

---

# Academia Open



*By Universitas Muhammadiyah Sidoarjo*

---

# Academia Open

Vol. 11 No. 1 (2026): June  
DOI: 10.21070/acopen.11.2026.14689

## Table Of Contents

<b>Journal Cover</b> .....	1
<b>Author[s] Statement</b> .....	3
<b>Editorial Team</b> .....	4
<b>Article information</b> .....	5
Check this article update (crossmark) .....	5
Check this article impact .....	5
Cite this article.....	5
<b>Title page</b> .....	6
Article Title .....	6
Author information .....	6
Abstract .....	6
<b>Article content</b> .....	7

## Originality Statement

The author[s] declare that this article is their own work and to the best of their knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the published of any other published materials, except where due acknowledgement is made in the article. Any contribution made to the research by others, with whom author[s] have work, is explicitly acknowledged in the article.

## Conflict of Interest Statement

The author[s] declare that this article was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Copyright Statement

Copyright © Author(s). This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

# Academia Open

Vol. 11 No. 1 (2026): June  
DOI: 10.21070/acopen.11.2026.14689

## EDITORIAL TEAM

### Editor in Chief

Mochammad Tanzil Multazam, Universitas Muhammadiyah Sidoarjo, Indonesia

### Managing Editor

Bobur Sobirov, Samarkand Institute of Economics and Service, Uzbekistan

### Editors

Fika Megawati, Universitas Muhammadiyah Sidoarjo, Indonesia

Mahardika Darmawan Kusuma Wardana, Universitas Muhammadiyah Sidoarjo, Indonesia

Wiwit Wahyu Wijayanti, Universitas Muhammadiyah Sidoarjo, Indonesia

Farkhod Abdurakhmonov, Silk Road International Tourism University, Uzbekistan

Dr. Hindarto, Universitas Muhammadiyah Sidoarjo, Indonesia

Evi Rinata, Universitas Muhammadiyah Sidoarjo, Indonesia

M Faisal Amir, Universitas Muhammadiyah Sidoarjo, Indonesia

Dr. Hana Catur Wahyuni, Universitas Muhammadiyah Sidoarjo, Indonesia

Complete list of editorial team ([link](#))

Complete list of indexing services for this journal ([link](#))

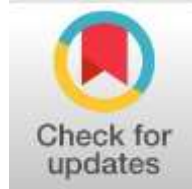
How to submit to this journal ([link](#))

# Academia Open

Vol. 11 No. 1 (2026): June  
DOI: 10.21070/acopen.11.2026.14689

## Article information

**Check this article update (crossmark)**



**Check this article impact (\*)**



**Save this article to Mendeley**



(\*) Time for indexing process is various, depends on indexing database platform

## Decision Tree Outperforms Naive Bayes For Infidelity Divorce Sentiment Classification: Decision Tree Mengungguli Naive Bayes Untuk Klasifikasi Sentimen Perceraian Akibat Perselingkuhan

**Parabelem Tinno Dolf Rompas, parabelemrompas@unima.ac.id**

*Program Studi Teknik Mesin, Universitas Negeri Manado, Tondano 95619, Indonesia*

**Sheila Sasoeng, 22210179@unima.ac.id (\*)**

*Program Studi Teknik Informatika, Universitas Negeri Manado, Tondano 95619, Indonesia*

**Gladly Caren Rorimpandey, gladlyrorimpandey@unima.ac.id**

*Program Studi Teknik Informatika, Universitas Negeri Manado, Tondano 95619, Indonesia*

(\*) Corresponding author

### Abstract

**General Background** Divorce due to infidelity has become a massive public issue widely exposed on social media platforms. **Specific Background** This surge in digital opinion generates unstructured textual data reflecting polarized moral perceptions, necessitating machine learning to map public sentiment. **Knowledge Gap** Previous sentiment analysis predominantly focused on public services, leaving a significant void in examining domestic conflicts like marital betrayal. **Aims** This study analyzes public perception on social media X and compares Naïve Bayes and Decision Tree algorithms on an imbalanced dataset. **Results** Processing 3,450 tweets using TF-IDF and SMOTE revealed that public opinion is dominantly Neutral (47.9%) and Negative (46.7%), with Positive sentiment at only 5.3%. The Decision Tree algorithm achieved 96.95% accuracy, significantly outperforming Naïve Bayes (73.59%). **Novelty** This research proves the Decision Tree model's superior stability and accuracy over Naïve Bayes in handling imbalanced textual data regarding sensitive moral issues. **Implications** The developed model and its interactive dashboard offer a robust foundation for content moderation and formulating contextual family resilience education programs.

### Highlights

- ♦ The rule-based algorithm achieved 96.95% accuracy, demonstrating exceptional stability when analyzing textual data.
- ♦ Public reactions on the X platform regarding marital betrayal are primarily divided into neutral and negative categories.
- ♦ Implementing the Synthetic Minority Over-sampling Technique effectively addressed dataset imbalances prior to model evaluation.

**Keywords** Decision Tree; Naive Bayes; Sentiment Analysis; Social Media X; Synthetic Minority Oversampling Technique

Published date: 2026-06-19

## PENDAHULUAN

Kasus perceraian yang akibat perselingkuhan Adalah sebuah fenomena yang tidak asing lagi bagi pengguna media sosial dan sekarang ini sedang meningkat di Indonesia yang telah bertransformasi menjadi isu publik yang sangat masif, terutama dengan adanya peningkatan eksposur di media sosial. Data dari Badan Pusat Statistik (BPS) dan Mahkamah Agung menunjukkan bahwa peningkatan kasus perceraian yang signifikan, dengan dominasi “cerai gugat” yang diajukan oleh pihak istri. Salah satu faktor utama pemicu keretakan rumah tangga ini Adalah perselingkuhan (Khairuddin, 2024). Tingginya angka perceraian di Indonesia merupakan sebuah fenomena sosial yang mengkhawatirkan. Data Badan Pusat Statistik (BPS) tahun 2024 mencatat sebanyak 394.608 kasus perceraian, di mana mayoritas disebabkan oleh perselisihan yang terus-menerus, masalah ekonomi, serta faktor sosial lainnya [1].

Perselingkuhan yang seringkali menjadi suatu pemicu berita viral dan banyak menarik perhatian publik di era sekarang ini kesedihan sering kali dijadikan komoditas di media sosial, Dimana orang memanfaatkannya untuk mencari perhatian dan keuntungan dengan upaya untuk kepopuleran. Fenomena ini menunjukkan bagaimana emosi diubah menjadi alat untuk menarik simpati atau popularitas. Di era digital, isu perselingkuhan terutama yang mengakibatkan figur publik cepat sekali menjadi berita viral. Media sosial X (sebelumnya twitter) telah menjadi platform utama bagi korban untuk mengungkap fakta perselingkuhan *spilling the tea* (menceritakan topik hangat) guna mencari keadilan atau sekedar dukungan publik. Fenomena ini menciptakan pergeseran di mana kesedihan dan masalah privasi kerap menjadi komoditas digital demi menarik simpati atau popularitas. Namun, di balik keraiamaian tersebut, dampak yang ditimbulkan sangatlah destruktif, Dimana mulai dari trauma psikologis bagi pasangan yang telah dikhianati hingga gangguan emosional pada anak yang kehilangan figure orang tua akibat perceraian [2].

Dalam komunitas Marah-Marah Twitter, netizen menyampaikan respons emosional secara terbuka, sehingga mencerminkan cara masyarakat memandang dan menanggapi isu perselingkuhan [3]. Fenomena ini berpotensi menimbulkan dampak social yang cukup signifikan di Indonesia sebagai negara hukum.

Untuk mengatasi dampak destruktif dari komodifikasi konflik domestik di ruang digital, penelitian ini menerapkan pendekatan Machine Learning untuk menganalisis sentimen perceraian akibat perselingkuhan di media sosial X menghasilkan data tekstual yang mencerminkan berbagai sentimen Masyarakat, baik berupa dukungan, kecaman, maupun opini yang netral. Sehingga tingginya polarisasi opini dari publik ini memerlukan analisis yang sistematis untuk memahami pola dari reaksi Masyarakat terhadap isu sosial tersebut [4]. Namun, dalam mengolah ribuan data teks secara manual tidaklah efisien, sehingga diperlukan pendekatan Machine Learning untuk melakukan klasifikasi sentimen secara otomatis dan juga akurat, menurut [5]. Jika data memiliki ketimpangan distribusi kelas (*imbalanced data*) diatasi melalui integrasi teknik Synthetic Minority Over-sampling Technique (SMOTE) sebelum fase pelatihan model [6]. Synthetic Minority Over-sampling Technique (SMOTE). Teknik ini diterapkan untuk menyeimbangkan proporsi data dengan cara menyintesis data baru pada kelas minoritas berdasarkan karakteristik kedekatan fiturnya, sehingga kedua algoritma yang diuji dapat mempelajari pola dari seluruh kategori sentimen secara adil dan objektif. Rangkaian proses ini diawali dengan pengumpulan data dari media sosial X pada tahap Data Selection [7]. Selanjutnya, klasifikasi sentimen dioptimalkan dengan membandingkan efisiensi tinggi algoritma Naïve Bayes pada data berdimensi besar dan transparansi aturan keputusan Decision Tree dalam memetakan fitur kata yang paling berpengaruh. Luaran berupa model klasifikasi sentimen dengan akurasi tinggi untuk memetakan polarisasi opini publik secara akurat, yang dapat digunakan sebagai dasar perumusan regulasi moderasi konten serta penyusunan program edukasi ketahanan keluarga yang lebih kontekstual di Indonesia [8].

Pada pengumpulan data atau Data Selection dari data historis di media sosial X dengan menggunakan Teknik proses crawling data, preprocessing, pembobotan TF-IDF, hingga evaluasi menggunakan Multi-Label Confusion Matrix (MLCM), maka nantinya dari penelitian ini diharapkan dapat memberikan sebuah gambaran yang akurat mengenai persepsi publik serta menentukan algoritma mana yang memiliki performa terbaik dalam memetakan sentimen masyarakat terhadap isu krusial ini [9].

Pada penelitian (Purnamasari & Agoestanto, 2024), telah menjelaskan Dari Analisis yang dilakukan pada media sosial mengenai transportasi publik jaklingko, ditemukan 6.001 data dari proses crawling kemudian diperoleh 4.017 data bersih sehingga penggunaan metode Naïve Bayes dan Decision Tree dalam mengklasifikasikan sentimen positif dan negatif menghasilkan Tingkat akurasi untuk metode Naïve Bayes sebesar 84,9% dan metode Decision Tree sebesar 84,2% hal ini membuktikan kedua algoritma ini saling melengkapi [10]. Dari beberapa penelitian terdahulu yang sudah bisa membuktikan efektivitas dari penggunaan algoritma Naïve Bayes dan Decision Tree dalam analisis sentimen pada berbagai domain, dan terdapat kesenjangan yang signifikan yang Dimana dari fokus penelitian yang sebelumnya diketahui masih didominasi oleh isu dari layanan publik, transportasi, bahkan kebijakan pemerintah [11].

Pada penelitian ini bertujuan untuk melakukan analisis sentimen terhadap kasus perceraian akibat perselingkuhan yang ada di sosial media terkhusus media sosial X (twitter) dengan menggunakan algoritma Naive Bayes dan Decision Tree dengan evaluasi Multi Label Confusion Matrix. dapat memberikan gambaran yang akurat mengenai persepsi dan polarisasi opini publik, sekaligus menentukan algoritma klasifikasi terbaik yang paling efektif, efisien, dan memiliki tingkat akurasi tertinggi dalam memetakan sentimen masyarakat terhadap isu krusial tersebut.

Penerapan analisis sentimen ini juga nantinya akan diimplementasikan pada dasbor agar dapat mengetahui hasil dari analisis sentimen dengan algoritma naive bayes dan decision tree dan evaluasi multi label confusion matrix.

## METODE

## A. Alat dan Bahan

Berikut adalah daftar alat dan bahan yang dibutuhkan dalam pelaksanaan penelitian:

### 1. Perangkat Keras:

- a. Laptop ASUS Vivobook Go 14/15
- b. AMD 3020e With Radeon Graphics
- c. RAM 8 GB

### 2. Perangkat Lunak:

- a. GOOGLE Colab
- b. Bahasa Pemrograman Python
- c. Web crawling
- d. Library Sastrawi
- e. Browser Chrome
- f. Sistem Operasi Window

### 3. Bahan Data Utama:

Data tweet (media sosial x) dengan jumlah total data 3.450 tweet. Kata kunci “cerai karena selingkuh” pada periode data dari bulan April 2024 sampai November 2025

## B. Metode Penelitian



**Gambar 1** Alur Penelitian

Metodologi yang akan digunakan dalam penelitian ini dengan mengikuti Langkah-langkah yang pada umumnya ditemukan dalam data mining, yang telah dikenal sebagai tahap Knowledge Discovery in Database (KDD) yang akan digunakan proses penambangan data untuk menemukan pola dan juga informasi penting dalam data. Sehingga tahap-tahap Knowledge Discovery in Database (KDD) Adalah data Selection, data preprocessing, data transformation, data mining, interpretation/evaluation [12].

## C. Pengumpulan Data

### 1. Crawling Data

Crawling data merupakan tahap dalam penelitian yang bertujuan untuk mengumpulkan atau mengunduh data dari server twitter berupa user dan tweet beserta atribut-atributnya.

### 2. Pelabelan Data

Dari data yang diperoleh selama proses crawling, setelah dataset melalui tahap pembersihan dengan menghilangkan simbol-simbol yang tidak relevan, selanjutnya dilakukan pelabelan data untuk menentukan apakah sentimen mengarah positif, negatif, atau netral. Proses pelabelan utama dilakukan secara otomatis menggunakan RapidMiner dengan pendekatan berbasis kamus sentimen (lexicon-based labeling). Namun, sebelum proses otomatis dilakukan, sebanyak 100 data terlebih dahulu dilabeli secara manual sebagai data latih untuk membantu sistem dalam mengenali pola awal sentimen.

Penggunaan pelabelan otomatis berbasis RapidMiner dan kamus sentimen dipilih karena lebih efisien dalam menangani dataset berukuran besar, mengurangi waktu yang dibutuhkan dibandingkan pelabelan manual penuh, serta meminimalkan bias subjektif antar anotator. Pendekatan ini juga memungkinkan terbentuknya dataset berlabel secara lebih konsisten yang dapat digunakan untuk mendukung proses pelatihan dan pengujian model machine learning.

### 3. Text Processing

Tahapan ini dilakukan untuk memastikan data yang diolah memiliki kualitas yang baik. Dengan cara data tweet mentah terlebih dahulu dilakukan proses:

a. Cleaning

b. Casefolding  
Stopword removal

c. Tokenizing

d. Stemming

### 4. TF-IDF

Setelah melewati tahap preprocessing, dataset yang berupa kumpulan kata dasar (token) ditransformasikan ke dalam representasi numerik menggunakan metode pembobotan Term Frequency-Inverse Document Frequency (TF-IDF).

### 5. Synthetic Minority Over-sampling Technique (SMOTE)

Dalam tahapan pra-pemrosesan lanjutan sebelum masuk ke proses pemodelan, teknik Synthetic Minority Over-sampling Technique (SMOTE) diterapkan secara spesifik hanya pada data latih (training data) untuk mengatasi potensi masalah ketidakseimbangan kelas (imbalanced data).

### 6. Data Mining

Tahap Data Mining merupakan inti dari proses Knowledge Discovery in Database (KDD) dalam penelitian ini. Pada tahap ini, dilakukan ekstraksi pola informasi dari dataset yang telah melewati tahap preprocessing dan transformasi (TF-IDF) untuk menghasilkan model klasifikasi yang mampu membedakan opini masyarakat secara otomatis

### 7. Algoritma Naïve Bayes

Pada tahap pemodelan yang pertama, penelitian ini mengimplementasikan algoritma Naïve Bayes untuk mengklasifikasikan sentimen masyarakat berdasarkan perhitungan probabilitas statistik. Algoritma ini dipilih karena rekam jejaknya yang sangat efisien dan optimal dalam memproses dataset teks berdimensi tinggi, seperti matriks hasil pembobotan TF-IDF.

### 8. Algoritma Decision Tree

Setelah tahap pembobotan fitur selesai, langkah selanjutnya adalah membangun model klasifikasi menggunakan algoritma Decision Tree. Implementasi ini dilakukan menggunakan kelas DecisionTreeClassifier yang tersedia dalam pustaka Scikit-learn pada bahasa pemrograman Python.

### 9. Multi-Label Confusion Matrix

Penggunaan Multi-Label Confusion Matrix MLCM bertujuan untuk memberikan penilaian performa yang lebih ringkas dan tidak ambigu dibandingkan metrik agregat seperti Hamming loss atau rata-rata presisi dan recall. Metrik agregat sering kali bersifat ambigu dalam mengidentifikasi hasil negatif palsu (FN) dan positif palsu (FP) karena sifatnya yang menjumlahkan hasil secara keseluruhan tanpa memperlihatkan distribusi kesalahan antar kelas secara spesifik

### 10. Perancangan Antar Muka Dasbor

Proses perancangan antarmuka dasbor mengadopsi prinsip desain yang berpusat pada pengguna, dimulai dengan pemetaan struktur informasi yang akan ditampilkan. Sistem dasbor ini dirancang menggunakan teknologi berbasis web modern, di mana antarmuka pengguna (frontend) dibangun menggunakan pustaka React.

### 11. Evaluasi

Tahap akhir dalam metodologi Knowledge Discovery in Database (KDD) pada penelitian ini adalah melakukan evaluasi kinerja model. Evaluasi ini bertujuan untuk mengukur sejauh mana tingkat keakuratan dan efektivitas algoritma Naïve Bayes dan Decision Tree dalam memprediksi serta mengklasifikasikan sentimen masyarakat pada platform X terkait isu [ISSN 2714-7444 \(online\)](https://doi.org/10.21070/acopen.11.2026.14689), <https://acopen.umsida.ac.id>, published by [Universitas Muhammadiyah Sidoarjo](https://www.muhammadiyahsidoarjo.ac.id)

perceraian akibat perselingkuhan.

HASIL DAN PEMBAHASAN

A. Pengumpulan Data

1. Crawling Data

Pada Tahap awal dari hasil penelitian ini adalah pemaparan mengenai proses pengumpulan data (data crawling) yang diambil dari platform X (sebelumnya Twitter). Proses ekstraksi data dilakukan dengan memanfaatkan library Tweet Harvest guna menjangkau tweet berbahasa Indonesia yang secara spesifik relevan dengan topik perceraian akibat perselingkuhan. Adapun batasan waktu penarikan data yang ditetapkan dalam penelitian ini mencakup unggahan pada rentang periode 1 April 2024 hingga 1 November 2025. Berdasarkan hasil eksekusi crawling dengan parameter tersebut, penelitian ini berhasil mengumpulkan dataset mentah sebanyak 3.450 tweet. Seluruh data yang diperoleh ini menjadi himpunan data primer yang selanjutnya akan digunakan untuk tahapan pelabelan dan pemrosesan data (preprocessing)



Gambar 2 Hasil Crawling Data

Pada gambar 2 merupakan hasil dari tahap awal dari pengumpulan data dalam melakukan penelitian untuk analisis sentimen kasus perceraian akibat perselingkuhan di media sosial X yang menghasilkan menghasilkan 3.450 data tweet.

2. Pelabelan Data

Penelitian ini diawali dengan tahap pengumpulan data yang menghasilkan 3.450 data tweet mentah dari hasil crawling menggunakan kata kunci "cerai karena selingkuh". Dari data mentah tersebut, tahapan pertama yang dilakukan adalah proses pelabelan otomatis menggunakan metode Lexicon-Based (berbasis kamus). Pelabelan ini sengaja dilakukan di awal guna menjaga keaslian konteks data sebelum ada karakter yang dihilangkan.

Berbeda dengan penggunaan VADER pada umumnya, proses pelabelan ini bekerja dengan mencocokkan token pada setiap tweet dengan daftar kata positif dan negatif yang telah didefinisikan secara spesifik untuk kasus perceraian. Hasil pelabelan menunjukkan bahwa opini publik didominasi oleh kategori Netral sebanyak 2.116 data, diikuti oleh kategori Negatif 1.255 data, dan kategori Positif 79 data.

Setelah seluruh data berhasil diberi label, barulah dataset tersebut melalui tahapan preprocessing, seperti pembersihan data (data cleaning) dan proses lanjutan lainnya. Tahapan preprocessing ini dilakukan agar algoritma selanjutnya dapat memproses pola sentimen secara lebih akurat tanpa adanya gangguan karakter atau noise yang tidak perlu.

B. Text Preprocessing

Setelah melakukan pelabelan data maka selanjutnya Langkah untuk melakukan text preprocessing yaitu cleaning, case folding, stopword removal, tokenizing dan stemming.

1. Cleaning

Tabel 1 Hasil Cleaning

Table with 3 columns: Sebelum, Cleaning, and Label. It shows two examples of text cleaning where characters like @ and # are removed and the text is normalized.

## 2. Casefolding

Adalah proses penyeragaman seluruh karakter huruf pada dataset menjadi huruf kecil (lowercase). Hal ini bertujuan agar sistem komputer membaca kata yang sama tanpa terpengaruh oleh kapitalisasi, misalnya kata "Selingkuh", "SELINGKUH", dan "selingkuh" akan dianggap sebagai satu entitas kata yang sama. Hasil Casefolding dapat dilihat pada tabel 2:

**Tabel 2 Hasil Caefolding**

Sebelum	Casefolding	Label
punya driver grab langganan karena kerja dulu di bandara rumah di bogor dr rumah ke damri naik grab karena bawa koper dituduh selingkuh ama suami ga pernah lagi sama driver yg itu Ganti kerjaan dituduh selingkuh ama kolega kantor cerai lah sudah cape	punya driver grab langganan karena kerja dulu di bandara rumah di bogor dr rumah ke damri naik grab karena bawa koper dituduh selingkuh ama suami ga pernah lagi sama driver yg itu ganti kerjaan dituduh selingkuh ama kolega kantor cerai lah sudah cape	Negatif
Belum ada info resmi atau nama yang disebut soal pacar Baskara selain Sherina dulu Cerai mereka tahun lalu memang karena sibuk kerja amp pertengkaran biasa bukan selingkuh kata pengadilan tapi repost Sherina hari ini bikin rumor selingkuh lagi rame Gak ada detail pihak	belum ada info resmi atau nama yang disebut soal pacar baskara selain sherina dulu cerai mereka tahun lalu memang karena sibuk kerja amp pertengkaran biasa bukan selingkuh kata pengadilan tapi repost sherina hari ini bikin rumor selingkuh lagi rame gak ada detail pihak	Negatif

## 3. Stopword Removal

Tahap ini bertujuan untuk membuang kata-kata umum yang sering muncul namun tidak memiliki makna spesifik atau bobot sentimen (stopword). Kata hubung dan kata penunjuk seperti "yang", "dan", "di", "ke", "dari", "ini", dan "itu" dihapus dari dataset agar mengurangi dimensi data dan mempercepat proses komputasi. Hasil Stopword Removal dapat dilihat pada tabel 3:

**Tabel 3 Hasil Stopword Removal**

Sebelum	Stopword Removal	Label
belum,ada,info,resmi,atau,nama,yang,disebut ,soal,pacar,baskara,selain,sherina,dulu,cerai, mereka,tahun,lalu,memang,karena,sibuk,kerj a,amp,pertengkaran,biasa,bukan,selingkuh,ka ta,pengadilan,tapi,repost,sherina,hari,ini,biki n,rumor,selingkuh,lagi,rame,gak,ada,detail,pi hak	info,resmi,nama,pacar,baskara,sherina,cerai,sibuk, kerja,amp,pertengkaran,selingkuh,pengadilan,repo st,sherina,bikin,rumor,selingkuh,rame,gak,detail	Negatif
ternyata,baskara,sama,sherina,cerai,karena,b askara,yang,selingkuh,astaga,gak,nyangka	baskara,sherina,cerai,baskara,selingkuh,astaga,gak ,nyangka	Negatif

## 4. Tokenizing

Teks kalimat yang sebelumnya berbentuk untaian karakter (string) utuh akan dipecah menjadi potongan kata tunggal yang berdiri sendiri, atau yang disebut sebagai token. Pemecahan ini dilakukan dengan menggunakan spasi sebagai karakter pemisah (delimiter). Hasil Tokenizing dapat dilihat pada tabel 4:

**Tabel 4 Hasil Tokenizing**

Sebelum	Tokonenizing	Label
bener dari kecil selalu jadi saksi ibuk bapak berantem udah gede lihat rumah tangganya tante yang hampir hancur dan udah nyaris mau cerai tapi gak jadi terus ada kerabat yang akhirnya cerai juga karena suaminya selingkuh	bener,dari,kecil,selalu,jadi,saksi,ibuk,bapak,berant em,udah,gede,lihat,rumah,tangganya,tante,yang,h ampir,hancur,dan,udah,nyaris,mau,cerai,tapi,gak,j adi,terus,ada,kerabat,yang,akhirnya,cerai,juga,kare na,suaminya,selingkuh	Negatif
siapa bilang cerai haram yg kubilang dosa kan menceraikan tanpa sebab yg jelas dlm islam itu termasuk menzhalimi pasangan dan itu dosa besar kalau cerai karena hal yg masuk akal misal pasangan selingkuh itu sah aja kan pembahasan dari tadi tiba ceraiin pasangan tanpa sebab	siapa,bilang,cerai,haram,yg,kubilang,dosa,kan,men ceraikan,tanpa,sebab,yg,jelas,dlm,islam,itu,termas uk,menzhalimi,pasangan,dan,itu,dosa,besar,kalau, cerai,karena,hal,yg,masuk,akal,misal,pasangan,seli ngkuh,itu,sah,aja,kan,pembahasan,dari,tadi,tiba,ce raiin,pasangan,tanpa,sebab	Negatif

## 5. Stemming

Tahap terakhir dari preprocessing adalah stemming, yaitu proses mengembalikan setiap kata yang memiliki imbuhan (awalan, sisipan, atau akhiran) menjadi bentuk kata dasarnya. Menggunakan library Sastrawi, kata berimbuhan seperti

"diselingkuhi" diubah menjadi kata dasar "selingkuh", dan "berhubungan" menjadi "hubung". Hasil Tokenizing dapat dilihat pada tabel 5:

**Tabel 5 Hasil Stemming**

Sebelum	Tokonenizing	Label
aneh,suami,cerai,selingkuh,wanita,menikah,suami,wanita,selingkuan,umur,kayaknya,palah,tua,muda	aneh suami cerai selingkuh wanita meni suami wanita selingkuh umur kayak palah tua mjda	Negatif
keliatan,adem,ayem,ributnya,ditutupin,kak,malu,hasil,ngerebut,berujung,berantem,cerai,tetangga,sdh,nikah,selingkuh,suami,orang,dipukulin,suaminya,tpi,bertahan,malu,cerai	liat adem ayem ribut ditutupin kak malu hasil ngerebut ujung antem cerai tetangga sdh nikah selingkuh suami orang dipukulin suami tpi tahan malu cerai	Negatif

### C.Membagi Data Latih dan Uji

Memasuki Tahap selanjutnya adalah pembagian data yang bertujuan untuk memisahkan dataset yang telah diproses menjadi dua subset, yaitu data latih (training data) dan data uji (testing data). Data latih digunakan oleh algoritma untuk mempelajari pola dan hubungan antar fitur dengan label sentimen, sedangkan data uji berfungsi untuk mengevaluasi kinerja model pada data yang belum pernah dilihat sebelumnya guna mendapatkan estimasi kemampuan generalisasi model yang akurat. Berdasarkan hasil pemrosesan pada notebook, dataset dibagi dengan proporsi 80% untuk data latih dan 20% untuk data uji. Dari total 3.273 sampel data yang siap diolah, diperoleh jumlah data latih sebanyak 2.618 sampel dan data uji sebanyak 655 sampel. Distribusi label pada data latih terdiri dari 1.594 data Netral, 966 data Negatif, dan 58 data Positif, yang akan menjadi dasar bagi model dalam mengenali karakteristik setiap kelas sentiment [13].

#### 1. TF-IDF

Dari Tahap ekstraksi fitur dalam penelitian ini dilakukan menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) untuk mentransformasi data teks kualitatif yang telah dibersihkan menjadi representasi numerik. Proses ini dikonfigurasi dengan parameter `max_features` sebanyak 5.000 kata unik, yang bertujuan untuk membentuk matriks pembobotan berdasarkan tingkat signifikansi setiap kata terhadap dokumen dalam dataset. Berdasarkan hasil eksekusi program, kata-kata seperti "selingkuh", "cerai", serta entitas terkait seperti "inara", "rusli", dan "virgoun" teridentifikasi sebagai fitur dengan akumulasi bobot tertinggi. Data numerik hasil pembobotan TF-IDF inilah yang kemudian menjadi input utama bagi algoritma machine learning untuk mempelajari pola distribusi sentimen dan melakukan prediksi klasifikasi secara otomatis pada tahap selanjutnya.

Berikut adalah 10 kata yang memiliki total akumulasi nilai bobot TF-IDF paling besar di dalam dataset Anda:

**Tabel 6 Top 5 Kata Berdasarkan Pembobotan TF-IDF:**

kata	Jumlah muncul	Skor TF-IDF
selingkuh	4207	0.051541
Cerai	702	0.038833
manfaat	2378	0.018928
zina	4992	0.000817
kecewa	1736	0.000483

#### 2. Penerapan Synthetic Minority Over-sampling Technique (SMOTE)

Berdasarkan pembagian data latih (training data) sebelumnya, diketahui bahwa distribusi kelas sangat tidak seimbang (imbalanced data), di mana sentimen Netral mendominasi dengan 1.594 sampel, diikuti oleh sentimen Negatif sebanyak 966 sampel, dan sentimen Positif yang sangat minoritas dengan hanya 58 sampel. Untuk mengatasi ketimpangan tersebut, teknik SMOTE (Synthetic Minority Over-sampling Technique) diterapkan secara khusus pada data latih. Perlu digarisbawahi bahwa SMOTE tidak sekadar menduplikasi atau menyalin ulang data minoritas yang sudah ada. Sebaliknya, metode ini bekerja secara lebih cerdas dengan menganalisis karakteristik kedekatan antar fitur pada data kelas minoritas, lalu menyintesis (menciptakan) titik-titik data buatan baru yang secara logis memiliki kemiripan pola dengan data aslinya.

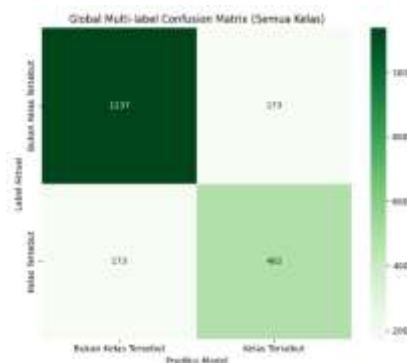
Setelah proses SMOTE berhasil dieksekusi, jumlah sampel pada kelas sentimen Negatif dan Positif ditingkatkan hingga setara dengan proporsi sampel pada kelas mayoritas (Netral). Dengan demikian, distribusi data latih pasca-SMOTE menjadi seimbang, yaitu 1.594 sampel untuk sentimen Netral, 1.594 sampel untuk sentimen Negatif, dan 1.594 sampel untuk sentimen Positif. Total keseluruhan data latih yang awalnya berjumlah 2.618 sampel meningkat menjadi 4.782 sampel data latih sintesis.

Peran spesifik SMOTE dalam tahapan ini sangatlah krusial untuk proses pembelajaran (training). Jika algoritma dilatih menggunakan data yang timpang, model akan cenderung bias; ia akan lebih sering menebak data baru sebagai sentimen Netral hanya karena label tersebut paling sering ia kenali selama masa pelatihan, dan mengabaikan sentimen Positif. Dengan adanya penyeimbangan proporsi melalui data sintesis SMOTE, ruang pemisah keputusan (decision boundary) pada algoritma Naïve Bayes dan Decision Tree dipaksa untuk ikut mengenali karakteristik dan pola kata dari sentimen minoritas.

Hasilnya, model tidak lagi didominasi oleh kelas mayoritas dan mampu melakukan klasifikasi terhadap ketiga kategori sentimen tersebut secara lebih adil, objektif, dan akurat [14].

### 3. Algoritma Naïve Bayes

Pada tahap ini, dilakukan pembuatan model menggunakan algoritma Multinomial Naïve Bayes terhadap dataset perceraian akibat perselingkuhan yang telah diekstraksi fiturnya menggunakan TF-IDF. Hasil pengujian menunjukkan bahwa model mampu mencapai tingkat akurasi sebesar 73.59%, yang mengindikasikan performa yang cukup baik dalam mengklasifikasikan opini publik di platform X. Berdasarkan analisis melalui confusion matrix dan classification report, model menunjukkan kemampuan prediksi yang sangat kuat pada kelas Netral dengan skor F1-score sebesar 0,84, diikuti oleh kelas Negatif dengan F1-score 0,71. Namun, model mengalami kesulitan dalam mengenali sentimen Positif (F1-score 0,00) dikarenakan adanya ketidakseimbangan data (imbalanced data), di mana jumlah sampel positif jauh lebih sedikit dibandingkan kelas lainnya. Secara kontekstual, hasil ini mencerminkan bahwa meskipun algoritma Naïve Bayes efektif dalam membedakan antara informasi netral dan sentimen negatif yang mendominasi dataset, keterbatasan jumlah data pada kelas minoritas menjadi tantangan dalam proses generalisasi model secara menyeluruh. Sehingga algoritma berbasis probabilitas ini secara otomatis memberikan bobot yang jauh lebih besar pada kategori negatif tersebut [15].



**Gambar 3** Heatmap Naïve Bayes

Gambar 3 tersebut memvisualisasikan hasil evaluasi kinerja agregat algoritma Naïve Bayes melalui Global Multi-label Confusion Matrix dalam mengklasifikasikan sentimen terkait perceraian ke dalam tiga kelas (Negatif, Netral, dan Positif). Dengan menerapkan pendekatan One-vs-Rest (OvR), matriks ini merangkul total pengujian dari 655 data uji ke dalam format evaluasi biner. Berdasarkan matriks tersebut, model berhasil melakukan prediksi yang tepat (true prediction) sebagaimana ditunjukkan oleh akumulasi nilai pada blok diagonal utamanya, yakni memprediksi secara akurat 1.137 pengujian sebagai "Bukan Kelas Tersebut" (True Negative) dan 482 pengujian sebagai "Kelas Tersebut" (True Positive). Meskipun demikian, pola sebaran angka pada matriks ini secara eksplisit menunjukkan adanya dampak dari ketidakseimbangan data (imbalanced data). Tingkat kesalahan klasifikasi (misclassification) pada model ini masih tergolong signifikan, di mana sistem mencatatkan 173 kesalahan tebak palsu (False Positive) pada blok kanan atas dan 173 kegagalan dalam mengenali kelas aktual (False Negative) pada blok kiri bawah. Distribusi kesalahan sebesar 173 data ini mengindikasikan bahwa pemodelan Naïve Bayes yang dibangun cukup kesulitan dan kurang sensitif dalam mengenali kelas minoritas (Positif), serta rentan mengalami bias prediksi akibat kemiripan bobot probabilitas fitur kata yang sangat didominasi oleh kelas mayoritas. Visualisasi ini merangkul total tebakan algoritma Multinomial Naïve Bayes Anda. Matriks ini membandingkan Label Aktual (kenyataan sebenarnya pada data, sumbu Y) dengan Prediksi Model (tebakan algoritma, sumbu X) [16].

Berikut adalah arti dari keempat kotak tersebut:

a.1137 (Kiri Atas - True Negative): Ini adalah Tebakan Benar. Model berhasil mengenali dengan tepat 1.137 data yang memang aslinya bukan bagian dari kelas yang sedang dievaluasi, dan model menebaknya "Bukan Kelas Tersebut". Warna hijau gelap menunjukkan mayoritas tebakan benar terkonsentrasi di sini, yang sejalan dengan dominasi kelas mayoritas (Netral/Negatif) pada data Anda.

b.482 (Kanan Bawah - True Positive): Ini juga Tebakan Benar. Model berhasil mengenali 482 data sesuai dengan kelas aslinya.

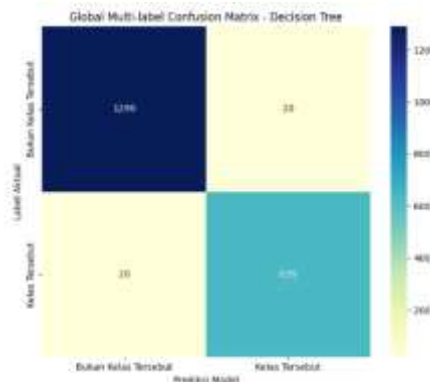
c.173 (Kanan Atas - False Positive): Ini adalah Tebakan Salah. Model keliru menebak 173 data masuk ke "Kelas Tersebut", padahal di dunia nyata data tersebut "Bukan Kelas Tersebut".

d.173 (Kiri Bawah - False Negative): Ini juga Tebakan Salah. Model melewatkan atau gagal mengenali 173 data yang aslinya adalah "Kelas Tersebut", dan malah menebaknya sebagai "Bukan Kelas Tersebut".

## 4. Algoritma Decision Tree

Pada tahap pengujian, algoritma Decision Tree digunakan untuk mengklasifikasikan sentiment terkait itu perceraian akibat perselingkuhan dan menghasilkan akurasi sebesar 96,95%. Berdasarkan classification report, model menunjukkan kinerja yang baik pada kelas Netral dengan F1-score 0,79 dan kelas Negatif dengan F1-score 0,70.

Namun, performa model pada kelas Positif masih rendah dengan F1-score 0,16. Temuan ini menunjukkan bahwa meskipun Decision Tree mampu mengidentifikasi pola kata yang spesifik, ketidakseimbangan data masih memengaruhi kemampuan model dalam mengenali kelas sentiment yang jumlah datanya terbatas [17].



**Gambar 4** Heatmap Decision Tree

Gambar 4 tersebut menampilkan visualisasi Multi-label Confusion Matrix yang digunakan untuk mengevaluasi kinerja agregat dari algoritma Decision Tree dalam mengklasifikasikan sentiment publik terkait perceraian ke dalam tiga kategori: Negatif, Netral, dan Positif. Menggunakan pendekatan One-vs-Rest (UvR), matriks ini merangkum total evaluasi dari 655 data uji ke dalam representasi biner. Model menunjukkan performa klasifikasi yang sangat superior dengan tingkat presisi yang tinggi. Hal ini dibuktikan secara eksplisit oleh tingginya angka prediksi tepat (true prediction) pada blok warna diagonal utama yang pekat, yakni model berhasil memprediksi secara akurat 1.290 pengujian sebagai “Bukan Kelas Tersebut” (True Negative) dan 635 pengujian sebagai “Kelas Tersebut” (True Positive).

Berkat ketelitian struktur aturan yang tajam pada algoritma ini, tingkat kesalahan prediksi (misclassification) dapat ditekan seminimal mungkin, di mana model hanya mencatatkan 20 kesalahan tebak palsu (False Positive) dan 20 kesalahan terlewat (False Negative). Secara keseluruhan, sebaran nilai agregat pada matriks ini mengonfirmasi keandalan dan stabilitas model Decision Tree dalam menangani ketidakseimbangan data sehingga mampu memetakan opini publik dengan sangat akurat tanpa mengalami bias terhadap kelas mayoritas. Matriks ini membandingkan Label Aktual (kenyataan sebenarnya pada data uji, sumbu vertikal) dengan Prediksi Model (tebakan algoritma Decision Tree, sumbu horizontal).

Angka-angka di dalamnya menunjukkan seberapa cerdas model Anda:

a.1290 (Kiri Atas - True Negative): Ini adalah Tebakan Benar. Model sangat sukses mengenali 1.290 data yang memang aslinya bukan bagian dari kelas yang sedang dievaluasi, dan model menebaknya dengan tepat. Warna biru tua yang pekat menjadi indikator visual bahwa model sangat menguasai area prediksi ini.

b.635 (Kanan Bawah - True Positive): Ini juga Tebakan Benar. Model berhasil mengenali 635 data dengan tepat sesuai kelas sentimen aslinya.

c.20 (Kanan Atas - False Positive): Ini adalah Tebakan Salah. Model hanya melakukan 20 kekeliruan dengan menebak suatu data sebagai "Kelas Tersebut", padahal di kenyataannya bukan.

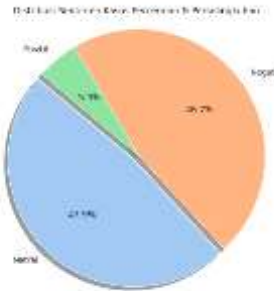
d.20 (Kiri Bawah - False Negative): Ini juga Tebakan Salah. Model hanya melewatkan 20 data yang seharusnya masuk ke "Kelas Tersebut", namun meleset dan ditebak bukan.

## 5. Multi-Label Confusion Matrix (MLCM)

Penggunaan Global Multi-label Confusion Matrix dalam penelitian ini dipilih karena mengadopsi pendekatan One-vs-Rest (OvR) yang mengagregasi performa seluruh kelas sentiment (Positif, Netral, dan Negatif) menjadi satu representasi biner tunggal yang komprehensif. Berbeda dengan confusion matrix multikelas standar yang menampilkan matriks persilangan 3x3 yang kompleks, agregasi 2x2 ini menjumlahkan seluruh metrik dasar untuk mengevaluasi kemampuan model dalam membedakan "Kelas Tersebut" dengan "Bukan Kelas Tersebut" secara keseluruhan. Pendekatan ini sangat krusial untuk menghasilkan perhitungan evaluasi berbasis Micro-Averaging, seperti akurasi, presisi, recall, dan khususnya F1-Score, yang secara efektif dapat menangani potensi masalah ketidakseimbangan jumlah data (imbalanced data) pada ekstraksi sentimen opini publik terkait perceraian, sehingga hasil evaluasi akhir model menjadi lebih objektif dan tidak bias terhadap kelas mayoritas [18].

Berdasarkan analisis Multilabel Confusion Matrix, performa model berbeda pada setiap kelas sentiment akibat ketidakseimbangan distribusi data. Algoritma Decision Tree memperoleh akurasi 96,95% dengan kemampuan yang baik

dalam mengklasifikasikan sentrimen Netral (429 prediksi benar) dan Negatif (56 prediksi benar). Sementara itu, Naïve Bayes mencapai akurasi 73,58% dengan mayoritas prediksi benar berada pada kelas Netral (281 data) dan Negatif (196 data). Meskipun keduanya menunjukkan kinerja yang baik pada kelas mayoritas, jumlah False Negative yang tinggi pada kelas Positif menunjukkan bahwa model masih kesulitan mengenali sentiment tersebut. Temuan ini mengindikasikan bahwa dominasi data Netral dan Negatif menyebabkan model cenderung memprediksi ke dua kelas tersebut dibandingkan kelas Positif.



Gambar 5 Pie Chart Penelitian

Gambar 5 yang menampilkan diagram lingkaran (pie chart) tersebut memvisualisasikan persentase distribusi sentimen publik di media sosial terkait kasus perceraian akibat perselingkuhan dari total keseluruhan dataset yang telah diolah. Berdasarkan diagram tersebut, opini masyarakat secara jelas didominasi oleh dua kategori utama dengan proporsi yang hampir setara, yakni sentimen Netral sebesar 47,9% dan sentimen Negatif sebesar 46,7%. Tingginya persentase sentimen Negatif secara logis mencerminkan banyaknya kecaman, luapan amarah, dan kekecewaan publik terhadap fenomena perselingkuhan, sedangkan tingginya sentimen Netral mengindikasikan banyaknya pengguna yang sekadar membagikan tautan berita, fakta, atau informasi objektif tanpa keberpihakan emosional. Sebaliknya, sentimen Positif berada pada proporsi yang sangat minim, yaitu hanya sebesar 5,3%; kondisi ini tidak hanya menunjukkan sedikitnya opini publik yang secara eksplisit memberikan dukungan moral atau motivasi, tetapi juga menjadi bukti visual yang kuat untuk mengonfirmasi adanya fenomena ketidakseimbangan kelas (imbalanced data) yang menjadi tantangan utama bagi algoritma machine learning pada fase pemodelan di penelitian ini.

## 6. Implementasi Dasbor Aktif

Hasil implementasi menunjukkan dasbor telah berhasil menyajikan data dari 3.450 cuitan yang telah diklasifikasikan ke dalam kategori netral, negatif, dan positif secara visual. Antarmuka dasbor menampilkan grafik perbandingan akurasi di mana Decision Tree menunjukkan performa unggul sebesar 96,95% dibandingkan Naïve Bayes yang memperoleh 73,59%. Selain itu, fitur simulasi pengujian teks dalam dasbor terbukti mampu menjalankan logika klasifikasi secara akurat, memberikan gambaran langsung kepada pengguna mengenai cara model memproses kata kunci seperti "selingkuh" atau "cerai" menjadi label sentimen tertentu.

Hasil implementasi antarmuka menunjukkan bahwa dasbor telah berhasil dibangun dan menyajikan data dari 3.450 cuitan yang telah diklasifikasikan ke dalam kategori netral, negatif, dan positif secara visual. Pengembangan dasbor ini memanfaatkan komponen visual dari pustaka Recharts dan elemen ikonografi dari Lucide-React untuk menghasilkan tampilan yang komunikatif dan responsive

Berikut adalah penjabaran hasil implementasi fitur pada dasbor:

a. Visualisasi Distribusi Sentimen: Dasbor secara otomatis merender pie chart interaktif yang menggambarkan hasil akhir dari opini publik terkait fenomena perceraian akibat perselingkuhan. Grafik ini mempermudah pengguna untuk melihat secara instan dominasi sentimen Netral dan Negatif dibandingkan sentimen Positif.

b. Komparasi Performa Model: Antarmuka dasbor menampilkan grafik perbandingan akurasi dalam bentuk diagram batang (bar chart). Komponen visual pada dasbor menunjukkan bahwa Decision Tree memiliki performa lebih unggul dengan akurasi 96,95%, dibandingkan Naïve Bayes yang mencapai 73,59%.

c. Fitur simulasi klasifikasi memungkinkan pengguna menguji teks secara langsung dan memperoleh prediksi sentimen secara real-time. Melalui fitur ini, pengguna dapat melihat bagaimana model mengolah kata kunci tertentu menjadi label sentimen, sehingga hasil machine learning dapat dimanfaatkan dengan mudah tanpa memerlukan kemampuan pemrograman [19].



**Gambar 6** Tampilan Halaman Utama Dasbor Analisis Sentimen

Gambar 6 di atas menampilkan halaman utama (beranda) dari dasbor visualisasi hasil analisis sentimen mengenai studi kasus perceraian akibat perselingkuhan. Pada halaman ini, disajikan dua metrik ringkasan yang paling krusial dari hasil penelitian:

- 1) Panel Distribusi Data Sentimen: Menggunakan representasi visual berupa diagram donat (donut chart), panel ini memberikan gambaran instan kepada pengguna mengenai proporsi opini publik. Secara visual, dapat dilihat dominasi blok warna abu-abu (Netral) dan merah (Negatif), sementara blok warna hijau (Positif) memiliki proporsi yang sangat kecil, merepresentasikan ketidakseimbangan kelas (imbalanced data) dari opini masyarakat.
- 2) Panel Akurasi Naïve Bayes vs Decision Tree: Menggunakan representasi diagram batang (bar chart), panel ini secara tegas mengomunikasikan hasil komparasi performa kedua algoritma penambangan data (data mining). Diagram batang secara visual membuktikan keunggulan signifikan dari algoritma Decision Tree (blok biru muda) dibandingkan algoritma Naïve Bayes (blok abu-abu) yang tercatat berada di angka 73,59%. Visualisasi ini menjawab tujuan utama penelitian secara langsung dan mudah dipahami.



**Gambar 7** Tampilan Visualisasi Heatmap Confusion Matrix pada Dasbor

Gambar 7 di atas memperlihatkan penjabaran detail dari evaluasi kinerja model melalui visualisasi Heatmap Multi-Label Confusion Matrix (MLCM) untuk kedua algoritma.

- 1) Pada sisi kiri, disajikan matriks untuk algoritma Naïve Bayes (Akurasi 73,59%). Dapat diamati bahwa model ini cukup kuat dalam mendeteksi kelas Netral (281 data tepat), namun mengalami kesulitan dalam mengenali sentimen Positif karena bias data minoritas.
- 2) Pada sisi kanan, matriks Decision Tree (Akurasi 96,95%) menunjukkan performa yang sangat superior. Hal ini dibuktikan dengan blok warna biru yang sangat pekat pada diagonal utama (seperti 222 prediksi tepat untuk Negatif dan 399 untuk Netral), yang menandakan tingkat presisi yang tinggi dan minimnya kesalahan prediksi (False Positives/False Negatives).

Selain grafik matriks, antarmuka ini juga menyertakan ringkasan analitik berbasis teks di bagian bawah setiap matriks. Teks tersebut berfungsi untuk mengedukasi pengguna dasbor mengenai keunggulan analitis dari masing-masing algoritma, seperti kecepatan komputasi pada Naïve Bayes dan ketelitian struktur aturan yang tajam pada Decision Tree.



**Gambar 8** Tampilan Simulasi Preprocessing dan Bobot TF-IDF pada Dasbor

Gambar 8 di atas mendemonstrasikan transparansi proses "di balik layar" (backend) dari sistem analisis sentimen yang disajikan ke dalam bentuk antarmuka pengguna (frontend). Panel ini terbagi menjadi dua bagian edukatif:

- 1) Panel Simulasi Preprocessing (Kiri): Modul ini memperlihatkan perjalanan transformasi teks secara bertahap. Teks cuitan mentah (Full Text) diperagakan saat melewati tahap Cleaning (pembersihan karakter), Casefolding (penyeragaman huruf kecil), hingga Stemming (pemecahan menjadi bentuk kata dasar dalam format array/list). Fitur "Ganti Contoh Data" memungkinkan pengguna untuk melihat simulasi ini pada berbagai variasi teks cuitan.
- 2) Panel Top Keywords / Bobot TF-IDF (Kanan): Panel ini mengekstraksi dan memvisualisasikan fitur kata yang memiliki pengaruh paling besar terhadap hasil klasifikasi sentimen berdasarkan perhitungan Term Frequency-Inverse Document Frequency (TF-IDF). Kata-kata kunci pembentuk konteks seperti "Selingkuh" (bobot 0.0515), "Cerai" (0.0388), dan "Manfaat" ditampilkan secara berurutan beserta representasi visual diagram batang (progress bar) untuk mengindikasikan tingkat signifikansi kata tersebut dalam memengaruhi keputusan model klasifikasi.



**Gambar 9** Tampilan Antarmuka Eksplorasi data Cuitan pada Dasbor

Gambar 9 di atas merepresentasikan antarmuka fitur Eksplorasi Data Cuitan pada dasbor interaktif yang telah dibangun. Panel ini secara transparan menampilkan hasil akhir dari proses pelabelan data. Terdapat daftar 3.273 baris cuitan (tweet) yang telah melalui tahap preprocessing dan masing-masing telah diklasifikasikan ke dalam label sentimen spesifik yang ditandai dengan badge warna (Positif dengan warna hijau, Netral dengan warna abu-abu, dan Negatif yang tidak terlihat di tangkapan layar ini namun tersedia di dalam sistem).

Untuk memudahkan analisis data, dasbor dilengkapi fitur pencarian dan penyaringan berdasarkan kategori sentimen (Semua, Positif, Negatif, dan Netral). Fitur ini membantu pengguna menampilkan data secara spesifik serta memvalidasi hasil, prediksi model dengan lebih mudah.

Selain sebagai media visualisasi hasil penelitian, dasbor interaktif ini menunjukkan penerapan analitik data secara praktis dalam mendukung pengambilan Keputusan. Integrasi grafik distribusi sentimen, confusion matrix, simulasi preprocessing, dan eskplorasi data cuitan memungkinkan pengguna memahami proses analisis sentiment secara lebih menyeluruh.

Dasbor tidak hanya menyajikan hasil akhir berupa nilai akurasi dan label sentimen, tetapi juga memberikan transparansi terhadap proses pengolahan data dan pembentukan model klasifikasi. Dengan demikian, sistem yang dibangun mampu menjembatani kebutuhan antara analisis data yang bersifat teknis dengan penyajian informasi yang mudah dipahami oleh pengguna nonteknis, sehingga hasil penelitian dapat dimanfaatkan secara lebih luas dalam kegiatan pemantauan opini publik maupun kajian sosial berbasis media digital.

#### D. Evaluasi

Tahap pengujian menunjukkan bahwa penerapan Synthetic Minority Over-sampling Technique (SMOTE) telah berfungsi krusial dalam mengatasi ketidakseimbangan distribusi data pada fase pelatihan. Berdasarkan proses tersebut, algoritma Decision Tree berhasil mencapai tingkat akurasi tertinggi sebesar 96,95%, mengungguli Multinomial Naïve Bayes yang hanya memperoleh akurasi 73,59%. Evaluasi kinerja model ini diukur secara komprehensif menggunakan Global Multi-label Confusion Matrix (MLCM). Keunggulan utama MLCM dalam penelitian ini terletak pada penerapan pendekatan One-vs-Rest (OvR) yang mengevaluasi setiap kelas secara biner, sehingga hasil pengukuran tidak terlalu dipengaruhi oleh dominasi kelas mayoritas. Hasil analisis menunjukkan bahwa Decision Tree memiliki performa yang konsisten dengan nilai F1Score sebesar 0,98 pada kelas Netral, 0,97 pada kelas Negatif, dan 0,85 pada kelas Positif. Sebaliknya, Naïve Bayes menunjukkan keterbatasan signifikan dalam mengenali sentimen minoritas Positif (F1-Score 0,19). Berbeda dengan matriks evaluasi Naïve Bayes yang menyumbang 173 kesalahan pada indikator salah prediksinya, Decision Tree mampu menekan angka kesalahan menjadi hanya 20.

Keunggulan signifikan algoritma Decision Tree ini disebabkan oleh perbedaan mendasar dalam memproses fitur teks. Decision Tree terbukti jauh lebih superior karena kemampuannya membangun struktur aturan keputusan yang tegas berdasarkan kata-kata spesifik hasil pembobotan TF-IDF (seperti "selingkuh" dan "cerai"), sehingga menghasilkan batas klasifikasi yang akurat dan stabil. Mengingat pengujian ini dilakukan setelah data diseimbangkan, tingkat kesalahan yang sangat kecil tersebut menjadi bukti kuat bahwa Decision Tree tidak goyah oleh masalah imbalanced data, sehingga mampu mengangkat nilai F1-Score secara drastis untuk seluruh kategori, termasuk kelas Positif. Sebaliknya, algoritma Naïve Bayes yang berlandaskan perhitungan probabilitas statistik rentan mengalami bias prediksi; dominasi kelas sentimen Netral dan Negatif pada dataset awal membuat model probabilitas ini secara otomatis memberikan bobot tebakan yang jauh lebih besar pada kelas mayoritas, sehingga tetap kesulitan dan kurang sensitif dalam mengidentifikasi sentimen kelas minoritas (Positif). Oleh karena itu, Decision Tree dipilih sebagai model terbaik karena kemampuannya yang lebih akurat dan menyeluruh dalam menangkap pola teks terkait isu perceraian dan perselingkuhan di platform X. Selanjutnya, model terbaik ini berhasil diimplementasikan ke dalam sebuah dasbor interaktif [20]. Pengaruh utama dari pembuatan dasbor ini adalah kemampuannya menerjemahkan hasil komputasi model algoritma yang rumit menjadi visualisasi data yang dinamis, interaktif, dan mudah dipahami oleh pengguna secara langsung. Secara keseluruhan, keuntungan dari penelitian ini adalah terciptanya sebuah instrumen analisis sentimen yang cerdas dan valid, yang dapat dimanfaatkan oleh berbagai pihak sebagai rujukan objektif dalam memetakan serta memahami pergeseran persepsi masyarakat terhadap dampak sosial perceraian di era digital.

## KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa analisis sentimen opini publik di platform X terkait isu perceraian akibat perselingkuhan menunjukkan dominasi sentimen Netral (2.116 data) dan Negatif (1.255 data), sementara sentimen Positif hanya berjumlah 79 data. Dalam pengujian model klasifikasi, algoritma Decision Tree terbukti memberikan performa yang jauh lebih unggul dengan tingkat akurasi mencapai 96,95% dan kemampuan generalisasi yang

baik di seluruh kelas sentimen dibandingkan dengan algoritma Multinomial Naïve Bayes yang hanya memperoleh akurasi 73,59%. Keberhasilan Decision Tree dalam mengenali kata kunci spesifik, seperti “selingkuh”, “cerai”, dan nama toko terkait melalui pembobotan TF-IDF, menunjukkan bahwa metode ini efektif untuk menganalisis teks media sosial yang berfokus pada topik tertentu. Namun, ketidakseimbangan data masih memengaruhi kinerja model pada kelas Positif sehingga diperlukan pengembangan lebih lanjut untuk meningkatkan sensitivitasnya. Selain itu, dasbor interaktif yang dikembangkan berhasil menyajikan hasil analisis sentimen secara dinamis dan informatif. Dasbor mampu menggambarkan dominasi sentimen Netral dan Negatif serta menunjukkan keunggulan Decision Tree sebagai model terbaik. Temuan ini membuktikan bahwa integrasi Decision Tree dengan antarmuka interaktif efektif dalam mendukung analisis opini publik di platform X.

Temuan penelitian ini juga menunjukkan bahwa analisis sentimen dapat dimanfaatkan oleh pemerintah, lembaga perlindungan perempuan dan anak, konselor keluarga, maupun organisasi sosial sebagai sarana pemantauan persepsi masyarakat terhadap isu perselingkuhan dan perceraian. Informasi yang diperoleh dari pemetaan opini publik tersebut dapat digunakan sebagai dasar dalam merancang program edukasi keluarga, kampanye pencegahan perselingkuhan, layanan konseling pranikah maupun pascanikah, serta strategi komunikasi publik yang lebih tepat sasaran untuk meminimalkan dampak sosial yang ditimbulkan.

Untuk penelitian selanjutnya, disarankan menerapkan teknik penanganan data tidak seimbang seperti SMOTE, Random Oversampling, atau metode cost-sensitive learning guna meningkatkan kemampuan model dalam mengenali kelas minoritas. Selain itu, penelitian mendatang dapat memanfaatkan model berbasis Deep Learning atau Transformer seperti IndoBERT untuk memperoleh representasi konteks bahasa yang lebih baik, serta mengintegrasikan analisis sentimen dengan analisis topik (topic modeling) sehingga tidak hanya mampu mengidentifikasi polaritas opini, tetapi juga mengungkap faktor-faktor utama yang melatarbelakangi persepsi masyarakat terhadap fenomena perceraian akibat perselingkuhan.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada seluruh pihak yang telah memberikan dukungan, bimbingan, dan bantuan selama proses penelitian serta penyusunan artikel ini. Apresiasi disampaikan kepada pembimbing, institusi terkait, keluarga, dan rekan-rekan atas kontribusi serta motivasi yang diberikan. Semoga hasil penelitian ini dapat bermanfaat dan memberikan kontribusi bagi pengembangan ilmu pengetahuan.

## References

- [1] P. Salsabila, “Al-Qur’an sebagai Resolusi Konflik: Studi Tafsir Kemenag,” vol. 4, pp. 127–142, 2024.
- [2] R. I. Prasadirini and M. Arifin, “Dampak Perselingkuhan Orang Tua terhadap Perilaku Sosial Anak (Studi Kasus Kecamatan Cepogo Kabupaten Boyolali),” *Al-Mabsut: Jurnal Studi Islam dan Sosial*, vol. 18, no. 2, pp. 365–380, 2024, doi: 10.56997/almabsut.v18i2.1463.
- [3] A. T. R. Farhana, “Eksresi Emosional Netizen dalam Komentar Berita Selingkuh di Komunitas Marah-Marah Twitter,” *Ilmu Budaya: Jurnal Bahasa, Sastra, Seni, dan Budaya*, vol. 9, no. 1, p. 123, 2025, doi: 10.30872/jbssb.v9i1.18883.
- [4] V. P. R. Ronaldo Gabriel Pemasi and P. T. D. Rompas, “Analisis Sentimen terhadap Review Aplikasi Hotel Online Menggunakan Metode K-Nearest Neighbor,” pp. 44–52.
- [5] B. Siswoyo, Z. A. Abas, A. N. C. Pee, R. Komalasari, and N. Suyatna, “Pemanfaatan Machine Learning untuk Klasifikasi Sentimen Pelanggan pada Media Sosial,” *Jurnal Teknologi Informasi Digital*, vol. 1, no. 1, pp. 29–34, 2025.
- [6] A. S. Suweleh, D. Susilowati, and K. Marzuki, “Penanganan Ketidakseimbangan Kelas Menggunakan Pendekatan Level Data,” vol. 20, no. 1, 2020, doi: 10.30812/matrik.v20i1.846.
- [7] M. A. Supit, A. A. Kenap, G. David, and P. Maramis, “Penerapan Algoritma Decision Tree C4.5 pada Layanan,” vol. 11, no. 1, pp. 89–99, 2026.
- [8] N. P. Sahay, K. Santa, and A. A. K., “Analisis Sentimen terhadap Aplikasi Tim Salut Menggunakan Metode Naïve Bayes Classifier Berbasis Web,” pp. 76–85, 2024.
- [9] I. P. W. K. Gumi and A. Syafrianto, “Perbandingan Algoritma Naïve Bayes dan Decision Tree pada Sentimen Analisis,” *IJCSR: The Indonesian Journal of Computer Science Research*, vol. 1, 2022.
- [10] A. Purnamasari and A. Agoestanto, “Modeling of Naïve Bayes and Decision Tree Algorithms to Analyze Sentiment Related to Jaklingko Public Transportation on Social Media X (Twitter),” vol. 5, no. 1, pp. 67–78, 2024, doi: 10.30598/ppcst.2024.knmxxii.67-78.
- [11] F. A. Gosal, P. T. D. Rompas, G. David, and P. Maramis, “Aplikasi Prediksi Nilai Tukar Rupiah terhadap Dolar Menggunakan Algoritma Autoregressive Integrated Moving Average,” vol. 9, no. 2, pp. 1–12, 2025.
- [12] S. A. R. Rizaldi, S. Alam, and I. Kurniawan, “Analisis Sentimen Pengguna Aplikasi JMO (Jamsostek Mobile) pada Google Play Store Menggunakan Metode Naive Bayes,” *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer*, vol. 2, no. 3, pp. 109–117, 2023, doi: 10.55123/storage.v2i3.2334.
- [13] M. Rizky Pratama, Y. R. Ramadhan, and M. A. Komara, “Analisis Sentimen BRImo dan BCA Mobile Menggunakan Support Vector Machine dan Lexicon Based.”
- [14] A. Safira, A. S. Masyarakat, and F. N. Hasan, “Analisis Sentimen Masyarakat terhadap Paylater Menggunakan Metode Naive Bayes Classifier,” *Jurnal Sistem Informasi*, vol. 5, no. 1, 2023.
- [15] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, “Analisis Klasifikasi Sentimen terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 7, no. 1, pp. 47–58, 2021, doi: 10.28932/jutisi.v7i1.3216.
- [16] T. Krisdiyanto, E. Maricha, and O. Nurharyanto, “Analisis Sentimen Opini Masyarakat Indonesia terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naive Bayes Classifiers,” vol. 7, no. 1, pp. 32–37, 2021.
- [17] A. P. Kusuma and A. S. Wibowo, “Analisis Sentimen pada Aplikasi X terhadap Kebijakan TAPERA Menggunakan Algoritma Naïve Bayes,” *Naratif: Jurnal Ilmiah Nasional Riset Aplikasi dan Teknik Informatika*, vol. 7, no. 1, pp. 56–63, 2025.

# Academia Open

Vol. 11 No. 1 (2026): June

DOI: 10.21070/acopen.11.2026.14689

- [18] Y. Ansori, K. Fahmi, and H. Holle, "Perbandingan Metode Machine Learning dalam Analisis Sentimen Twitter," *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, vol. 10, no. 4, pp. 1–6, 2022, doi: 10.26418/justin.v10i4.51784.
- [19] S. E. Audi Rahma, "Analisis Sentimen Pengguna Twitter terhadap Fluktuasi Harga Ethereum Menggunakan NLP dan Machine Learning," vol. 10, no. 1, pp. 1195–1203, 2026.
- [20] Y. Desnelita, N. Nasution, L. Suryati, and F. Zoromi, "Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak Seimbang," vol. 21, no. 3, 2022, doi: 10.30812/matrik.v21i3.1726.