Jurnal Pendidikan dan Teknologi Indonesia (JPTI)
Vol. 5, No. 9, September 2025, Hal. 2706-2718
DOI: https://doi.org/10.52436/1.jpti.1044
p-ISSN: 2775-4227
e-ISSN: 2775-4219

ANALISIS SENTIMEN PUBLIK TERHADAP REVISI UNDANG-UNDANG TENTARA NASIONAL INDONESIA DI TWITTER MENGGUNAKAN NAIVE BAYES DAN RANDOM FOREST

Alfred Hasiholan Simamora[*1], Ryan Randy Suryono[2]

[1]Fakultas Teknik dan Ilmu Komputer, Program Studi Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia
[2]Fakultas Teknik dan Ilmu Komputer, Program Studi Magister Ilmu Komputer, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia
Email: [1]alfred_hasiholan_simamora@teknokrat.ac.id, [2]ryan@teknokrat.ac.id

Abstrak

Media sosial X (sebelumnya Twitter) menjadi ruang diskusi intens terkait rencana Revisi Undang-Undang Tentara Nasional Indonesia (RUU TNI), yang memunculkan berbagai opini publik. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap RUU TNI dengan mengelompokkan opini pengguna ke dalam kategori positif dan negatif. Data sebanyak 5.646 unggahan dikumpulkan melalui teknik crawling berbasis kata kunci, lalu diproses melalui tahapan pembersihan, tokenisasi, stopword removal, dan stemming. Fitur diekstraksi menggunakan pendekatan lexicon-based Bahasa Indonesia dan TF-IDF, serta penanganan ketidakseimbangan data dilakukan menggunakan class weight. Dua algoritma machine learning, Naive Bayes dan Random Forest, digunakan untuk membandingkan performa klasifikasi. Hasil menunjukkan bahwa Random Forest memberikan akurasi tertinggi sebesar 84,51%, dengan f1-score 91% pada kelas negatif. Naive Bayes mencatat akurasi 83,63%, dengan f1-score 90% untuk kelas negatif. Kedua model mengalami kesulitan dalam mendeteksi sentimen positif, namun Random Forest terbukti lebih andal secara keseluruhan. Temuan menunjukkan dominasi sentimen negatif, mencerminkan kekhawatiran masyarakat terhadap implikasi revisi undang-undang. Penelitian ini memberikan kontribusi akademik dalam pemanfaatan analisis sentimen untuk memahami persepsi publik terhadap isu kebijakan nasional secara cepat dan terukur**.**

Kata kunci: *Analisis Sentimen, Media Sosial X, Naive Bayes, Random Forest, RUU TNI, Sentimen Publik.*

Analysis of Public Sentiment Towards the Revision of the Indonesian National Army Law on Twitter Using Naive Bayes and Random Forest

Abstract

Social media X (formerly Twitter) has become a space for intense discussion regarding the plan to revise the Indonesian National Army Law (RUU TNI), which has generated various public opinions. This study aims to analyze public sentiment towards the TNI Bill by grouping user opinions into positive and negative categories. Data of 5,646 posts were collected through keyword-based crawling techniques, then processed through the stages of cleaning, tokenization, stopword removal, and stemming. Features were extracted using Indonesian lexicon-based approach and TF-IDF, and data imbalance was handled using class weight. Two machine learning algorithms, Naive Bayes and Random Forest, were used to compare classification performance. The results showed that Random Forest gave the highest accuracy of 84.51%, with an f1-score of 91% in the negative class. Naive Bayes recorded an accuracy of 83.63%, with an f1-score of 90% for the negative class. Both models had difficulty detecting positive sentiment, but Random Forest proved more reliable overall. The findings showed a predominance of negative sentiment, reflecting the public's concern over the implications of the law revision. This research provides an academic contribution in the utilization of sentiment analysis to understand public perceptions of national policy issues in a quick and measurable manner.

Keywords: *Naive Bayes, Public Sentiment, Random Forest, Sentiment Analysis, Social Media X, TNI Bill.*

p-ISSN: 2775-4227
e-ISSN: 2775-4219

1. PENDAHULUAN

Kemajuan teknologi informasi dan komunikasi dalam dua puluh tahun terakhir telah menyebabkan transformasi yang berarti dalam cara interaksi sosial dan penyampaian pendapat masyarakat. Media sosial, sebagai salah satu produk teknologi yang paling menonjol, telah berfungsi sebagai media utama bagi publik untuk bertukar informasi, berdialog, dan menyampaikan opini mereka mengenai beragam topik. Salah satu platform yang sangat berpengaruh adalah X (yang sebelumnya disebut Twitter), yang memungkinkan penyebaran informasi secara cepat, luas, dan interaktif[1]. Fenomena ini terlihat pada lonjakan diskusi tentang rencana pengesahan Revisi Undang-Undang Tentara Nasional Indonesia (RUU TNI) pada Maret 2025. Lonjakan perbincangan publik tersebut mengindikasikan perhatian yang besar dari masyarakat terhadap kebijakan yang dianggap berdampak langsung pada stabilitas sosial, politik, dan keamanan nasional. Dalam konteks ini, diperlukan alat analisis yang mampu menangkap dan menginterpretasikan opini publik secara komprehensif, salah satunya adalah analisis sentimen[2]. Analisis sentimen adalah pendekatan yang menggunakan data dari teks untuk mengidentifikasi dan memahami kecenderungan emosi atau opini masyarakat terhadap suatu isu tertentu. Dengan memanfaatkan data besar yang tersedia di media sosial, analisis ini dapat memberikan wawasan yang lebih dalam tentang bagaimana masyarakat merespons suatu kebijakan atau peristiwa[3]. Pendekatan ini relevan untuk kasus Undang-Undang Tentara Nasional Indonesia (RUU TNI), di mana opini masyarakat tersebar luas dengan nuansa yang sangat beragam, mulai dari dukungan hingga kritik tajam. Dalam penelitian ini, analisis sentimen difokuskan untuk memetakan persepsi publik terhadap Undang-Undang Tentara Nasional Indonesia (RUU TNI) ke dalam dua kategori utama, yaitu sentimen positif dan negatif. Pemahaman terhadap pola sentimen ini tidak hanya bermanfaat bagi pemerintah untuk mengevaluasi kebijakan, tetapi juga bagi pemangku kepentingan lainnya yang ingin memahami dinamika sosial di era digital[4].

Studi ini memanfaatkan dua algoritma machine learning, yakni Naive Bayes dan Random Forest, untuk menganalisis opini masyarakat. Pemilihan kedua algoritma ini didasarkan pada keunggulannya dalam menangani data teks, seperti kesederhanaan dan efisiensi Naive Bayes dalam memodelkan distribusi probabilitas kata, serta kekuatan Random Forest dalam mengolah dataset kompleks dengan fitur-fitur yang beragam[5]. Kedua model tersebut diharapkan dapat memberikan perspektif yang berbeda terkait efektivitas pengelompokan sentimen, sehingga hasil penelitian mampu mencerminkan opini publik dengan tingkat akurasi yang tinggi. Namun, belum ada penelitian yang secara khusus menganalisis sentimen publik terhadap Undang-Undang Tentara Nasional Indonesia (UU TNI) berdasarkan data media sosial dengan membandingkan kinerja dua algoritma machine learning ini. Penelitian ini mengisi celah ini. Lebih jauh, penelitian ini bertujuan untuk mengidentifikasi kelemahan dan keunggulan masing-masing model dalam konteks analisis teks berbahasa Indonesia. Untuk mendukung proses analisis, data dikumpulkan dari media sosial X menggunakan metode crawling dengan kata kunci yang relevan, seperti "RUU TNI," "#RUUTNI," dan "#TolakRUUTNI."[6]. Periode pengumpulan data mencakup waktu sebelum dan sesudah puncak diskusi publik, yaitu dari 15 Maret hingga 20 April 2025.

Rentang waktu ini dipilih untuk memastikan bahwa data mencakup dinamika opini masyarakat secara holistik, baik pada fase awal diskusi maupun setelah kebijakan menjadi topik utama. Tahapan pengolahan data dilakukan secara menyeluruh, mencakup pembersihan elemen tidak relevan seperti URL dan tanda baca, normalisasi teks melalui case folding, tokenisasi untuk memecah teks menjadi unit analisis, serta penghapusan kata-kata umum (stopword removal) yang tidak memiliki signifikansi analisis[7]. Proses ini dilanjutkan dengan stemming menggunakan algoritma yang relevan untuk Bahasa Indonesia, sehingga menghasilkan data yang bersih, terstruktur, dan siap digunakan untuk pemodelan klasifikasi. Oleh karena itu, tujuan utama penelitian ini adalah untuk memberikan analisis sentimen yang menyeluruh terhadap opini publik tentang RUU TNI dengan menggunakan Naive Bayes dan Random Forest. Selain itu, juga akan menilai manfaat dan kekurangan masing-masing pendekatan dalam konteks data berbahasa Indonesia. Hal yang diharapkan dari hasil penelitian adalah bahwa mereka akan memberikan bantuan praktis dalam proses membangun teknik untuk menganalisis sentimen. Mereka juga akan berfungsi sebagai referensi bagi pengambil kebijakan untuk mendapatkan pemahaman yang lebih baik tentang pendapat masyarakat dengan lebih cepat dan mendalam.

Melalui serangkaian tahapan ini, penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam memahami dinamika opini publik terhadap isu-isu kebijakan nasional. Lebih jauh, hasil penelitian ini tidak hanya relevan dalam konteks revisi Undang-Undang Tentara Nasional Indonesia (RUU TNI), tetapi juga dapat dijadikan sebagai model analisis untuk isu-isu kebijakan lain yang menjadi perhatian masyarakat luas.

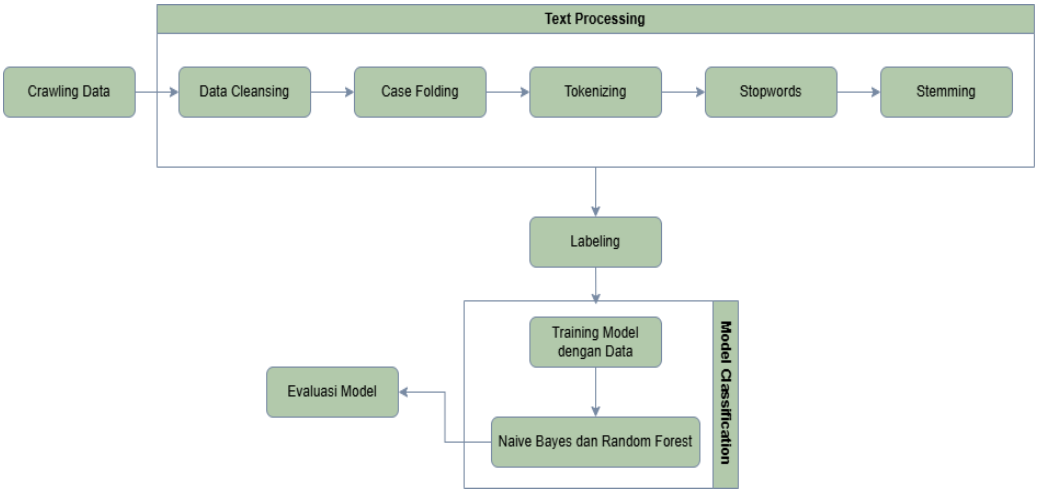
2. METODE PENELITIAN

Studi ini menerapkan pendekatan kuantitatif komparatif yang bertujuan untuk mengelompokkan pandangan masyarakat terkait Revisi Undang-Undang Tentara Nasional Indonesia (RUU TNI) yang muncul di media sosial X (sebelumnya Twitter) ke dalam dua kategori utama, yaitu sentimen positif dan negatif. Penelitian ini menggunakan data unggahan dari pengguna media sosial dan mengaplikasikan metode pembelajaran mesin untuk

p-ISSN: 2775-4227
e-ISSN: 2775-4219

membandingkan kinerja dua algoritma klasifikasi, yaitu Naive Bayes dan Random Forest [8]. Desain penelitian ini bersifat eksploratif dan aplikatif, karena tidak hanya menganalisis kecenderungan sentimen publik, tetapi juga mengevaluasi efektivitas dari masing-masing algoritma yang digunakan dalam mengolah data teks secara sistematis. Fokus penelitian ini terletak pada pengolahan data berbasis teks secara statistik dan berbasis fitur numerik, dengan metode non-linguistik yang tidak membutuhkan pemrosesan bahasa alami tingkat tinggi (seperti parsing sintaksis atau analisis dependensi) [9].

Python digunakan untuk menjalankan seluruh tahapan penelitian ini. Ini didukung oleh beberapa library seperti pandas dan numpy untuk manipulasi data, scikit-learn untuk pengajaran dan evaluasi model machine learning, dan Sastrawi untuk proses stemming pada teks Bahasa Indonesia.



Gambar 1. Alur Penyelesaian Masalah

2.1. Pengumpulan Data

Data dalam penelitian ini diambil dari media sosial X (yang sebelumnya disebut Twitter), yang dipilih sebagai sumber informasi karena merupakan salah satu platform utama untuk menyampaikan opini publik secara langsung dan real-time. Untuk mendapatkan data yang relevan, diterapkan teknik Tweet-Harvest, yaitu cara pengumpulan data secara otomatis dari media sosial dengan memanfaatkan akses terhadap Application Programming Interface (API) yang disediakan oleh platform itu [10]. Metode ini memungkinkan pengumpulan data dalam skala besar dengan cara yang efisien berdasarkan kata kunci tertentu yang berhubungan dengan topik penelitian. Pengumpulan data dilaksanakan antara 15 Maret hingga 20 April 2025, yang mencakup waktu sebelum dan sesudah puncak perdebatan publik terkait rencana pengesahan Revisi Undang-Undang Tentara Nasional Indonesia (RUU TNI), yang terjadi sekitar tanggal 20 Maret 2025 [11]. Rentang waktu ini dipilih untuk mendapatkan gambaran yang lebih komprehensif mengenai dinamika opini publik, baik pada saat isu mulai mencuat maupun setelahnya. Kata kunci yang digunakan dalam proses crawling disesuaikan dengan istilah-istilah yang sering muncul dalam percakapan publik seputar Undang-Undang Tentara Nasional Indonesia (RUU TNI), seperti "RUU TNI", "Revisi TNI", dan berbagai variasi hashtag yang relevan. Dari proses crawling tersebut, terkumpul sebanyak 5.645 data unggahan yang memuat opini atau tanggapan publik terhadap isu tersebut. Data yang dikumpulkan berupa teks unggahan (tweet) yang berbahasa Indonesia, dan diseleksi untuk memastikan hanya data yang mengandung opini aktual yang disertakan dalam tahap analisis.

2.2. Preprocessing data

Pra-pemrosesan data (Preprocessing) adalah serangkaian langkah yang dilakukan untuk menyiapkan data mentah agar dapat digunakan secara efektif dalam analisis atau pemodelan machine learning. Pada konteks analisis teks, pra-pemrosesan bertujuan untuk menyederhanakan dan menyusun ulang teks mentah menjadi bentuk yang lebih terstruktur, bersih, dan relevan bagi algoritma[12]. Proses ini penting karena teks mentah sering kali mengandung elemen-elemen yang tidak relevan, redundan, atau sulit diproses langsung oleh komputer, seperti tanda baca, kata-kata umum (stopwords), dan berbagai bentuk morfologi kata.

a) Pembersihan Data
   Menghapus unsur-unsur yang tidak penting seperti simbol, angka, URL, emoji, dan karakter non-alfabet, serta menormalisasi teks dengan mengurangi jumlah huruf besar.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

b) Tokenisasi
Memecah teks menjadi bagian-bagian terkecil, seperti kata atau simbol, untuk lebih mudah menganalisisnya.
c) Penghapusan Stopword
Menghapus kata-kata umum seperti "dan", "yang", atau "di" yang tidak relevan untuk analisis.
d) Stemming
Untuk menyatukan variasi kata yang memiliki makna yang sama, gunakan alat seperti stemmer Sastrawi untuk mengubah kata ke bentuk dasarnya.

Tahapan *preprocessing* dilakukan menggunakan bahasa pemrograman *Python* dengan bantuan *library* seperti *re* untuk pembersihan teks, nltk untuk tokenisasi dan *stopword removal*, serta *Sastrawi* untuk *stemming*.

2.3. Handle Imbalance

Pada tahap ini, penanganan ketidakseimbangan kelas (class imbalance) menjadi perhatian utama, mengingat distribusi data yang tidak seimbang antara sentimen negatif dan positif. Untuk mengatasi masalah ini, digunakan teknik class weights pada model klasifikasi. Dengan menggunakan class weights, bobot lebih besar diberikan pada kelas minoritas (sentimen positif), sehingga model dapat lebih sensitif terhadap kelas tersebut dan menghindari bias terhadap kelas mayoritas (sentimen negatif)[13]. Teknik ini membantu model untuk memberikan perhatian yang lebih seimbang pada kedua kelas, sehingga dapat menghasilkan prediksi yang lebih adil dan akurat. Dua pendekatan utama digunakan untuk mengekstraksi fitur dalam penelitian ini: Frekuensi Frequency-Inverse Document (TF-IDF) dan pendekatan berbasis kamus sentimen (lexicon-based). [14]. Meskipun kedua teknik ini berkonsentrasi pada transformasi teks menjadi representasi numerik atau berbasis kata-kata tertentu, penekanan di bagian ini adalah pada TF-IDF, yang memberikan perhatian lebih besar pada kata-kata yang sering muncul di satu dokumen tetapi jarang ditemukan di dokumen lain. Pendekatan TF-IDF memungkinkan model menangkap informasi yang lebih relevan dan mengurangi efek kata-kata yang sering muncul di seluruh korpus[15]. Gabungan dari teknik ini diharapkan dapat meningkatkan akurasi model dalam membedakan sentimen positif dan negatif, serta membantu model bekerja lebih efektif meskipun menghadapi ketidakseimbangan kelas.

2.4. Labeling

Pada tahap ini, proses pemberian label dilakukan untuk menandai sentimen yang terkandung dalam setiap unggahan. Labeling ini penting untuk membangun dataset yang akan digunakan sebagai target variabel dalam model klasifikasi. Setiap unggahan diberi label Negatif atau Positif berdasarkan analisis sentimen yang dilakukan terhadap teks unggahan tersebut. Labeling dilakukan dengan menggunakan pendekatan lexicon-based yang memanfaatkan kamus sentimen berbahasa Indonesia. Kamus ini berisi daftar kata-kata yang dikategorikan sebagai positif atau negatif, dan digunakan untuk menentukan kecenderungan sentimen dalam unggahan[16]. Jika unggahan mengandung lebih banyak kata bernuansa negatif, maka unggahan tersebut diberi label Negatif. Sebaliknya, jika unggahan mengandung lebih banyak kata bernuansa positif, maka diberi label Positif. Proses ini dilakukan dengan tujuan untuk menyediakan data yang terlabel dengan baik sebagai input bagi model klasifikasi. Keakuratan dalam pemberian label sangat penting, karena kesalahan dalam proses labeling dapat memengaruhi kinerja model dalam memprediksi sentimen pada data baru. Oleh karena itu, meskipun pendekatan ini bersifat otomatis, tetap diperlukan pengecekan dan validasi untuk memastikan bahwa label yang diberikan sesuai dengan konteks sentimen yang terkandung dalam unggahan.

2.5. Klasifikasi Sentimen

Dalam penelitian ini, langkah klasifikasi sentimen bertujuan untuk mengelompokkan pendapat publik dari X di media sosial ke dalam dua kategori: positif dan negatif. Dua algoritma pengajaran mesin Naive Bayes dan Random Forest digunakan untuk melakukan klasifikasi data setelah melewati tahapan pra-pemrosesan dan ekstraksi fitur. Algoritma-algoritma ini dipilih karena kemampuan mereka untuk menangani teks Bahasa Indonesia dan mengelola sejumlah besar data yang beragam[17]. Sementara Random Forest dipilih karena kemampuan untuk membuat model prediksi yang stabil dengan menggabungkan banyak pohon keputusan, Naive Bayes digunakan karena kesederhanaannya dalam menghitung peluang klasifikasi berdasarkan distribusi kata. Masing-masing model diuji untuk kinerja dalam mengklasifikasikan sentimen. Ini dilakukan dengan menggunakan metrik evaluasi seperti akurasi, precision, recall, dan skor f1. Hasil klasifikasi digunakan untuk menentukan kecenderungan opini publik terhadap RUU TNI dan mengevaluasi kekuatan masing-masing model dalam menangani sifat bahasa media sosial yang dinamis.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

2.6. Evaluasi Model

Sejauh mana algoritma yang digunakan dapat mengidentifikasi sentimen dengan tepat, evaluasi model klasifikasi dilakukan. Dalam penelitian ini, metrik evaluasi umum untuk klasifikasi teks seperti akurasi, ketepatan, recall, dan skor f1 digunakan untuk mengevaluasi kinerja model. Metrik-metrik ini menunjukkan keseimbangan antara kemampuan model untuk menemukan sentimen yang tepat dan menghindari kesalahan klasifikasi, terutama dalam kondisi data di mana mungkin ada ketidakseimbangan antara sentimen positif dan negatif [18]. Proses evaluasi dilakukan dengan membandingkan performa dua algoritma, yaitu Naive Bayes dan Random Forest, terhadap data hasil ekstraksi fitur. Penekanan khusus diberikan pada f1-score karena metrik ini memperhitungkan keseimbangan antara precision dan recall, sehingga cocok digunakan dalam konteks klasifikasi dua kelas. Dengan evaluasi ini, dapat diketahui algoritma mana yang lebih efektif digunakan dalam analisis sentimen publik terhadap Undang-Undang Tentara Nasional Indonesia (RUU TNI), serta memberikan dasar dalam memilih model terbaik untuk diterapkan pada data serupa di masa depan[19].

3. HASIL DAN PEMBAHASAN

3.1. HASIL

Bagian ini berbicara tentang hasil dari tindakan yang telah dilakukan, termasuk pembersihan data, visualisasi kata kunci, dan klasifikasi perasaan. Data yang dianalisis berasal dari media sosial yang berhubungan dengan Rancangan Undang-Undang Tentara Nasional Indonesia dan Kepolisian Negara Republik Indonesia. Proses teknis seperti penghapusan kata umum, pemisahan kata, dan visualisasi menggunakan wordcloud dan diagram frekuensi digunakan untuk mendapatkan hasil. Dengan menggunakan algoritma Support Vector Machine, sentimen kemudian diklasifikasikan menjadi positif, negatif, dan netral. Analisis ini memberikan gambaran umum tentang opini publik tentang Rancangan Undang-Undang yang dibahas.

3.1.1. Preprocessing Data

Tahap preprocessing data merupakan langkah awal yang sangat penting dalam pengolahan teks. Langkah ini bertujuan untuk membersihkan data dari elemen-elemen yang tidak relevan, menyusun data agar lebih terstruktur, dan mempersiapkannya untuk analisis lebih lanjut. Preprocessing membantu meningkatkan kualitas data, yang secara langsung memengaruhi hasil akhir dari analisis atau model yang diterapkan. Langkah-langkah ini dilakukan secara sistematis untuk mengurangi noise, meningkatkan efisiensi, dan memastikan konsistensi dalam pemrosesan teks.Proses preprocessing terdiri dari beberapa tahapan utama yang melibatkan pembersihan, transformasi, dan normalisasi data teks. Setiap tahapan memiliki fungsi spesifik untuk memastikan data lebih representatif terhadap informasi penting yang dibutuhkan dalam analisis. Berikut adalah penjelasan dari setiap tahap preprocessing:

a) Cleansing

Data dipersiapkan untuk analisis lebih lanjut dengan melakukan tahapan preprocessing untuk menghilangkan elemen yang tidak relevan. Menghapus tag, URL, emotikon, tanda baca, dan karakter non-alfabet adalah beberapa langkah yang dilakukan.

Tabel 1. Cleansing

Tweet	Cleaning
Rasanya baru Isu RUU TNI ini aksi penolakan terjadi sampai di kota-kota kecil di daerah2 Hari ini ada di Kediri Bojonegoro dan Bogor. #CabutUUTNI #TolakRUUPolri https://t.co/UPHuHBRQeZ	Rasanya baru Isu RUU TNI ini aksi penolakan terjadi sampai di kota kota kecil di daerah2 Hari ini ada di Kediri Bojonegoro dan Bogor. cabutuutni tolakruupolri
AYOO BACA YANG BELUM TAU KENAPA KITA HARUS MENOLAK RUU TNI & RUU POLRI!! LAWAN SAMPAI MENANG #CabutUUTNI #TolakRUUPolri #IndonesiaGelap https://t.co/nnqzX9yrvm	AYOO BACA YANG BELUM TAU KENAPA KITA HARUS MENOLAK RUU TNI RUU POLRI!! LAWAN SAMPAI MENANGcabutuutni tolakruupolri indonesiagelap

b) Case Folding

Salah satu metode normalisasi teks adalah case folding, di mana setiap huruf dalam teks diubah menjadi format huruf kecil atau lowercase. Langkah ini sering digunakan dalam pengolahan teks, terutama dalam analisis teks atau pemrosesan bahasa alami (NLP), untuk memastikan perbandingan kata atau pencocokan pola konsisten.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Tabel 2. Case Folding

Tweet	Case Folding
Rasanya baru Isu RUU TNI ini aksi penolakan terjadi sampai di kota kota kecil di daerah2 Hari ini ada di Kediri Bojonegoro dan Bogor. cabutuutni tolakruupolri	rasanya baru isu ruu tni ini aksi penolakan terjadi sampai di kota kota kecil di daerah hari ini ada di kediri bojonegoro dan bogor cabutuutni tolakruupolri
AYOO BACA YANG BELUM TAU KENAPA KITA HARUS MENOLAK RUU TNI RUU POLRI!! LAWAN SAMPAI MENANGcabutuutni tolakruupolri indonesiagelap	ayoo baca yang belum tahu kenapa kita harus tolak ruu tni amp ruu polri lawan sampai menang cabutuutni tolakruupolri indonesiagelap

c) Tokenizing

Proses membagi teks menjadi bagian yang disebut "token", yang dapat berupa kata, frasa, kalimat, atau bahkan karakter, tergantung pada aplikasi. Tokenizing adalah langkah awal yang sangat penting dalam pengolahan bahasa alami (NLP) karena teks mentah harus diubah menjadi bentuk yang lebih terstruktur sehingga algoritma atau model dapat menganalisisnya lebih lanjut.

Tabel 3. Tokenizing

Tweet	Tokenizing
rasanya baru isu ruu tni ini aksi penolakan terjadi sampai di kota kota kecil di daerah hari ini ada di kediri bojonegoro dan bogor cabutuutni tolakruupolri	['rasanya', 'baru', 'isu', 'ruu', 'tni', 'ini', 'aksi', 'penolakan', 'terjadi', 'sampai', 'di', 'kota', 'kota', 'kecil', 'di', 'daerah', 'hari', 'ini', 'ada', 'di', 'kediri', 'bojonegoro', 'dan', 'bogor', 'cabutuutni', 'tolakruupolri']
ayoo baca yang belum tahu kenapa kita harus tolak ruu tni amp ruu polri lawan sampai menang cabutuutni tolakruupolri indonesiagelap	['ayoo', 'baca', 'yang', 'belum', 'tahu', 'kenapa', 'kita', 'harus', 'tolak', 'ruu', 'tni', 'amp', 'ruu', 'polri', 'lawan', 'sampai', 'menang', 'cabutuutni', 'tolakruupolri', 'indonesiagelap']

d) Stopword

Dalam pengolahan bahasa alami (NLP), stopword adalah istilah yang mengacu pada kata-kata umum yang sering muncul dalam teks, tetapi biasanya tidak memiliki nilai informasi yang signifikan untuk analisis. Dalam bahasa Indonesia, stopwords seperti "dan", "yang", "di", "ke", "ini", "itu", dan sebagainya. Sebaliknya, dalam bahasa Inggris, stopwords seperti "the", "is", "at", "on", "in", dan sebagainya.

Tabel 4. Stopwords

Tweet	Stopword
['rasanya', 'baru', 'isu', 'ruu', 'tni', 'ini', 'aksi', 'penolakan', 'terjadi', 'sampai', 'di', 'kota', 'kota', 'kecil', 'di', 'daerah', 'hari', 'ini', 'ada', 'di', 'kediri', 'bojonegoro', 'dan', 'bogor', 'cabutuutni', 'tolakruupolri']	['isu', 'ruu', 'tni', 'aksi', 'penolakan', 'kota', 'kota', 'daerah', 'kediri', 'bojonegoro', 'bogor', 'cabutuutni', 'tolakruupolri']
['ayoo', 'baca', 'yang', 'belum', 'tahu', 'kenapa', 'kita', 'harus', 'tolak', 'ruu', 'tni', 'amp', 'ruu', 'polri', 'lawan', 'sampai', 'menang', 'cabutuutni', 'tolakruupolri', 'indonesiagelap']	['ayoo', 'baca', 'tolak', 'ruu', 'tni', 'amp', 'ruu', 'polri', 'lawan', 'menang', 'cabutuutni', 'tolakruupolri', 'indonesiagelap']

e) Stemming

Proses stemming adalah proses mengubah kata-kata dalam teks ke bentuk dasarnya (root word atau stem) dengan menghilangkan imbuhan seperti awalan, akhiran, atau sisipan. Tujuan utama stemming adalah untuk membuat variasi kata menjadi bentuk yang lebih konsisten, sehingga lebih mudah untuk menganalisis data teks untuk tujuan pencarian informasi, klasifikasi teks, atau analisis sentimen.

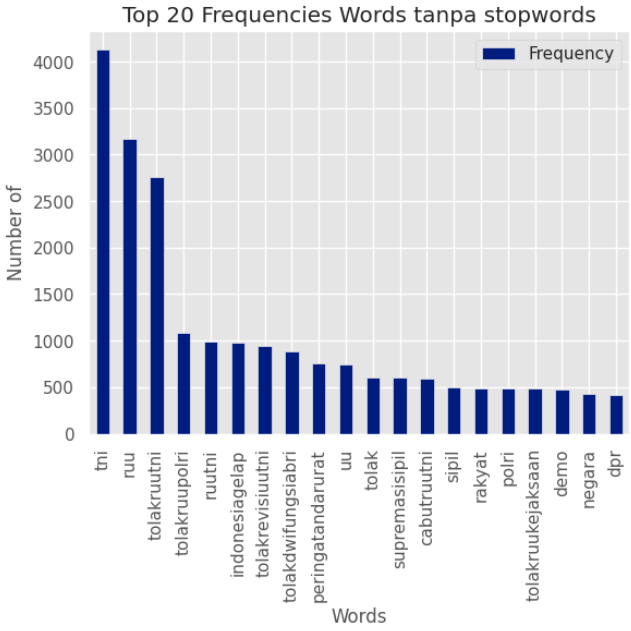
p-ISSN: 2775-4227
e-ISSN: 2775-4219

Tabel 5. Stemming

Tweet	Stemming
['isu', 'ruu', 'tni', 'aksi', 'penolakan', 'kota', 'kota', 'daerah', 'kediri', 'bojonegoro', 'bogor', 'cabutuutni', 'tolakruupolri']	isu ruu tni aksi tolak kota kota daerah diri bojonegoro bogor cabutuutni tolakruupolri
['ayoo', 'baca', 'tolak', 'ruu', 'tni', 'amp', 'ruu', 'polri', 'lawan', 'menang', 'cabutuutni', 'tolakruupolri', 'indonesiagelap']	ayoo baca tolak ruu tni amp ruu polri lawan menang cabutuutni tolakruupolri indonesiagelap

3.1.2. Tahapan Visualisasi Word Frequence

Setelah tahap preprocessing data selesai, langkah selanjutnya dalam analisis teks ini adalah visualisasi frekuensi kata untuk mendapatkan gambaran umum tentang topik dan kata kunci yang paling sering muncul dalam data. Analisis frekuensi kata penting karena dapat membantu peneliti memahami fokus utama dalam opini publik yang dianalisis [20]. Proses ini memberikan wawasan tentang topik atau tema yang dominan dalam data yang dianalisis. Berikut adalah hasil analisis frekuensi kata tanpa stopwords:



Gambar 2. Kata Yang Sering Muncul

Setelah menghapus stopwords, hasil analisis lebih fokus pada kata-kata bermakna yang relevan dengan konteks data. Kata "tni" (4.124 kali), "ruu" (3.166 kali), dan "tolakruutni" (2.761 kali) tetap menjadi yang paling dominan, menegaskan pentingnya isu terkait kebijakan TNI. Selain itu, kata-kata lain seperti "tolakrevisiuutni" (946 kali), "tolakdwifungsiabri" (887 kali), dan "peringatandarurat" (750 kali) menunjukkan adanya narasi yang kuat terkait dengan isu kebijakan dan fungsi TNI dalam struktur pemerintahan. Kata "tolak" (606 kali) juga sering muncul, mencerminkan nada protes atau penolakan dalam data ini.

3.1.3. Tahapan Visualisasi WordCloud

Wordcloud ini menggambarkan perbincangan publik yang intens terkait isu Undang-Undang Tentara Nasional Indonesia (RUU TNI) dan peran militer di Indonesia. Kata "tni", "ruu", dan "tolakruutni" menjadi yang paling dominan, mencerminkan kekhawatiran masyarakat terhadap arah kebijakan pertahanan dan keamanan[21]. Dengan mempertimbangkan stopwords, kata-kata seperti "yang", "dan", "tidak", dan "ini" muncul cukup sering, menandakan struktur kalimat yang mengekspresikan opini dan penolakan. Tanpa stopwords, terlihat lebih jelas istilah yang merepresentasikan sentimen negatif seperti "tolakruupolri", "indonesiagelap", "tolakrevisiuutni", hingga "tolakdwifungsiabri". Di sisi lain, muncul juga kata-kata yang bernuansa aspirasi positif seperti "supremasisipil", "sipil", "rakyat", dan "negara", yang menunjukkan dorongan untuk memperkuat peran sipil

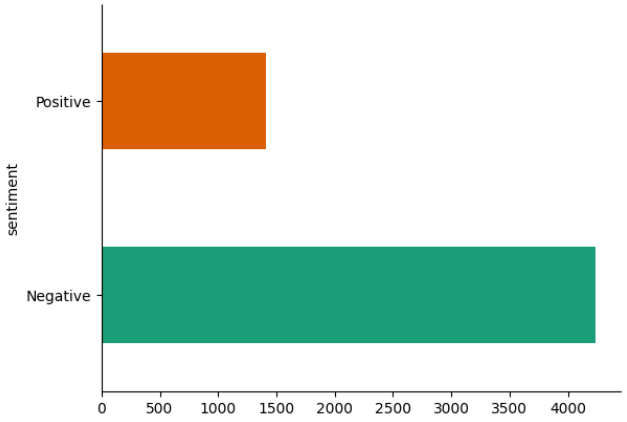
p-ISSN: 2775-4227
e-ISSN: 2775-4219

dalam demokrasi dan pemerintahan. Kata "demo", "dpr", dan "cabutruutni" menunjukkan aksi nyata dan jalur politik yang ditempuh oleh masyarakat. Wordcloud ini merupakan potret gabungan dari kekhawatiran dan harapan, penolakan terhadap militerisme, serta tuntutan akan supremasi sipil dan demokrasi yang lebih kuat.

Gambar 3. WordCloud Semua Sentiment

3.1.4. Tahap Pengujian

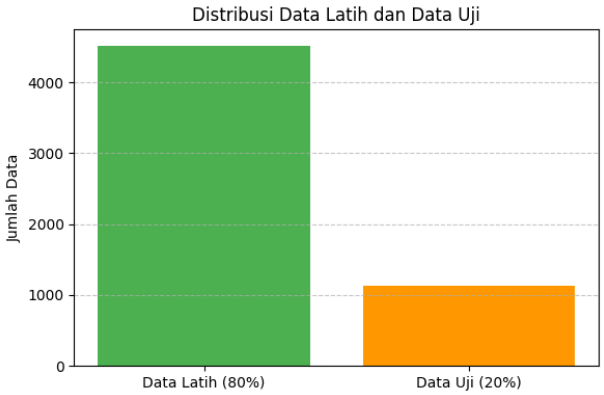
Pada titik ini, data dibagi menjadi dua komponen utama: set pelatihan (80%) dan set uji (20%). Terdapat 5.646 entri dalam dataset yang digunakan, dengan 4.238 data berlabel negatif dan 1.408 data berlabel positif. Agar model dapat dilatih dan diuji secara adil, pemisahan dilakukan secara acak dengan distribusi proporsi kelas tetap.



Gambar 4. Sentiment

Hasil pemisahan memberikan 4.516 data untuk melatih model dan 1.130 untuk pengujian. Data latih ini digunakan untuk melatih model untuk mengidentifikasi pola yang membedakan sentimen positif dan negatif dalam teks yang telah diekstraksi fiturnya. Sementara itu, data uji digunakan untuk mengukur kinerja model pada data yang belum pernah dipelajari sebelumnya.

p-ISSN: 2775-4227
e-ISSN: 2775-4219



Gambar 5. Distribusi Data Latih dan Data Uji

Namun, dalam dataset ini terdapat ketidakseimbangan kelas yang signifikan, di mana kelas Negative mendominasi dibandingkan kelas Positive. Untuk mengatasi masalah ini, digunakan class weights pada kedua model, yaitu Naive Bayes dan Random Forest. Dengan menyesuaikan bobot kelas, model lebih sensitif terhadap kelas minoritas (Positive), sehingga dapat meningkatkan kemampuan prediksi pada kelas tersebut tanpa perlu melakukan oversampling data. Setelah pelatihan, kedua model diuji pada data uji untuk mengevaluasi akurasi dan kinerjanya. Ini dilakukan dengan menggunakan metrik seperti akurasi, ketepatan, recall, skor f1, dan matrix confusion. Hasilnya menunjukkan seberapa baik model dapat mengklasifikasikan sentimen dengan benar, terutama untuk mendeteksi kelas minoritas. Metode ini diharapkan dapat menjalankan model dengan baik meskipun ada ketidakseimbangan kelas, memberikan hasil yang lebih adil antara kelas negatif dan positif.

3.1.5. Klasifikasi Model

Pada tahap klasifikasi, dua model yang diuji dalam penelitian ini adalah Naive Bayes dan Random Forest. Kedua model ini digunakan untuk memprediksi sentimen pada dataset yang telah diproses dan diekstraksi fiturnya menggunakan TF-IDF. Hasil evaluasi menunjukkan bahwa kedua model memiliki performa yang baik namun dengan karakteristik yang berbeda dalam hal akurasi, precision, recall, dan f1-score untuk kelas Negatif dan Positif. Naive Bayes, model probabilistik berbasis distribusi multinomial, menghasilkan akurasi sebesar 83.45%. Model ini menunjukkan bahwa untuk kelas negatif, akurasi 83%, recall 98%, dan skor f1 90%; untuk kelas positif, akurasi 85%, tetapi recall hanya 35%, yang menunjukkan bahwa model ini gagal menemukan tweet dengan sentimen positif sebagian besar. Sebaliknya, Random Forest—model ensemble dengan banyak pohon keputusan—menunjukkan akurasi 84.96%. Precision untuk kelas negatif pada model ini lebih tinggi, dengan 85% precision dan 97% recall, menghasilkan f1-score 91%, yang menunjukkan kemampuan model ini untuk mengidentifikasi tweet negatif dengan sangat baik. Namun, dengan precision 82% dan recall 46%, model ini juga kesulitan menemukan tweet dengan sentimen positif, yang menghasilkan f1-score 59% untuk kelas positif.

TABEL 6. Hasil Klasifikasi Model

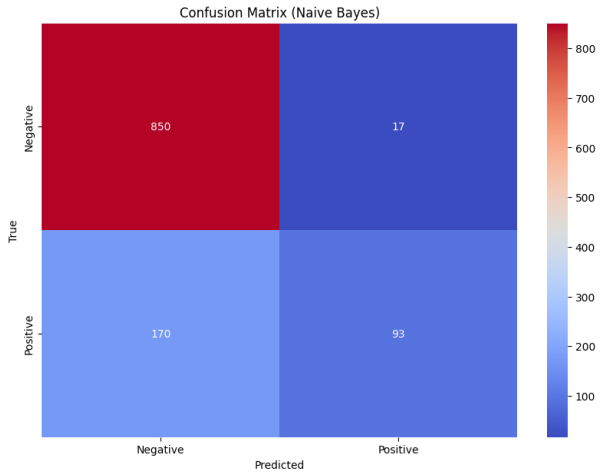
Model	Akurasi	Precision (Negatif)	Precision (Positif)	Recall (Negatif)	Recall (Positif)	F1-Score (Negatif)	F1-Score (Positif)
Naïve Bayes	83.45%	83%	85%	98%	35%	90%	50%
Random Forest	84.96%	85%	82%	97%	46%	91%	59%

Hasil ini menunjukkan bahwa meskipun Random Forest sedikit lebih unggul dalam akurasi keseluruhan, kedua model masih menunjukkan ketidakseimbangan dalam hal klasifikasi kelas Positif. Model Naive Bayes lebih sensitif dalam mengidentifikasi tweet positif dengan precision yang lebih tinggi, namun gagal mendeteksi sebagian besar tweet positif yang sebenarnya ada. Di sisi lain, Random Forest lebih efektif dalam mengidentifikasi tweet negatif, namun mengalami kesulitan dalam hal recall untuk tweet positif. Untuk meningkatkan kinerja model pada kelas positif, beberapa pendekatan seperti peningkatan sampling atau penyesuaian bobot kelas dapat dipertimbangkan, selain mencoba teknik lain seperti SMOTE untuk mengatasi masalah ketidakseimbangan kelas. Secara keseluruhan, meskipun kedua model sudah menunjukkan akurasi yang baik, pemahaman lebih mendalam tentang distribusi kelas dalam dataset dan penyesuaian model lebih lanjut masih diperlukan untuk mencapai hasil yang lebih seimbang antara kedua kelas.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

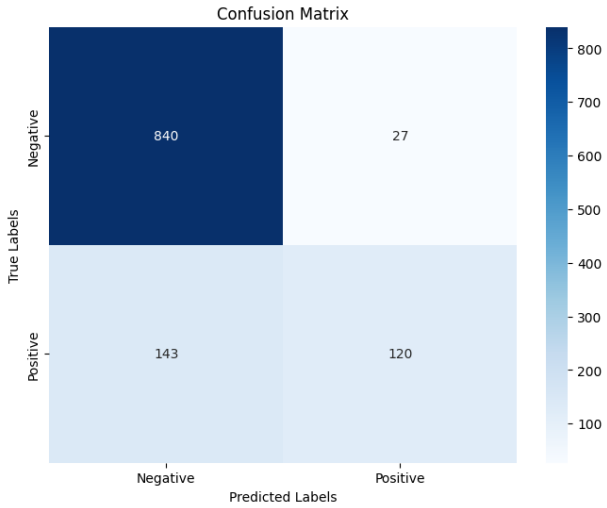
3.1.6. Confusion Matrix

Untuk memberikan gambaran yang lebih jelas tentang bagaimana model berfungsi dalam mengklasifikasikan data ke dalam dua kelas, yaitu Negatif dan Positif, matriks confusion menunjukkan jumlah prediksi yang benar dan salah untuk masing-masing kelas.



Gambar 6. Confusion Matrix Model Naive Bayes

Pada model Naive Bayes, dari total 867 data berlabel Negatif, sebanyak 850 berhasil diklasifikasikan dengan benar, sementara 17 salah diklasifikasikan sebagai Positif. Namun, dari 263 data berlabel Positif, hanya 93 yang berhasil diklasifikasikan dengan benar, sedangkan 170 lainnya salah diklasifikasikan sebagai Negatif. Hal ini menunjukkan bahwa model cenderung bias terhadap kelas Negatif, dan memiliki recall yang rendah untuk kelas Positif (hanya 35%).



Gambar 7. Confusion Matrix Model Random Forest

Sedangkan pada model Random Forest, dari 867 data Negatif, sebanyak 840 diklasifikasikan dengan benar, dan 27 salah diklasifikasikan sebagai Positif. Dari 263 data Positif, model berhasil mengklasifikasikan 120 data dengan benar dan salah mengklasifikasikan 143 data sebagai Negatif. Dibandingkan dengan Naive Bayes, model Random Forest memiliki recall yang lebih tinggi pada kelas Positif (46%), meskipun masih belum seimbang dengan kelas Negatif. Berdasarkan confusion matrix ini, dapat disimpulkan bahwa kedua model lebih andal dalam mendeteksi sentimen Negatif. Namun, Random Forest memiliki keunggulan dalam mendeteksi sentimen Positif dibandingkan Naive Bayes, walaupun keduanya tetap menghadapi tantangan dalam mengklasifikasikan data Positif secara akurat. Hal ini kembali menunjukkan dampak dari ketidakseimbangan kelas, meskipun telah ditangani menggunakan class weight.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

3.2. PEMBAHASAN

Bagian ini berbicara tentang hasil dari tindakan yang telah dilakukan, termasuk pembersihan data, visualisasi kata kunci, dan klasifikasi perasaan. Data yang dianalisis berasal dari media sosial yang berhubungan dengan Rancangan Undang-Undang Tentara Nasional Indonesia dan Kepolisian Negara Republik Indonesia. Proses teknis seperti penghapusan kata umum, pemisahan kata, dan visualisasi menggunakan wordcloud dan diagram frekuensi digunakan untuk mendapatkan hasil. Dengan menggunakan algoritma Support Vector Machine, sentimen kemudian diklasifikasikan menjadi positif, negatif, dan netral. Analisis ini memberikan gambaran umum tentang opini publik tentang Rancangan Undang-Undang yang dibahas.

3.2.1. Interpretasi Hasil

Hasil penelitian menunjukkan bahwa algoritma Random Forest mengklasifikasikan data yang digunakan dengan lebih baik daripada algoritma Naive Bayes. Random Forest menunjukkan kestabilan prediksi yang lebih baik dan akurasi yang lebih tinggi. Ini menunjukkan bahwa Random Forest mampu menangani variasi dan kompleksitas data dengan lebih baik. Random Forest secara teknis adalah algoritma pembelajaran kelompok yang menggabungkan banyak pohon keputusan (decision trees) yang dibangun secara acak (bagging). Setiap pohon belajar dari set data yang berbeda, sehingga model akhir lebih tahan terhadap gangguan dan kesalahan daripada model tunggal seperti Naive Bayes. Kelebihan ini terutama terlihat pada data dengan dimensi fitur yang tinggi, di mana interaksi antar fitur seringkali kompleks dan sulit ditangkap oleh algoritma yang mengasumsikan independensi fitur seperti Naive Bayes. Karena prediksi akhir merupakan rata-rata dari banyak pohon yang berbeda, Random Forest juga mengurangi risiko overfitting yang sering terjadi pada pohon keputusan tunggal. Oleh karena itu, model ini lebih umum digunakan dan dapat memberikan kinerja yang stabil saat diuji pada data baru.

3.2.2. Pembandingan dengan Studi Sebelumnya

Hasil ini sejalan dengan banyak studi sebelumnya yang telah menunjukkan bahwa Random Forest lebih baik dalam banyak domain klasifikasi. Menurut Breiman [1], pencipta Random Forest, teknik ini meningkatkan akurasi prediksi dan mengurangi overfitting dengan menggabungkan berbagai model pohon keputusan secara acak. [1]. Selain itu, penelitian menunjukkan bahwa Random Forest dapat menangani data yang memiliki banyak fitur dan memiliki kemampuan untuk secara otomatis menemukan fitur penting, yang membuatnya sangat berguna untuk aplikasi dunia nyata yang kompleks [2]. Selain itu, studi sebelumnya membandingkan Random Forest dengan Naive Bayes dalam klasifikasi data teks dan menemukan bahwa Random Forest lebih akurat dalam menangkap pola-pola non-linear dan interaksi antar kata yang Naive Bayes tidak dapat [5]. Dengan cara yang sama, penelitian mereka tentang data medis menunjukkan bahwa Random Forest lebih baik dalam menangani dataset besar yang memiliki banyak fitur dan heterogen. Itu juga menunjukkan bahwa itu tahan terhadap data yang mengandung noise dan nilai yang tidak ada. Akibatnya, keandalan prediksi meningkat. [6]. Dari perspektif metodologi, Naive Bayes mengasumsikan bahwa semua fitur bersifat independen satu sama lain, yang seringkali tidak realistis dalam data nyata yang memiliki korelasi antar fitur. Akibatnya, Naive Bayes kurang efektif dalam menangani fitur yang saling berinteraksi dan berdimensi besar, sementara Random Forest tidak memerlukan asumsi independensi tersebut dan dapat membangun model yang lebih fleksibel.

3.2.3. Penjelasan Keunggulan Teknis Random Forest

Random Forest unggul secara teknis karena beberapa faktor. Pertama, model ini menggunakan metode bootstrap aggregating, atau bagging, yang secara acak memilih subset data dan subset fitur untuk membangun masing-masing pohon keputusan, sehingga setiap pohon memiliki variasi yang berbeda dan model akhir menjadi lebih stabil dan tidak mudah overfit. Kedua, Random Forest dapat mengatasi data dengan banyak fitur karena setiap pohon hanya melihat sebagian dari fitur, yang berarti bahwa Random Forest. Ini sangat membantu dalam interpretasi model dan pengambilan keputusan. Keempat, karena agregasi hasil dari berbagai pohon keputusan dapat mengurangi dampak data anomali, Random Forest lebih tahan terhadap data yang tidak seimbang dan berisik. Random Forest banyak digunakan dalam berbagai industri, termasuk biomedis, penggalian teks, pengenalan pola, dan sistem rekomendasi, di mana data sangat kompleks dan beragam.

4. KESIMPULAN

Hasil penelitian terdiri dari pengumpulan dan pra-pemrosesan data, klasifikasi sentimen, dan evaluasi performa model. Tujuan utama penelitian ini adalah untuk mengetahui bagaimana opini publik tentang Revisi Undang-Undang Tentara Nasional Indonesia (RUU TNI) diwakili melalui media sosial. Selain itu, penelitian ini

p-ISSN: 2775-4227
e-ISSN: 2775-4219

juga bertujuan untuk mengukur seberapa baik dua algoritma pembelajaran mesin bekerja dalam mengklasifikasikan sentimen. Untuk meningkatkan kualitas teks, proses preprocessing data, yang mencakup perbaikan, case folding, tokenizing, stopword removal, dan stemming, sangat penting. Tahapan ini membantu mengurangi noise dan memastikan data lebih bersih serta konsisten untuk analisis lanjutan. Visualisasi frekuensi kata dan wordcloud menunjukkan bahwa isu terkait "RUU TNI" dan "tolakruutni" mendominasi percakapan publik. Kata-kata bermuatan negatif seperti "tolakrevisiuutni" dan "tolakdwifungsiabri" memperjelas bahwa sentimen terhadap isu tersebut cenderung kritis. Di sisi lain, muncul juga aspirasi terhadap supremasi sipil melalui kata-kata seperti "rakyat" dan "negara". Dataset yang digunakan bersifat tidak seimbang, dengan jumlah data sentimen negatif jauh lebih banyak dibandingkan positif. Untuk mengatasi ketidakseimbangan ini, digunakan pendekatan class weights pada kedua model yang diuji, yakni Naive Bayes dan Random Forest. Pada tahap pengujian model, Random Forest menunjukkan kinerja yang lebih baik dibandingkan Naive Bayes, dengan tingkat akurasi sebesar 84,96% dibandingkan 83,45% pada Naive Bayes. Random Forest juga memiliki f1-score yang lebih tinggi, khususnya dalam mendeteksi kelas positif yang lebih sulit karena minoritas. Berdasarkan hasil confusion matrix, Random Forest menghasilkan prediksi yang lebih stabil untuk kedua kelas, meskipun masih terdapat tantangan dalam meningkatkan recall untuk sentimen positif. Secara keseluruhan, penggunaan preprocessing yang tepat, pemilihan metode handling imbalance dengan class weights, serta pemilihan model Random Forest memberikan hasil yang optimal dalam klasifikasi sentimen terhadap isu Undang-Undang Tentara Nasional Indonesia (RUU TNI) ini. Model berhasil menangkap nuansa kritik dan aspirasi publik, meskipun masih ada ruang untuk perbaikan terutama dalam meningkatkan sensitivitas terhadap kelas minoritas. Kontribusi akademik dari penelitian ini adalah pengembangan framework model analisis opini publik berbasis media sosial, yang dapat digunakan untuk memahami respon masyarakat terhadap isu kebijakan nasional secara cepat dan efisien. Namun, penelitian ini memiliki beberapa keterbatasan, seperti dataset yang terbatas pada isu tertentu dan dominasi sentimen negatif yang dapat memengaruhi generalisasi model. Penelitian selanjutnya dapat mengeksplorasi metode NLP berbasis deep learning serta memperluas dataset ke isu-isu publik lainnya.

DAFTAR PUSTAKA

[1] D. A. Wulandari, R. Rohmat Saedudin, and R. Andreswari, "Analisis Sentimen Media Sosial Twitter Terhadap Reaksi Masyarakat Pada Ruu Cipta Kerja Menggunakan Metode Klasifikasi Algoritma Naive Bayes Analysis of Twitter Social Media Sentiment on the Public'S Reaction To the Drafts of Job Creation Law Using the Classification Method Naive Bayes Algorithm," *e-Proceeding Eng.*, vol. 8, no. 5, pp. 9007–9016, 2021.

[2] T. N. Wijaya, R. Indriati, and M. N. Muzaki, "Analisis Sentimen Opini Publik Tentang Undang-Undang Cipta Kerja Pada Twitter," *Jambura J. Electr. Electron. Eng.*, vol. 3, no. 2, pp. 78–83, 2021, doi: 10.37905/jjeee.v3i2.10885.

[3] K. Nugroho, F. N. Hasan, P. Korespondensi, : Firman, N. Hasan, and R. Artikel, "Analisis Sentimen Masyarakat Mengenai RUU Perampasan Aset Di Twitter Menggunakan Metode Naïve Bayes Analysis of Public Sentiment Regarding RUU Perampasan Aset on Twitter Using Naïve Bayes Method," *SMATIKA STIKI Inform. J.*, vol. 13, no. 2, p. 13, 2023.

[4] E. Fitri, "Sentiment Analysis of the Ruangguru Application Using Naive Bayes, Random Forest and Support Vector Machine Algorithms," *J. Transform.*, vol. 18, no. 1, p. 71, 2020.

[5] C. Mario, R. R. Suryono, U. T. Indonesia, and B. Lampung, "PUBLIC SENTIMENT ANALYSIS ON DIRTY VOTE MOVIE ON YOUTUBE USING RANDOM FOREST AND NAÏVE BAYES ANALISIS SENTIMEN PUBLIK PADA FILM DIRTY VOTE DI BAYES," vol. 10, no. 1, pp. 111–122, 2025.

[6] S. Fachlevi and A. A. Unde, "Analisis Sentimen Publik pada Tagar # KawalPutusanMK di Media Sosial X," vol. 5, no. 10, pp. 4232–4241, 2024.

[7] G. Darmawan, S. Alam, and M. I. Sulistyo, "Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi Mypertamina Pada Google Playstore Menggunakan Metode Naïve Bayes," *STORAGE – J. Ilm. Tek. dan Ilmu Komput.*, vol. 2, no. 3, pp. 100–108, 2023.

[8] M. M. Mala Olhang, S. Achmadi, and F. . A. Wibisono, "Analisis Sentimen Pengguna Twitter Terhadap Covid-19 Di Indonesia Menggunakan Metode Naive Bayes Classifier (Nbc)," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 4, no. 2, pp. 214–221, 2020, doi: 10.36040/jati.v4i2.2695.

[9] O. P. Zusrotun, A. C. Murti, and R. Fiati, "Analisis Sentimen Terhadap Belajar Online pada Media Sosial Twitter Menggunakan Algoritma Naive Bayes," *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 3, pp. 310–319, 2022, doi: 10.23887/janapati.v11i3.49160.

[10] R. I. Permatasari, M. A. Fauzi, P. P. Adikara, E. Dewi, and L. Sari, "Analisis Sentimen Film pada Twitter

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Berbahasa Indonesia Menggunakan Ensemble Features dan Naïve Bayes," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 11, pp. 5921–5927, 2018, [Online]. Available: http://j-ptiik.ub.ac.id

[11] A. Prasatya and N. Hendrastuty, "Analisis Sentimen : Perbandingan Performa Algoritma Naive Bayes , Support Vector Machine , Random Forest , dan K-Nearest Neighbor Dalam Pemecatan Shin Tae Yong pada Media X," vol. 6, no. 4, 2025, doi: 10.47065/bits.v6i4.6987.

[12] R. M. Awangga and N. H. Khonsa', "Analisis Performa Algoritma Random Forest dan Naive Bayes Multinomial pada Dataset Ulasan Obat dan Ulasan Film," *InComTech J. Telekomun. dan Komput.*, vol. 12, no. 1, p. 60, 2022, doi: 10.22441/incomtech.v12i1.14770.

[13] I. N. Husada and H. Toba, "Pengaruh Metode Penyeimbangan Kelas Terhadap Tingkat Akurasi Analisis Sentimen pada Tweets Berbahasa Indonesia," *J. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 2, pp. 400–413, 2020, doi: 10.28932/jutisi.v6i2.2743.

[14] Y. Asri, W. N. Suliyanti, D. Kuswardani, and M. Fajri, "Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile," *Petir*, vol. 15, no. 2, pp. 264–275, 2022, doi: 10.33322/petir.v15i2.1733.

[15] D. N. Agustia, R. R. Suryono, U. T. Indonesia, L. Ratu, and K. B. Lampung, "COMPARISON OF NAÏVE BAYES , RANDOM FOREST , AND LOGISTIC REGRESSION ALGORITHMS FOR SENTIMENT ANALYSIS ONLINE GAMBLING KOMPARASI ALGORITMA NAÏVE BAYES , RANDOM FOREST , DAN LOGISTIC REGRESION UNTUK ANALISIS," vol. 10, no. 1, pp. 284–295, 2025.

[16] J. S. Teknologi and A. Sentimen, "Lexicon-Based Approach Pada Analisis Sentimen Ulasan Airbnb Menggunakan Vader Sentiment Lexicon-Based Approach to Sentiment Analysis of Airbnb Reviews Using Vader Sentiment," vol. 5, no. 4, pp. 558–566, 2024.

[17] A. Setiawan and R. R. Suryono, "Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes," *Edumatic J. Pendidik. Inform.*, vol. 8, no. 1, pp. 183–192, 2024, doi: 10.29408/edumatic.v8i1.25667.

[18] U. Suriani, "Penerapan Data Mining untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4.5," *Journalcisa*, vol. 3, no. 2, pp. 55–66, 2023, [Online]. Available: http://jesik.web.id/index.php/jesik/article/view/91

[19] Muhammad Fernanda Naufal Fathoni, Eva Yulia Puspaningrum, and Andreas Nugroho Sihananto, "Perbandingan Performa Labeling Lexicon InSet dan VADER pada Analisa Sentimen Rohingya di Aplikasi X dengan SVM," *Modem J. Inform. dan Sains Teknol.*, vol. 2, no. 3, pp. 62–76, 2024, doi: 10.62951/modem.v2i3.112.

[20] F. A. Artanto, F. Teknik, U. Muhammadiyah, and P. Pekalongan, "Implementasi Algoritma Random Forest dan Model Bag of Words Dalam Analisis Sentimen Mengenai E-Materai," vol. 4, no. 2, pp. 139–145, 2024, doi: 10.54259/satesi.v4i2.3240.

[21] B. F. S. Supriyanto and S. Rosalin, "Analisis Sentimen Program Merdeka Belajar dengan Text Analysis Wordcloud & Word Frequency," *J. Minfo Polgan*, vol. 12, no. 1, pp. 25–32, 2023, doi: 10.33395/jmp.v12i1.12312.