi model pohon, yang selanjutnya diubah menjadi serangkaian aturan pengambilan keputusan (rules) (Rizkia et al., 2019). Setelah dilakukannya pembobotan kata dengan TF-IDF, selanjutnya data diklasifikasikan menggunakan metode Decision Tree CART (Classification and Regression Tree) dalam analisis sentimen tentang perspektif konsumsi dari kedua jenis rokok. Berikut flowchart yang digunakan didalam penelitian ini.



Gambar 2. Flowchart Penelitian

Salah satu jenis Decision Tree yang banyak digunakan adalah CART (Classification and Regression Tree), yang membagi data berdasarkan nilai atribut dengan cara meminimalkan impuritas menggunakan indeks Gini (Gini Index).

$$Gini(D) = 1 - \sum_{i=1}^n p_i^2$$

Dalam rumus tersebut, D merupakan himpunan data pada suatu node,  $p_i$  adalah proporsi data untuk kelas ke- $i$  dalam node tersebut, dan  $n$  adalah jumlah total kelas. Nilai Gini mendekati nol menunjukkan bahwa seluruh data dalam node termasuk ke dalam satu kelas yang sama, sedangkan nilai Gini mendekati satu menunjukkan bahwa data tersebar merata ke beberapa kelas. Dengan demikian, pemilihan atribut pemisah terbaik dalam proses pembentukan pohon keputusan dilakukan dengan memilih atribut yang menghasilkan nilai Gini terkecil.

### 3.5 Pengujian

Evaluasi dalam penelitian ini menggunakan metode *Confusion Matrix* untuk menilai kinerja model klasifikasi, dengan metrik seperti akurasi, *recall*, dan *F1-score*. Rumus perhitungan akurasi, *precision*, *recall*, dan *F1-score* disajikan pada bagian berikut.

$$\text{Akurasi} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \times 100\%$$

$$\text{Precision} = \frac{TP}{(TP + FP)} \times 100\%$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \times 100\%$$

$$F1\text{-score} = 2 \times \frac{(\text{Recall} \times \text{Precision})}{(\text{Recall} + \text{Precision})} \times 100\%$$

Dalam mengevaluasi kinerja klasifikasi, *True Positive* (TP) adalah data yang tepat diklasifikasikan sebagai kelas positif, sedangkan *True Negative* (TN) merupakan data yang benar diklasifikasikan sebagai kelas negatif. *False Positive* (FP) terjadi ketika data negatif keliru diprediksi sebagai positif, sementara *False Negative* (FN) adalah data positif yang salah diklasifikasikan sebagai negatif.

#### 4. Hasil dan Pembahasan

Penelitian ini memanfaatkan algoritma Decision Tree untuk menganalisis pandangan masyarakat terhadap konsumsi rokok tembakau dan rokok elektrik di media sosial X. Data dikumpulkan melalui proses crawling dengan menggunakan kata kunci seperti #rokok, #tembakau, #vape, dan #rokokelektrik, yang menghasilkan 1.012 tweet terkait rokok tembakau dan 1.004 tweet mengenai rokok elektrik. Tanya ChatGPT (Santoso, 2021). Data yang dikumpulkan kemudian disimpan dalam format .csv dan melalui serangkaian proses preprocessing seperti pembersihan teks, tokenisasi, dan normalisasi. Metode berbasis leksikon digunakan untuk secara otomatis mengidentifikasi sentimen yang terkandung dalam setiap tweet, yang kemudian dikelompokkan ke dalam dua kategori utama: sentimen positif dan negatif. Sebelum dataset dibagi menjadi data pelatihan dan data pengujian, metode TF-IDF digunakan untuk melakukan proses pembobotan kata. Data pelatihan digunakan untuk membangun Model Decision Tree, yang kemudian dievaluasi dengan data pengujian menggunakan metrik seperti akurasi, presisi, recall, dan skor F1. Hasil evaluasi menunjukkan bahwa model ini dapat mengelompokkan opini masyarakat dengan baik dan dapat memberikan gambaran umum tentang sentimen publik terhadap rokok tembakau dan rokok elektrik di platform media sosial. Contoh data yang digunakan dalam penelitian ini ditunjukkan dalam Tabel 1.

Tabel 1. Sampel data rokok tembakau

No	Username	Tweet
1	@mochama32623730	@InovasiTembakau Yg di bahas masalah kematian masalah rokok. Dibalik itu tidak dibahas berapa juta yg dihidupi oleh tembakau mulai dari hulu sampai hilir.
2	@favoritecyborg	Dan perlu kita akui juga secara ekonomi politik sosial budaya rokok punya banyak dampak positif bahkan dari era zaman pra kemerdekaan indonesia. Bisa habis 3 SKS kalo di jabarin soal ini. Jadi ya diem aja lah. saling menghargai satu sama lain aja pada inti nya.
...	...	...
1012	@mariwpossa	@andiicalilyas Rokok bahaya banget jdi stop merokok yaa guys

Tabel 2. Sampel data rokok elektrik

No	Username	Tweet
1	@Thousmi	Mau rokok ataupun vape sama aja bahaya.
2	@ttpbhgia04	@namaku_adalahhh Bukan dibolehin sih mungkin nyokapnya lebih mengerti bahwa vape lebih bahaya dari pada rokok
...	...	....
1004	@KamandaSurya	Suka gk paham sama org yg mikir asap rokok elektrik sama rokok konvensional itu beda. Sama-sama asap jadi merokoklah di tmpt yang tepat.

Tahap berikutnya adalah preprocessing data, yang diawali dengan proses cleaning untuk membersihkan data komentar yang telah dikumpulkan. Pada titik ini, elemen yang tidak penting dihapus. Ini termasuk tagar, sebutan, nama pengguna, tautan, alamat email, angka, tanda baca, dan emotikon. Tabel 3 dan 4 menunjukkan contoh hasil data yang telah dibersihkan selama tahap preprocessing.

Tabel 3. Data hasil cleaning rokok tembakau

<i>Tweet</i>	<i>Cleaning</i>
@InovasiTembakau Yg di bahas masalah kematian masalah rokok. Dibalik itu tidak dibahas berapa juta yg dihidupi oleh tembakau mulai dari hulu sampai hilir. Dan perlu kita akui juga secara ekonomi politik sosial budaya rokok punya banyak dampak positif bahkan dari era zaman pra kemerdekaan indonesia. Bisa habis 3 SKS kalo di jabarin soal ini. Jadi ya diem aja lah. saling menghargai satu sama lain aja pada inti nya.	Yg di bahas masalah kematian masalah rokok Dibalik itu tidak dibahas berapa juta yg dihidupi oleh tembakau mulai dari hulu sampai hilir Dan perlu kita akui juga secara ekonomi politik sosial budaya rokok punya banyak dampak positif bahkan dari era zaman pra kemerdekaan indonesia Bisa habis SKS kalo di jabarin soal ini Jadi ya diem aja lah saling menghargai satu sama lain aja pada inti nya

Tabel 4. Data hasil cleaning rokok elektrik

<i>Tweet</i>	<i>Cleaning</i>
Mau rokok ataupun vape sama aja bahaya. @namaku_adalahhh Bukan dibolehin sih mungkin nyokapnya lebih mengerti bahwa vape lebih bahaya dari pada rokok @sirsazly Vape lebih bahaya	Mau rokok ataupun vape sama aja bahaya Bukan dibolehin sih mungkin nyokapnya lebih mengerti bahwa vape lebih bahaya dari pada rokok Vape lebih bahaya

Tabel 5 dan 6 menunjukkan contoh hasil data yang telah diproses melalui tahap case folding, yang merupakan langkah kedua dalam proses preprocessing. Tahap case folding mengubah seluruh huruf teks menjadi huruf kecil yang sama dengan format penulisan.

Tabel 5. Data hasil case folding rokok tembakau

<i>Cleaning</i>	<i>Case Folding</i>
Yg di bahas masalah kematian masalah rokok Dibalik itu tidak dibahas berapa juta yg dihidupi oleh tembakau mulai dari hulu sampai hilir Dan perlu kita akui juga secara ekonomi politik sosial budaya rokok punya banyak dampak positif bahkan dari era zaman pra kemerdekaan indonesia Bisa habis SKS kalo di jabarin soal ini Jadi ya diem aja lah saling menghargai satu sama lain aja pada inti nya	yg di bahas masalah kematian masalah rokok dibalik itu tidak dibahas berapa juta yg dihidupi oleh tembakau mulai dari hulu sampai hilir dan perlu kita akui juga secara ekonomi politik sosial budaya rokok punya banyak dampak positif bahkan dari era zaman pra kemerdekaan indonesia bisa habis sks kalo di jabarin soal ini jadi ya diem aja lah saling menghargai satu sama lain aja pada inti nya

Tabel 6. Data hasil case folding rokok elektrik

<i>Cleaning</i>	<i>Case Folding</i>
Mau rokok ataupun vape sama aja bahaya Bukan dibolehin sih mungkin nyokapnya lebih mengerti bahwa vape lebih bahaya dari pada rokok	mau rokok ataupun vape sama aja bahaya bukan dibolehin sih mungkin nyokapnya lebih mengerti bahwa vape lebih bahaya dari pada rokok

Normalisasi adalah tahap ketiga dari proses preprocessing. Tahap ini dilakukan untuk mengubah kata-kata dalam teks menjadi bentuk baku yang sesuai dengan kaidah Bahasa Indonesia berdasarkan KBBI, dan ini



dilakukan dengan menggunakan file kamus berbahasa Indonesia (kamuskatabaku.xlsx) yang dapat ditemukan di GitHub. Tabel 7 dan 8 menunjukkan contoh data yang telah melalui tahap normalisasi.

Tabel 7. Data hasil normalization rokok tembakau

<i>Case Folding</i>	<i>Normalization</i>
yg di bahas masalah kematian masalah rokok dibalik itu tidak dibahas berapa juta yg dihidupi oleh tembakau mulai dari hulu sampai hilir dan perlu kita akui juga secara ekonomi politik sosial budaya rokok punya banyak dampak positif bahkan dari era zaman pra kemerdekaan indonesia bisa habis sks kalo di jabarin soal ini jadi ya diem aja lah saling menghargai satu sama lain aja pada inti nya	yang di bahas masalah kematian masalah rokok dibalik itu tidak dibahas berapa juta yang dihidupi oleh tembakau mulai dari hulu sampai hilir dan perlu kita akui juga secara ekonomi politik sosial budaya rokok punya banyak dampak positif bahkan dari era zaman pra kemerdekaan indonesia bisa habis sks kalau di jabarin soal ini jadi ya diam saja lah saling menghargai satu sama lain saja pada inti ya

Tabel 8. Data hasil normalization rokok elektrik

<i>Case Folding</i>	<i>Normalization</i>
mau rokok ataupun vape sama aja bahaya bukan dibolehin sih mungkin nyokapnya lebih mengerti bahwa vape lebih bahaya dari pada rokok	mau rokok ataupun vape sama saja bahaya bukan dibolehin sih mungkin nyokapnya lebih mengerti bahwa vape lebih bahaya dari pada rokok

Langkah keempat dalam proses preprocessing adalah tokenizing, yang bertujuan untuk memisahkan setiap kata dalam sebuah kalimat atau tweet menjadi unit-unit kata secara terpisah. Contoh hasil data yang telah diproses melalui tahap tokenizing dapat ditemukan pada Tabel 9 dan 10.

Tabel 9. Data hasil tokenizing rokok tembakau

<i>Normalization</i>	<i>Tokenizing</i>
yang di bahas masalah kematian masalah rokok dibalik itu tidak dibahas berapa juta yang dihidupi oleh tembakau mulai dari hulu sampai hilir  dan perlu kita akui juga secara ekonomi politik sosial budaya rokok punya banyak dampak positif bahkan dari era zaman pra kemerdekaan indonesia bisa habis sks kalau di jabarin soal ini jadi ya diam saja lah saling menghargai satu sama lain saja pada inti ya	['yang', 'di', 'bahas', 'masalah', 'kematian', 'masalah', 'rokok', 'dibalik', 'itu', 'tidak', 'dibahas', 'berapa', 'juta', 'yang', 'dihidupi', 'oleh', 'tembakau', 'mulai', 'dari', 'hulu', 'sampai', 'hilir']  ['dan', 'perlu', 'kita', 'akui', 'juga', 'secara', 'ekonomi', 'politik', 'sosial', 'budaya', 'rokok', 'punya', 'banyak', 'dampak', 'positif', 'bahkan', 'dari', 'era', 'zaman', 'pra', 'kemerdekaan', 'indonesia', 'bisa', 'habis', 'sks', 'kalau', 'di', 'jabarin', 'soal', 'ini', 'jadi', 'ya', 'diam', 'saja', 'lah', 'saling', 'menghargai', 'satu', 'sama', 'lain', 'saja', 'pada', 'inti', 'ya']

Tabel 10. Data hasil tokenizing rokok elektrik

<i>Normalization</i>	<i>Tokenizing</i>
mau rokok ataupun vape sama saja bahaya bukan dibolehin sih mungkin nyokapnya lebih mengerti bahwa vape lebih bahaya dari pada rokok	['mau', 'rokok', 'ataupun', 'vape', 'sama', 'saja', 'bahaya'] ['bukan', 'dibolehin', 'sih', 'mungkin', 'nyokapnya', 'lebih', 'mengerti', 'bahwa', 'vape', 'lebih', 'bahaya', 'dari', 'pada', 'rokok']

Tujuan dari proses selanjutnya adalah untuk menghilangkan kata-kata seperti kata ganti dan kata hubung dari dokumen. Tabel 11 dan 12 menunjukkan contoh data yang telah melalui proses preprocessing data tahap ini.

Tabel 11. Data hasil stopwords removal rokok tembakau

<i>Tokenizing</i>	<i>Stopword Removal</i>
-------------------	-------------------------

['yang', 'di', 'bahas', 'masalah', 'kematian', 'masalah', 'rokok', 'dibalik', 'itu', 'tidak', 'dibahas', 'berapa', 'juta', 'yang', 'dihidupi', 'oleh', 'tembakau', 'mulai', 'dari', 'hulu', 'sampai', 'hilir']	['bahas', 'kematian', 'rokok', 'dibalik', 'dibahas', 'juta', 'dihidupi', 'tembakau', 'hulu', 'hilir']
['dan', 'perlu', 'kita', 'akui', 'juga', 'secara', 'ekonomi', 'politik', 'sosial', 'budaya', 'rokok', 'punya', 'banyak', 'dampak', 'positif', 'bahkan', 'dari', 'era', 'zaman', 'pra', 'kemerdekaan', 'indonesia', 'bisa', 'habis', 'sks', 'kalau', 'di', 'jabarin', 'soal', 'ini', 'jadi', 'ya', 'diam', 'saja', 'lah', 'saling', 'menghargai', 'satu', 'sama', 'lain', 'saja', 'pada', 'inti', 'ya']	['akui', 'ekonomi', 'politik', 'sosial', 'budaya', 'rokok', 'dampak', 'positif', 'era', 'zaman', 'pra', 'kemerdekaan', 'indonesia', 'habis', 'sks', 'jabarin', 'ya', 'diam', 'menghargai', 'inti', 'ya']

Tabel 12. Data hasil stopwords removal rokok elektrik

<i>Tokenizing</i>	<i>Stopword Removal</i>
['mau', 'rokok', 'ataupun', 'vape', 'sama', 'saja', 'bahaya']	['rokok', 'vape', 'bahaya']
['bukan', 'dibolehin', 'sih', 'mungkin', 'nyokapnya', 'lebih', 'mengerti', 'bahwa', 'vape', 'lebih', 'bahaya', 'dari', 'pada', 'rokok']	['dibolehin', 'sih', 'nyokapnya', 'mengerti', 'vape', 'bahaya', 'rokok']

Tahapan terakhir dalam proses preprocessing adalah stemming, yang bertujuan untuk menghilangkan seluruh bentuk imbuhan pada kata-kata dalam dokumen, seperti awalan, akhiran, maupun bentuk jamak, sehingga diperoleh kata dasar. Contoh hasil data yang telah melalui tahap stemming ditampilkan dalam Tabel 13 dan 14.

Tabel 13. Data hasil stemming rokok tembakau

<i>Stopword Removal</i>	<i>Stemming</i>
['bahas', 'kematian', 'rokok', 'dibalik', 'dibahas', 'juta', 'dihidupi', 'tembakau', 'hulu', 'hilir']	bahas mati rokok balik bahas juta hidup tembakau hulu hilir
['akui', 'ekonomi', 'politik', 'sosial', 'budaya', 'rokok', 'dampak', 'positif', 'era', 'zaman', 'pra', 'kemerdekaan', 'indonesia', 'habis', 'sks', 'jabarin', 'ya', 'diam', 'menghargai', 'inti', 'ya']	akui ekonomi politik sosial budaya rokok dampak positif era zaman pra merdeka indonesia habis sks jabarin ya diam harga inti ya

Tabel 14. Data hasil stemming rokok elektrik

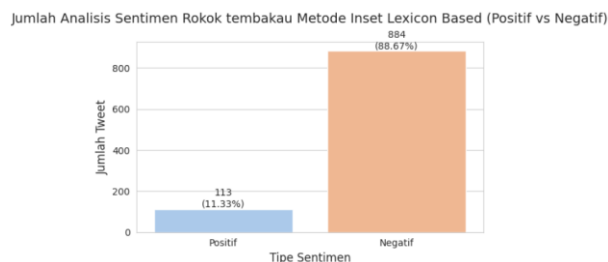
<i>Stopword Removal</i>	<i>Stemming</i>
['rokok', 'vape', 'bahaya']	rokok vape bahaya
['dibolehin', 'sih', 'nyokapnya', 'mengerti', 'vape', 'bahaya', 'rokok']	boleh sih nyokapnya ngerti vape bahaya rokok

Pelabelan data dilakukan menggunakan pendekatan leksikon berbasis skor sentimen yang diperoleh dari GitHub, dengan klasifikasi ke dalam dua kategori utama: positif dan negatif. Nilai sentimen dihitung dalam skala antara -1 hingga +1, di mana teks yang memiliki skor  $\geq 0$  diklasifikasikan sebagai "Positif", sedangkan teks dengan skor di bawah 0 dikategorikan sebagai "Negatif". Pendekatan ini mempermudah proses analisis sentimen dengan hanya menggunakan dua kelas. Contoh hasil proses pelabelan ditampilkan pada Tabel 15 dan 16.

Tabel 15. Data hasil pelabelan rokok tembakau

<i>Stemming</i>	Skor	Label
bahas mati rokok balik bahas juta hidup tembakau hulu hilir	0	Positif
akui ekonomi politik sosial budaya rokok dampak positif era zaman pra merdeka indonesia habis sks jabarin ya diam harga inti ya	2	Positif

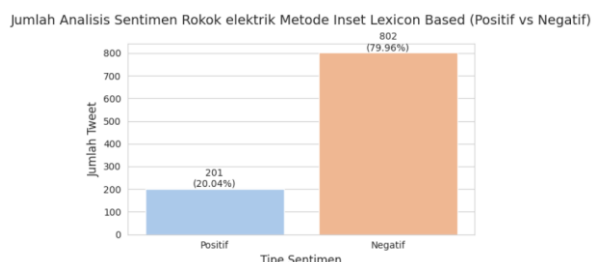




Gambar 3. Pemberian label dengan lexicon based rokok tembakau

Tabel 16. Data hasil pelabelan rokok elektrik

<i>Stemming</i>	Skor	Label
rokok vape bahaya	-2	Negatif
boleh sih nyokapnya ngerti vape bahaya rokok	-2	Negatif



Gambar 4. Pemberian label dengan lexicon based rokok elektrik

Metode TF-IDF digunakan untuk menentukan pentingnya suatu kata dalam dokumen. Metode ini mencakup perhitungan frekuensi kata (TF), frekuensi dokumen balik (IDF), dan pengalihan untuk mendapatkan bobot akhir (W). Kata-kata dengan bobot tinggi dianggap lebih relevan dalam konteks dokumen. Metode ini diterapkan pada tweet yang telah dipraolah untuk rokok elektrik dan tembakau. Dalam proses ini, kata-kata seperti "rokok bahaya banget, stop rokok ya guys" dan "vape bahaya" digunakan. Tabel 17 dan 18 menunjukkan hasil perhitungan TF dan DF.

Tabel 17. Perhitungan nilai TF dan DF rokok tembakau

<i>Term</i>	<i>TF</i>			<i>DF</i>
	D1	D2	D3	
bahas	2	0	0	1
mati	1	0	0	1
rokok	1	1	2	3
balik	1	0	0	1
juta	1	0	0	1
hidup	1	0	0	1
tembakau	1	0	0	1
hulu	1	0	0	1
hilir	1	0	0	1
rebah	1	0	0	1
sakit	1	0	0	1
bahaya	0	1	1	2
b banget	1	0	0	1
stop	1	0	0	1
ya	1	0	0	1
guys	1	0	0	1

Tabel 18. Perolehan nilai TF dan DF untuk data rokok elektrik

<i>Term</i>	<i>TF</i>			<i>DF</i>
	D1	D2	D3	
boleh	1	0	0	1
sih	1	0	0	1
nyokapnya	1	0	0	1
ngerti	1	0	0	1

vape	1	1	2	3
bahaya	1	1	1	3
rokok	1	0	2	2
ya	0	0	1	1
tinggal	0	0	1	1
mikir	0	0	1	1
saran	0	0	1	1
gua	0	0	1	1
mending	0	0	2	1
nyerutu	0	0	1	1
nder	0	0	1	1
haha	0	0	1	1
canda	0	0	1	1
orang	0	0	1	1

Berikut adalah contoh perhitungan term pertama, yaitu "bahas", yang telah diketahui  $N = 3$ , dan  $DF = 1$ .

$$IDF(bahas) = \log\left(\frac{3}{1}\right) = 0.4771$$

Tabel 19. Perhitungan nilai IDF rokok tembakau

Term	DF	IDF
bahas	1	0.4771
mati	1	0.4771
rokok	3	0
...	...	...
guys	1	0.4771

Tabel 20. Perhitungan nilai IDF rokok elektrik

Term	DF	IDF
boleh	1	0.4771
sih	1	0.4771
nyokapnya	1	0.4771
...	...	...
orang	1	0.4771

Contoh perhitungan awal ditunjukkan pada term "bahas" dengan nilai TF sebesar 2, dan  $IDF = 0.4771$ .

$$TF - IDF(bahas) = 2 \times 0.4771 = 0.9542$$

Tabel 21. Perhitungan nilai TF-IDF rokok tembakau

Term	TF-IDF		
	D1	D2	D3
bahas	0.9542	0	0
mati	0.4771	0	0
rokok	0	0	0
...	...	...	...
guys	0.4771	0	0

Tabel 22. Perhitungan nilai TF-IDF rokok elektrik

Term	TF-IDF		
	D1	D2	D3
boleh	0.4771	0	0
sih	0.4771	0	0
nyokapnya	0.4771	0	0
...	...	...	...
orang	0	0	0.4771

Setelah tahap pembobotan kata menggunakan metode TF-IDF selesai dilakukan, langkah berikutnya adalah mengklasifikasikan data uji dengan algoritma Decision Tree jenis Classification and Regression Tree (CART). Metode ini membentuk struktur pohon keputusan berdasarkan atribut yang paling relevan dalam memisahkan kelas sentimen. Proses klasifikasi dilakukan menggunakan data latih yang telah melalui tahap

preprocessing dan pelabelan sebelumnya. Contoh perhitungan klasifikasi terhadap data uji disajikan untuk menggambarkan penerapan model dalam memprediksi sentimen masyarakat terkait konsumsi rokok tembakau dan rokok elektrik.

Tabel 23. Contoh data untuk klasifikasi decision tree

Dokumen	Tweet	Label
D1	bahas mati rokok balik bahas juta hidup tembakau hulu hilir	Positif
D2	rebah sakit rokok bahaya	Negatif
D3	rokok bahaya banget stop rokok ya guys	Positif

Pada contoh data yang telah disajikan pada tabel 23, diketahui terdapat 3 dokumen, dimana 2 dokumen yang memiliki nilai “positif”, dan 1 dokumen yang memiliki nilai “negatif”. Selanjutnya dianggap dokumen bernilai “positif” menjadi label 1, dan dokumen bernilai “negatif” menjadi label 0. Kemudian dihitung Gini parent atau nilai Gini awal.

2 Positif = label 1

1 Negatif = label 0

$$\begin{aligned}
 Gini_{parent} &= 1 - \sum (p_i)^2 \\
 Gini_{parent} &= 1 - (p_1)^2 - (p_0)^2 \\
 &= 1 - \left(\frac{2}{3}\right)^2 - \left(\frac{1}{3}\right)^2 \\
 &= 1 - \frac{4}{9} - \frac{1}{9} = \frac{2}{9} \\
 &= 0.444
 \end{aligned}$$

Setelah nilai Gini parent didapat, kemudian sebagai contoh akan digunakan kata yang paling banyak muncul dari ketiga dokumen, yaitu kata “bahaya” dan “rokok” untuk dihitung nilai Gini dan nilai Gain dari masing-masing kata tersebut, akan tetapi untuk kata “rokok” karena muncul di setiap dokumen maka nilai Gain nya 0, dan kata “bahaya” muncul pada 2 dokumen, yaitu dokumen 2 dan 3, maka contoh perhitungannya sebagai berikut.

“bahaya” = 1 (D1 dan D3) dengan label positif

“bahaya” = 0 (D2) dengan label negatif

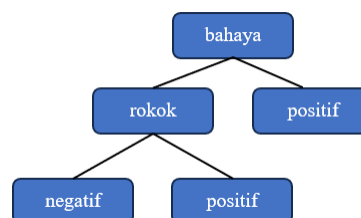
$$\begin{aligned}
 Gini_{bahaya1} &= 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 \\
 &= 1 - 0.25 - 0.25 = 0.5
 \end{aligned}$$

$$Gini_{bahaya0} = 1 - (1)^2 = 0$$

$$\begin{aligned}
 Gini_{split} &= \frac{2}{3} \times 0.5 + \frac{1}{3} \times 0 = \frac{1}{3} \\
 &= 0.333
 \end{aligned}$$

$$\begin{aligned}
 Gain &= Gini_{parent} - Gini_{split} \\
 &= 0.444 - 0.333 \\
 &= 0.111
 \end{aligned}$$

Karena kata “bahaya” memiliki nilai gain tertinggi dan kata “rokok” memiliki nilai terendah dari semua kata di dokumen 1, 2 dan 3. Maka kedua kata tersebut akan dipilih untuk membangun sebuah Decision Tree sederhana.



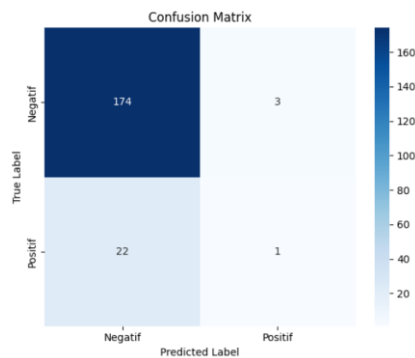
Gambar 5. Decision Tree CART sederhana

Penjelasan pada Gambar 4.3 adalah pohon dimulai dari kata “bahaya”. Misalkan pada sebuah dokumen, apakah terdapat kata “bahaya” atau tidak, apabila tidak ada maka dokumen bernilai “positif”. Namun apabila terdapat kata “bahaya” Masih belum pasti apakah dokumen “positif” atau “negatif”, jadi ditambahkan node baru yaitu “rokok”. Jika didalam dokumen terdapat kata “rokok” maka dokumen bernilai “negatif”, tetapi jika tidak, masih ada kemungkinan dokumen bernilai “positif”. Maka didapatkan hasil dari uji prediksi tiap dokumen dengan metode decision tree cart yang disajikan pada Tabel 24 berikut.

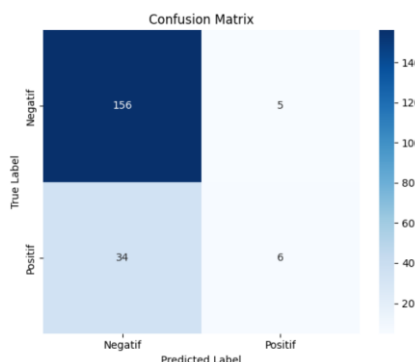
Tabel 24. Hasil uji untuk klasifikasi decision tree

Dokumen	Tweet	Label awal	Label akhir
D1	bahas mati rokok balik bahas juta hidup tembakau hulu hilir	Positif	Positif
D2	rebah sakit rokok bahaya	Negatif	Negatif
D3	rokok bahaya banget stop rokok ya guys	Positif	Negatif

Dari hasil uji yang sudah dilakukan dengan klasifikasi decision tree cart maka didapat akurasi sebesar 66.7% karena sudah benar menebak 2 dari 3 dokumen yang diberikan. Setelah proses pengujian dengan algoritma Decision Tree selesai, diperoleh hasil klasifikasi berupa label sentimen dari data uji berdasarkan model yang telah dilatih. Hasil prediksi selanjutnya dibandingkan dengan label sebenarnya untuk menghitung performa model menggunakan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Mengingat distribusi data tidak seimbang, di mana sentimen negatif jauh lebih dominan, maka evaluasi difokuskan pada kelas negatif. Confusion matrix disajikan untuk menunjukkan performa model dalam mengklasifikasikan sentimen.



Gambar 6. Confusion matrix rokok tembakau



Gambar 7. Confusion matrix rokok elektrik

Gambar 6 dan Gambar 7 menampilkan hasil klasifikasi yang digunakan untuk menghitung nilai accuracy, precision, recall, dan F1-score berdasarkan rumus yang telah ditentukan. Langkah awal perhitungan difokuskan pada sentimen terhadap "Rokok Tembakau".

$$\text{Akurasi} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \times 100\% = \frac{(1 + 174)}{(1 + 174 + 3 + 22)} = \frac{175}{200} = 87,50\%$$

$$\text{Precision} = \frac{TN}{(TN + FN)} \times 100\% = \frac{174}{(174 + 22)} = \frac{174}{196} = 88,77\%$$

$$\text{Recall} = \frac{TN}{(TN + FP)} \times 100\% = \frac{174}{(174 + 3)} = \frac{174}{177} = 98,30\%$$

$$\text{F1-score} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \times 100\%$$

$$= 2 \times \frac{(88,77\% \times 98,30\%)}{(88,77\% + 98,30\%)} \times 100\%$$

$$= 2 \times \frac{87,26\%}{187,07\%} \times 100\%$$

$$= 93,29\%$$

Selanjutnya pada sentiment "Rokok Elektrik".

$$\text{Akurasi} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \times 100\% = \frac{(6 + 156)}{(6 + 156 + 5 + 34)} = \frac{162}{201} = 80,60\%$$

$$\text{Precision} = \frac{TN}{(TN + FN)} \times 100\% = \frac{156}{(156 + 34)} = \frac{156}{190} = 82,10\%$$

$$\text{Recall} = \frac{TN}{(TN + FP)} \times 100\% = \frac{156}{(156 + 5)} = \frac{156}{161} = 96,89\%$$

$$F1\text{-score} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \times 100\%$$

$$= 2 \times \frac{(82,10\% \times 96,89\%)}{(82,10\% + 96,89\%)} \times 100\%$$

$$= 2 \times \frac{79,54\%}{178,99\%} \times 100\%$$

$$= 88,87\%$$

Dari hasil penghitungan terhadap data rokok tembakau, terdapat 200 data uji yang dianalisis, dengan akurasi mencapai 87,50%, presisi 88,78%, recall 98,31%, dan F1-score sebesar 93,30%. Sementara itu, untuk data rokok elektrik yang terdiri dari 201 data uji, diperoleh hasil akurasi sebesar 80,60%, presisi 82,11%, recall 96,89%, dan F1-score sebesar 88,89%.

## 5. Kesimpulan

Penelitian ini menganalisis sentimen masyarakat terhadap konsumsi rokok tembakau dan rokok elektrik di media sosial X menggunakan metode *Decision Tree* tipe CART. Data dikumpulkan melalui *crawling*, menghasilkan 997 tweet untuk rokok tembakau dan 1.003 untuk rokok elektrik setelah duplikasi dihapus. Data dibagi menjadi data latih dan uji (tembakau: 797/200; elektrik: 803/201). Pelabelan dilakukan dengan metode *lexicon-based*, menghasilkan 113 positif dan 884 negatif untuk rokok tembakau, serta 201 positif dan 802 negatif untuk rokok elektrik. Hasil klasifikasi menunjukkan performa yang baik: akurasi 87,50%, presisi 88,78%, recall 98,31%, dan F1-score 93,30% untuk rokok tembakau; serta akurasi 80,60%, presisi 82,11%, recall 96,89%, dan F1-score 88,89% untuk rokok elektrik. Sentimen negatif mendominasi pada kedua jenis rokok.

## 6. Kontribusi Penulis

M Rizky Ramadhan: Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, dan Writing – original draft.

Abdul Halim Hasugian: Supervision, Validation, Writing – review & editing, dan Conceptualization.

## 7. Declaration of Competing Interest

Penulis menyatakan tidak ada konflik kepentingan.

## 8. Referensi

- Aditya, D. R., Supriyati, E., & Listyorini, T. (2022). Analisis Sentimen Pengguna Twitter Terhadap Rokok Elektrik (Vape) Di Indonesia Menggunakan Metode Naïve Bayes. *JIPi (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 7(1), 43–50. <https://doi.org/10.29100/jipi.v7i1.2145>
- Bagus, D., & Cahyaningrum, C. P. E. (2023). Perbedaan Patologi Anatomi Tingkat Kerusakan Alveoli Paru dengan Paparan Asap Rokok Konvensional dan Rokok Elektrik. *ARTERI : Jurnal Ilmu Kesehatan*, 4(1), 29–36. <https://doi.org/10.37148/arteri.v4i1.245>
- Hutapea, D. S. M., & Fasya, T. K. (2021). Rokok Elektrik (Vape) sebagai Gaya Hidup Perokok Masa Kini di Kota Lhokseumawe. *Jurnal Ilmu Sosial Dan Ilmu Politik Malikussaleh (JSPM)*, 2(1), 92. <https://doi.org/10.29103/jspm.v2i1.3696>
- Mahirah, R., Aramico, B., & Arifin, V. N. (2024). Faktor-Faktor Yang Berhubungan Dengan Perilaku Merokok Pada Mahasiswa Pria. 5(1), 2.1: 44-53. <https://doi.org/10.34305/jphi.v5i01.1388>
- Nurhamidah, S., Sopiah, P., & Astuti, A. P. K. (2024). PENGARUH MEROKOK ELEKTRIK TERHADAP PENYAKIT LAMBUNG (GERD DAN GASTRITIS) : LITERATURE REVIEW Syifa Nurhamidah, Popi Sopiah, Ayu Prameswari Kusuma Astuti Universitas Pendidikan Indonesia. *Healthy Tadulako Journal*, 10(3), 483–492.
- Petiwi, M. I., Triayudi, A., & Sholihati, I. D. (2022). Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine. *Jurnal Media Informatika Budidarma*, 6(1), 542. <https://doi.org/10.30865/mib.v6i1.3530>
- Pratama, E. A., Hellyana, C. M., & Fadlilah, N. I. (2022). Perbandingan 3 Algoritma Klasifikasi Data Mining Dalam Pro-Kontra Bahaya Rokok Elektrik. *Jurnal Teknoinfo*, 16(1), 93. <https://doi.org/10.33365/jti.v16i1.1534>
- Rizkia, S., Budi Setiawan, E., & Puspandari, D. (2019). Analisis Sentimen Kepuasan Pelanggan Terhadap Internet Provider Indihome di Twitter Menggunakan Metode Decision Tree dan Pembobotan TF-IDF. *E-Proceeding of Engineering*, 6(Agustus), 9683–9693.
- Santoso, T. G. (2021). Analisis Sentimen Pada Tweet Dengan Tagar #Bpjsrasarentenir Menggunakan

- Metode Support Vectore Machine (Svm). (*Doctoral Dissertation, Universitas Islam Riau*), 12–13.
- Surbakti, A. Q., Hayami, R., & Al Amien, J. (2021). Analisa Tanggapan Terhadap Psbb Di Indonesia Dengan Algoritma Decision Tree Pada Twitter. *Jurnal CoSciTech (Computer Science and Information Technology)*, 2(2), 91–97. <https://doi.org/10.37859/coscitech.v2i2.2851>
- Tuasamu, Z., Lewaru, N. A. I. M., Idris, M. R., Syafaat, A. B. N., Faradilla, F., Fadlan, M., Nadiva, P., & Efendi, R. (2023). Analisis Sistem Informasi Akuntansi Siklus Pendapatan Menggunakan DFD Dan Flowchart Pada Bisnis Porobico. *Jurnal Bisnis Dan Manajemen(JURBISMAN)*, 1(2), 495–510. <https://ejournal.lapad.id/index.php/jurbisman/article/view/181>
- Wijoyo A, Saputra A, Ristanti S, Sya'ban S, Amalia M, & Febriansyah R. (2024). Pembelajaran Machine Learning. *OKTAL (Jurnal Ilmu Komputer Dan Science)*, 3(2), 375–380. <https://journal.mediapublikasi.id/index.php/oktal/article/view/2305>