

Perbandingan Algoritma Naive Bayes, Random Forest, dan Support Vector Machine Terhadap Pandangan Masyarakat Mengenai Revisi Undang-Undang TNI di Instagram

Royhan Nasrul, Aditia Yudhistira*

Fakultas Teknik dan Ilmu Komputer, Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ¹royhan_nasrul@teknokrat.ac.id, ^{2,*}aditiayudhistira@teknokrat.ac.id

Email Penulis Korespondensi: aditiayudhistira@teknokrat.ac.id

Submitted: 04/08/2025; Accepted: 04/09/2025; Published: 30/09/2025

Abstrak—Revisi Undang-Undang Tentara Nasional Indonesia (UU TNI) yang disahkan pada tahun 2025 menimbulkan kontroversi di masyarakat, terutama terkait isu supremasi sipil dan potensi dominasi militer. Seiring meningkatnya penggunaan media sosial sebagai ruang ekspresi publik, platform seperti Instagram menjadi sarana utama bagi masyarakat untuk menyampaikan opini terkait isu tersebut. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap revisi UU TNI dengan memanfaatkan algoritma klasifikasi teks, yaitu Naive Bayes, Random Forest, dan Support Vector Machine (SVM). Data diperoleh dari 28.669 komentar Instagram dan dianalisis melalui tahapan crawling, preprocessing, dan pelabelan. Untuk mengatasi ketidakseimbangan data, kemudian diterapkan teknik SMOTE. Selanjutnya, dilakukan klasifikasi menggunakan ketiga algoritma dengan evaluasi berdasarkan akurasi, precision, recall, dan F1-score. Hasil penelitian setelah SMOTE menunjukkan bahwa algoritma SVM memiliki performa terbaik dengan akurasi 92%, diikuti oleh Random Forest (88%) dan Naive Bayes (76%). Dengan demikian, SVM dinilai paling efektif dalam menangkap pola sentimen publik secara objektif. Penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan kajian opini publik digital dan mendukung proses evaluasi kebijakan pertahanan nasional.

Kata Kunci: Algoritma Klasifikasi; Instagram; Naive Bayes; Random Forest; Sentiment Analysis; Support Vector Machine

Abstract—The revision of the Indonesian National Army Law (TNI Law), enacted in 2025, sparked widespread controversy within society, particularly concerning issues of civilian supremacy and potential military dominance. With the growing use of social media as a platform for public expression, platforms such as Instagram have become the primary medium for the public to voice their opinions regarding this issue. This study aims to analyze public sentiment toward the revision of the TNI Law by utilizing text classification algorithms, namely Naive Bayes, Random Forest, and Support Vector Machine (SVM). Data was collected from 28,669 Instagram comments and analyzed through stages of data crawling, preprocessing, and labeling. To address data imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. Subsequently, classification was performed using the three algorithms, with evaluation metrics including accuracy, precision, recall, and F1-score. The results after SMOTE demonstrated that the SVM algorithm delivered the best performance with an accuracy of (92%), followed by Random Forest at (88%), and Naive Bayes at (76%). Consequently, SVM was deemed the most effective in capturing patterns of public sentiment objectively. This research is expected to contribute to the advancement of digital public opinion studies and support the evaluation process of national defense policies

Keywords: Classification Algorithm; Instagram; Naive Bayes; Random Forest; Sentiment Analysis; Support Vector Machine

1. PENDAHULUAN

Revisi terhadap Undang-Undang Tentara Nasional Indonesia (UU TNI) yang telah disahkan pada 20 Maret 2025 menimbulkan kontroversi dan kekhawatiran di berbagai kalangan masyarakat sipil, khususnya dari kelompok yang berfokus pada isu demokrasi dan hak asasi manusia. Salah satu pihak yang secara terbuka menyuarakan penolakan adalah Kaukus Indonesia untuk Kebebasan Akademik (KIKA), yang menilai bahwa revisi tersebut berpotensi melemahkan prinsip-prinsip negara hukum dan supremasi sipil, serta dapat membuka ruang bagi terjadinya impunitas di lingkungan militer [1]. Substansi revisi yang dianggap problematik meliputi sejumlah pasal penting, antara lain Pasal 7 yang memperluas tugas Tentara Nasional Indonesia dalam Operasi Militer Selain Perang (OMSP), Pasal 47 yang memberikan peluang bagi prajurit aktif untuk menduduki jabatan-jabatan sipil, serta Pasal 53 yang menaikkan batas usia pensiun prajurit. Perubahan-perubahan ini dinilai oleh sejumlah pihak sebagai indikasi menguatnya kembali peran militer di luar sektor pertahanan, yang berpotensi menyerupai kondisi dominasi militer pada masa Orde Baru [2].

Lebih lanjut, substansi revisi tersebut dinilai bertentangan dengan prinsip-prinsip yang terkandung dalam instrumen hak asasi manusia internasional seperti *International Covenant on Civil and Political Rights* (ICCPR) dan *Convention Against Torture* (CAT), yang telah disetujui oleh Pemerintah Indonesia. Ketidaksesuaian ini menimbulkan kekhawatiran akan penurunan akuntabilitas militer, pelemahan perlindungan terhadap kebebasan sipil, serta gangguan terhadap proses konsolidasi demokrasi di Indonesia [3]. Perkembangan transformasi digital telah mengubah pola komunikasi publik secara signifikan, di mana media sosial seperti Instagram kini memainkan peran penting sebagai ruang publik digital. Melalui platform ini, masyarakat berpartisipasi secara aktif dalam menyampaikan pendapat, kritik, serta aspirasi mereka terhadap berbagai isu kebijakan publik, termasuk dalam topik terkini mengenai revisi Undang-Undang Tentara Nasional Indonesia (UU TNI). Oleh karena itu, penting untuk mengevaluasi persepsi masyarakat terhadap revisi UU ini, khususnya melalui analisis terhadap opini publik yang berkembang di media sosial sebagai ruang ekspresi demokratis modern [4].

Untuk mencapai tujuan tersebut, pendekatan analisis sentimen (*sentiment analysis*), yang didukung oleh kemajuan teknologi *deep learning* dan *Natural Language Processing* (NLP), menyediakan kerangka metodologis yang komprehensif. Metode ini memungkinkan pemrosesan dan pengelompokan otomatis terhadap data teks tidak terstruktur berdasarkan polaritas sentimen, yakni positif, negatif, atau netral [5]. Dalam konteks penelitian ini, analisis sentimen digunakan untuk mengkaji opini masyarakat melalui komentar yang diunggah di media sosial Instagram, yang kerap menjadi wadah penyampaian pendapat publik secara terbuka. Berbagai algoritma klasifikasi seperti *Naive Bayes*, *Random Forest*, dan *Support Vector Machine* (SVM) dipilih untuk dibandingkan kinerjanya dalam mengelompokkan data komentar berdasarkan sentimen. Dengan cara ini, diharapkan penerapan tiga jenis model algoritma pada bidang analisis sentimen dapat memberikan dasar evaluasi berbasis data yang objektif dalam menggambarkan persepsi publik terhadap isu-isu strategis seperti revisi Undang-Undang TNI.

Beberapa penelitian terdahulu yang mengkaji topik serupa maupun berbeda menggunakan berbagai model algoritma yang telah digunakan secara luas dalam bidang analisis sentimen. Seperti salah satu penelitian yang menggunakan model algoritma *Support Vector Machine* (SVM) dengan data yang diperoleh dari platform media sosial X selama periode 1 hingga 31 Maret 2025, serta menerapkan pendekatan *Term Frequency-Inverse Document Frequency* (TF-IDF), menunjukkan bahwa model SVM mampu mencapai rata-rata akurasi sebesar 78,99%. Selain itu, nilai *F1-Score* tertinggi tercatat pada kategori sentimen negatif, yakni sebesar 83%, dari total 500 data tweet yang dianalisis. Distribusi data menunjukkan bahwa 58,99% merupakan sentimen negatif dan 41,01% merupakan sentimen positif [6].

Sementara itu, studi lain oleh Elin Haerani dan tim yang juga meneliti topik serupa, namun menggunakan algoritma *Naive Bayes Classifier*, menganalisis 1.547 data komentar dan memperoleh akurasi yang lebih tinggi, yaitu 83,74%. Model ini juga menunjukkan performa yang kuat dengan *precision* sebesar 81,17%, *recall* 87,86%, dan *F1-Score* 84,38%. Mengetahui metode yang paling akurat. Melalui pendekatan ini, tulisan diharapkan dapat memberikan kontribusi terhadap kajian opini publik digital dan pemanfaatan teknologi dalam evaluasi kebijakan pertahanan [7]. Sementara itu, studi dengan fokus berbeda namun tetap relevan dalam konteks analisis komentar di media sosial, menerapkan algoritma *Random Forest* untuk mendeteksi komentar bermuatan negatif dan unsur *cyberbullying* di platform Instagram. Hasilnya menunjukkan bahwa model ini mampu mencapai akurasi sebesar 84%, yang menunjukkan efektivitasnya dalam mengidentifikasi komentar dengan potensi perundungan daring [8].

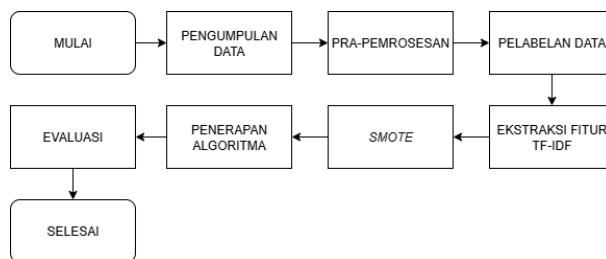
Mengacu pada latar belakang penelitian terdahulu dengan menerapkan ketiga pendekatan algoritmik *Naive Bayes*, *Random Forest* dan *Support Vector Machine*, dapat disimpulkan bahwa masing-masing memiliki keunggulan tersendiri dalam analisis sentimen terhadap data teks di media sosial. Oleh karena itu, penelitian ini bertujuan untuk membandingkan kinerja ketiga algoritma tersebut dalam menganalisis sentimen publik terhadap isu strategis, yakni Revisi Undang-Undang Tentara Nasional Indonesia (RUU TNI), dengan menggunakan data komentar dari media sosial Instagram. Melalui pendekatan ini, diharapkan hasil penelitian dapat memberikan kontribusi terhadap pengembangan kajian opini publik digital serta pemanfaatan teknologi dalam proses evaluasi kebijakan pertahanan. Selain itu, temuan dari penelitian ini juga diharapkan mampu memberikan masukan yang konstruktif bagi para pemangku kebijakan dalam memahami persepsi masyarakat secara lebih objektif terhadap kebijakan yang sedang dirancang maupun yang akan datang.

2. METODOLOGI PENELITIAN

Penelitian ini menerapkan pendekatan kuantitatif deskriptif dengan memanfaatkan teknik *text mining* dan *machine learning* untuk melakukan analisis sentimen terhadap opini masyarakat di media sosial Instagram. Data diperoleh melalui pengumpulan komentar pengguna Instagram dari berbagai postingan yang membahas isu Revisi Undang-Undang Tentara Nasional Indonesia (UU TNI), yang diketahui telah memicu berbagai reaksi dan perdebatan di kalangan masyarakat. Objek dalam penelitian ini adalah komentar-komentar dari warganet yang merepresentasikan sentimen publik, yang dikumpulkan secara purposif dari akun-akun atau unggahan yang relevan dengan tema pembahasan. Proses pengambilan data dilakukan dengan teknik *web scraping* menggunakan *tools* yang mendukung pengumpulan data teks dari Instagram.

2.1 Tahapan Penelitian

Penelitian ini diawali dengan pengumpulan data sejumlah 28.669 komentar publik dari media sosial Instagram. Data mentah tersebut kemudian melewati tahap pra-pemrosesan (*preprocessing*) yang meliputi pembersihan data (*data cleaning*) untuk menghilangkan *noise* seperti tanda baca, angka, dan URL, normalisasi teks ke format huruf kecil (*lowercase*), serta penghapusan *stopwords*. Setelah data bersih dan terstruktur, dilakukan pelabelan untuk mengkategorikan setiap komentar ke dalam sentimen positif, negatif, atau netral. Untuk mengatasi masalah ketidakseimbangan distribusi kelas pada dataset, penelitian ini mengaplikasikan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) guna menambah data secara sintesis pada kelas minoritas. Tahap analisis melibatkan implementasi tiga algoritma klasifikasi, yaitu *Support Vector Machine* (SVM), *Random Forest*, dan *Naive Bayes*. Kinerja ketiga model tersebut dievaluasi secara komparatif pada dataset sebelum dan sesudah penerapan SMOTE. Evaluasi performa didasarkan pada metrik akurasi, *precision*, *recall*, dan *F1-score*, dengan hasil yang divisualisasikan melalui *confusion matrix*. Keseluruhan alur metodologi penelitian ini dirangkum pada Gambar 1.



Gambar 1. Diagram alur penelitian

2.2 Crawling Data

Tahap pengumpulan data, sebagai langkah awal dalam penelitian ini, dilakukan dengan teknik *crawling* untuk mengumpulkan komentar publik dari platform media sosial Instagram. Proses ekstraksi data ini dijalankan secara otomatis dengan memanfaatkan layanan pihak ketiga, yaitu *platform* spesialis *web crawling* *EasyComment.ai*. Penggunaan platform ini bertujuan untuk memastikan efisiensi dan skalabilitas dalam pengumpulan data bervolume besar, yang kemudian menghasilkan sebuah dataset mentah (*raw dataset*) sebagai dasar untuk tahap analisis selanjutnya. *Crawling* atau *scraping* dalam bidang analisis sentimen adalah sebuah proses atau tahapan awal mengumpulkan data text dari berbagai sumber, salah satu contoh nya adalah media sosial [9].

2.3 Pra-Pemrosesan

Sebelum dilakukan proses analisis sentimen, data komentar terlebih dahulu diproses melalui tahapan pra-pemrosesan untuk meningkatkan kualitas, kebersihan, dan konsistensi data teks. Tahapan ini penting dilakukan dalam pengolahan data teks agar dapat diubah menjadi bentuk yang terstruktur dan dapat diterima oleh algoritma klasifikasi. Adapun tahapan-tahapan pra-pemrosesan yang dilakukan dalam penelitian ini merujuk pada praktik umum dalam bidang *Natural Language Processing (NLP)* dan *text mining* [10], sebagai berikut :

- Penghapusan data duplikat dan NaN: Setelah data terkumpul, langkah pra-pemrosesan awal difokuskan pada pembersihan dataset. Ini melibatkan penghapusan semua komentar duplikat untuk memastikan keunikan data serta membuang entri yang bernilai kosong (NaN). Tindakan ini esensial untuk mengurangi bias akibat repetisi dan meningkatkan validitas data sebelum masuk ke tahap analisis [11].
- Data Cleaning*: Pada tahap ini dilakukan pembersihan karakter-karakter yang tidak relevan seperti simbol, tanda baca, angka, dan tautan (URL) dari setiap komentar. Tujuan dari tahap ini adalah untuk menghilangkan *noise* atau gangguan dalam data yang dapat menurunkan akurasi model [12]. Tabel 1 berikut merupakan contoh hasil dari *Data Cleaning*.

Tabel 1. Contoh hasil data cleaning

Data Mentah	Data Cleaning
Kek na oke2 aja tuh ruu tni. Point mana yg bikin demo? Tni dibawah presiden? Ga terimakasih? Tni mengamankan presiden dan wapres beserta keluarganya? Iri hatikah? Mengatasi aksi terorisme? Krn pd rencana teroris?	Kek na oke aja tuh ruu tni Point mana yg bikin demo Tni dibawah presiden Ga terimakasih Tni mengamankan presiden dan wapres beserta keluarganya Iri hatikah Mengatasi aksi terorisme Krn pd rencana teroris
kasus TNI belakangan ini ialah : bunuh bos rental, bunuh casis TNI, bunuh anak kecil di aceh, bunuh bos showroom di aceh, bunuh polisi lg tugas, serang mako polres tarakan, di tilang lantas ngajak berantem, suplay senjata ke rekannya di papua sheaarga 1,3 M, disuruh masuk peradilan umum ga mau tapi pengen jabatan sipil umum gilaa si satuan ini. gile gileeee 🤔 ga kebayang kl RUU TNI di Sahkan bgmana nasib sipil 🤔	kasus TNI belakangan ini ialah bunuh bos rental bunuh casis TNI bunuh anak kecil di aceh bunuh bos showroom di aceh bunuh polisi lg tugas serang mako polres tarakan di tilang lantas ngajak berantem suplay senjata ke rekannya di papua sheaarga M disuruh masuk peradilan umum ga mau tapi pengen jabatan sipil umum gilaa si satuan ini gile gileeee ga kebayang kl RUU TNI di Sahkan bgmana nasib sipil
kenapa perihala ruu perampasan aset ga ada yg ngedukung kek pas ngeprotes ruu tni ini ya 🤔 komplikasi kah? 🤔	kenapa perihala ruu perampasan aset ga ada yg ngedukung kek pas ngeprotes ruu tni ini ya komplikasi kah

- Case Folding*: Tahap selanjutnya adalah *case folding*, yaitu proses dalam pemrosesan teks yang bertujuan untuk mengubah seluruh huruf dalam suatu kalimat menjadi huruf kecil (*lowercase*). Proses ini diperlukan karena penggunaan huruf kapital dalam teks tidak selalu konsisten, sehingga diperlukan normalisasi bentuk huruf untuk memastikan keseragaman data dan menghindari duplikasi makna pada kata yang sama [13]. Pada Tabel 2 berikut ini merupakan contoh hasil dari *Case Folding*.

Tabel 2. Contoh Hasil Case Folding

Data Cleaning	Case Folding
Kek na oke aja tuh ruu tni Point mana yg bikin demo Tni dibawah presiden Ga terimakasih Tni mengamankan presiden dan wapres beserta keluarganya Iri hatikah Mengatasi aksi terorisme Krn pd rencana teroris kasus TNI belakangan ini ialah bunuh bos rental bunuh casis TNI bunuh anak kecil di aceh bunuh bos showroom di aceh bunuh polisi lg tugas serang mako polres tarakan di tilang lantas ngajak berantem suplay senjata ke rekannya di papua sejaarga M disuruh masuk peradilan umum ga mau tapi pengen jabatan sipil umum gilaa si satuan ini gile gileeee ga kebayang kl RUU TNI di Sahkan bgmana nasib sipil kenapa perihala ruu perampasan aset ga ada yg ngedukung kek pas ngeprotes ruu tni ini ya komplikasi kah	kek na oke aja tuh ruu tni point mana yg bikin demo tni dibawah presiden ga terimakasih tni mengamankan presiden dan wapres beserta keluarganya iri hatikah mengatasi aksi terorisme krn pd rencana teroris kasus tni belakangan ini ialah bunuh bos rental bunuh casis tni bunuh anak kecil di aceh bunuh bos showroom di aceh bunuh polisi lg tugas serang mako polres tarakan di tilang lantas ngajak berantem suplay senjata ke rekannya di papua sejaarga m disuruh masuk peradilan umum ga mau tapi pengen jabatan sipil umum gilaa si satuan ini gile gileeee ga kebayang kl ruu tni di sahan bgmana nasib sipil kenapa perihala ruu perampasan aset ga ada yg ngedukung kek pas ngeprotes ruu tni ini ya komplikasi kah

d. *Tokenizing*: *Tokenizing* adalah proses dalam pra-pemrosesan teks yang bertujuan untuk memecah kalimat atau dokumen menjadi unit-unit kata yang lebih kecil yang disebut sebagai token. Setiap token biasanya merepresentasikan satu kata atau istilah dalam teks. Tahapan ini penting dalam analisis teks karena memungkinkan sistem untuk menganalisis dan memproses setiap kata secara terpisah. Dengan melakukan tokenisasi, data teks menjadi lebih terstruktur dan dapat diubah menjadi representasi numerik yang dapat digunakan dalam proses ekstraksi fitur dan pemodelan *machine learning* [14]. Pada tabel 3 dibawah ini merupakan contoh hasil dari *tokenizing*.

Tabel 3. Contoh hasil tokenizing

Case Folding	Tokenizing
kek na oke aja tuh ruu tni point mana yg bikin demo tni dibawah presiden ga terimakasih tni mengamankan presiden dan wapres beserta keluarganya iri hatikah mengatasi aksi terorisme krn pd rencana teroris	['kayak', 'nya', 'oke', 'saja', 'tuh', 'ruu', 'tni', 'point', 'mana', 'yang', 'bikin', 'demo', 'tni', 'dibawah', 'presiden', 'tidak', 'terimakasih', 'tni', 'mengamankan', 'presiden', 'dan', 'wapres', 'beserta', 'keluarganya', 'iri', 'hatikah', 'mengatasi', 'aksi', 'terorisme', 'karena', 'pada', 'rencana', 'teroris']
kasus TNI belakangan ini ialah : bunuh bos rental, bunuh casis TNI, bunuh anak kecil di aceh, bunuh bos showroom di aceh, bunuh polisi lg tugas, serang mako polres tarakan, di tilang lantas ngajak berantem, suplay senjata ke rekannya di papua sejaarga 1,3 M, disuruh masuk peradilan umum ga mau tapi pengen jabatan sipil umum gilaa si satuan ini. gile gileeee 😞 ga kebayang kl RUU TNI di Sahkan bgmana nasib sipil 😞	['kasus', 'tni', 'belakangan', 'ini', 'ialah', 'bunuh', 'bos', 'rental', 'bunuh', 'casis', 'tni', 'bunuh', 'anak', 'kecil', 'di', 'aceh', 'bunuh', 'bos', 'showroom', 'di', 'aceh', 'bunuh', 'polisi', 'lagi', 'tugas', 'serang', 'mako', 'polres', 'tarakan', 'di', 'tilang', 'lantas', 'mengajak', 'berantem', 'suplay', 'senjata', 'ke', 'rekannya', 'di', 'papua', 'sehaarga', 'sama', 'disuruh', 'masuk', 'peradilan', 'umum', 'tidak', 'mau', 'tapi', 'pengin', 'jabatan', 'sipil', 'umum', 'gila', 'sih', 'satuan', 'ini', 'gila', 'gileeee', 'tidak', 'terbayang', 'kalau', 'ruu', 'tni', 'di', 'sahkan', 'bagaimana', 'nasib', 'sipil']
kenapa perihala ruu perampasan aset ga ada yg ngedukung kek pas ngeprotes ruu tni ini ya 🤔 komplikasi kah? 🤔	['kenapa', 'perihala', 'ruu', 'perampasan', 'aset', 'tidak', 'ada', 'yang', 'mendukung', 'kayak', 'pas', 'ngeprotes', 'ruu', 'tni', 'ini', 'ya', 'komplikasi', 'kah']

e. *Stopword Removal*: Tahap berikutnya adalah *stopword removal*, yaitu proses penghapusan kata-kata umum dalam teks yang dianggap tidak memiliki kontribusi signifikan terhadap analisis, seperti kata sambung, kata depan, atau partikel. Langkah ini dilakukan untuk menyederhanakan data dan meningkatkan efisiensi pemrosesan tanpa mengurangi makna utama dari teks [15]. Pada Tabel 4 dibawah ini merupakan contoh hasil *stopword removal*.

Tabel 4. Contoh hasil stopword removal.

Tokenizing	Stopword Removal
['kayak', 'nya', 'oke', 'saja', 'tuh', 'ruu', 'tni', 'point', 'mana', 'yang', 'bikin', 'demo', 'tni', 'dibawah', 'presiden', 'tidak', 'terimakasih', 'tni', 'mengamankan', 'presiden', 'dan', 'wapres', 'beserta', 'keluarganya', 'iri', 'hatikah', 'mengatasi', 'aksi', 'terorisme', 'karena', 'pada', 'rencana', 'teroris']	['kayak', 'nya', 'oke', 'tuh', 'ruu', 'tni', 'point', 'bikin', 'demo', 'tni', 'dibawah', 'presiden', 'terimakasih', 'tni', 'mengamankan', 'presiden', 'wapres', 'beserta', 'keluarganya', 'iri', 'hatikah', 'mengatasi', 'aksi', 'terorisme', 'rencana', 'teroris']



Tokenizing	Stopword Removal
['kasus', 'tni', 'belakangan', 'ini', 'ialah', 'bunuh', 'bos', 'rental', 'bunuh', 'casis', 'tni', 'bunuh', 'anak', 'kecil', 'di', 'aceh', 'bunuh', 'bos', 'showroom', 'di', 'aceh', 'bunuh', 'polisi', 'lagi', 'tugas', 'serang', 'mako', 'polres', 'tarakan', 'di', 'tilang', 'lantas', 'mengajak', 'berantem', 'suplay', 'senjata', 'ke', 'rekannya', 'di', 'papua', 'sehaarga', 'sama', 'disuruh', 'masuk', 'peradilan', 'umum', 'tidak', 'mau', 'tapi', 'pengin', 'jabatan', 'sipil', 'umum', 'gila', 'sih', 'satuan', 'ini', 'gila', 'gileeee', 'tidak', 'terbayang', 'kalau', 'ruu', 'tni', 'di', 'sahkan', 'bagaimana', 'nasib', 'sipil']	['tni', 'bunuh', 'bos', 'rental', 'bunuh', 'casis', 'tni', 'bunuh', 'anak', 'aceh', 'bunuh', 'bos', 'showroom', 'aceh', 'bunuh', 'polisi', 'tugas', 'serang', 'mako', 'polres', 'tarakan', 'tilang', 'lantas', 'mengajak', 'berantem', 'suplay', 'senjata', 'rekannya', 'papua', 'sehaarga', 'disuruh', 'masuk', 'peradilan', 'pengin', 'jabatan', 'sipil', 'gila', 'sih', 'satuan', 'gila', 'gileeee', 'terbayang', 'ruu', 'tni', 'sahkan', 'nasib', 'sipil']
['kenapa', 'perihala', 'ruu', 'perampasan', 'aset', 'tidak', 'ada', 'yang', 'mendukung', 'kayak', 'pas', 'ngeprotes', 'ruu', 'tni', 'ini', 'ya', 'komplikasi', 'kah']	['perihala', 'ruu', 'perampasan', 'aset', 'mendukung', 'kayak', 'pas', 'ngeprotes', 'ruu', 'tni', 'ya', 'komplikasi', 'kah']

f. *Stemming*: Langkah terakhir dari pra-pemrosesan data *Stemming* adalah proses dalam pra-pemrosesan teks yang bertujuan untuk mengubah kata turunan menjadi bentuk dasarnya (akar kata). Teknik ini digunakan untuk menyamakan berbagai variasi morfologis dari sebuah kata agar dapat dianalisis sebagai satu entitas yang sama. Misalnya, kata "menjalankan", "berjalan", dan "dijalankan" akan direduksi menjadi bentuk dasar "jalan" [16]. Pada Tabel 5 merupakan contoh hasil dari *Stemming*.

Tabel 5. Contoh hasil stemming

Stopword Removal	Stemming
['kayak', 'nya', 'oke', 'tuh', 'ruu', 'tni', 'point', 'bikin', 'demo', 'tni', 'dibawah', 'presiden', 'terimakasih', 'tni', 'mengamankan', 'presiden', 'wapres', 'beserta', 'keluarganya', 'iri', 'hatikah', 'mengatasi', 'aksi', 'terorisme', 'rencana', 'teroris']	kayak nya oke tuh ruu tni point bikin demo tni bawah presiden terima tni aman presiden wapres serta keluarga iri hati atas aksi terorisme rencana teroris
['tni', 'bunuh', 'bos', 'rental', 'bunuh', 'casis', 'tni', 'bunuh', 'anak', 'aceh', 'bunuh', 'bos', 'showroom', 'aceh', 'bunuh', 'polisi', 'tugas', 'serang', 'mako', 'polres', 'tarakan', 'tilang', 'lantas', 'mengajak', 'berantem', 'suplay', 'senjata', 'rekannya', 'papua', 'sehaarga', 'disuruh', 'masuk', 'peradilan', 'pengin', 'jabatan', 'sipil', 'gila', 'sih', 'satuan', 'gila', 'gileeee', 'terbayang', 'ruu', 'tni', 'sahkan', 'nasib', 'sipil']	tni bunuh bos rental bunuh cas tni bunuh anak aceh bunuh bos showroom aceh bunuh polisi tugas serang mako polres tara tilang lantas ajak antem suplay senjata rekan papua sehaarga suruh masuk adil pengin jabat sipil gila sih satu gila gileeee bayang ruu tni sahkan nasib sipil
['perihala', 'ruu', 'perampasan', 'aset', 'mendukung', 'kayak', 'pas', 'ngeprotes', 'ruu', 'tni', 'ya', 'komplikasi', 'kah']	perihala ruu ampas aset dukung kayak pas ngeprotes ruu tni ya komplikasi kah

2.4 Pelabelan Data

Pelabelan data adalah tahapan dalam pemrosesan data yang bertujuan untuk memberikan identitas atau klasifikasi pada setiap data mentah berdasarkan karakteristik atau isi dari data tersebut. Dalam analisis sentimen dan *text mining*, pelabelan dilakukan dengan menentukan apakah suatu teks termasuk dalam sentimen positif, negatif, atau netral. Proses ini memiliki peran penting dalam pengembangan model *machine learning* berbasis *supervised learning*, karena model tersebut hanya dapat belajar secara optimal jika diberi data yang telah diklasifikasikan sebelumnya untuk kemudian digunakan dalam proses prediksi pada data yang belum dikenal [17]. Pada Tabel 6 merupakan contoh hasil dari pelabelan sentimen komentar pada penelitian.

Tabel 6. Contoh Hasil Pelabelan

Data Mentah	Label
Kek na oke2 aja tuh ruu tni. Point mana yg bikin demo? Tni dibawah presiden? Ga terimakasih? Tni mengamankan presiden dan wapres beserta keluarganya? Iri hatikah? Mengatasi aksi terorisme? Krn pd rencana teroris?	Positif
kasus TNI belakangan ini ialah : bunuh bos rental, bunuh casis TNI, bunuh anak kecil di aceh, bunuh bos showroom di aceh, bunuh polisi lg tugas, serang mako polres tarakan, di tilang lantas ngajak berantem, suplay senjata ke rekannya di papua sehaarga 1,3 M, disuruh masuk peradilan umum ga mau tapi pengen jabatan sipil umum gilaa si satuan ini. gile gileeee 😡 ga kebayang kl RUU TNI di Sahkan bgmana nasib sipil 😡	Negatif
ngedukung kek pas ngeprotes ruu tni ini ya 🤔 komplikasi kah? 🤔	Netral

Berikut ini adalah tampilan visualisasi dari hasil pelabelan berupa *wordcloud* dari tiga jenis sentimen analisis pada topik RUU TNI yang dapat dilihat pada Gambar 2, 3, dan 4. Gambar 2 menunjukkan visualisasi *wordcloud* untuk sentimen positif, yang menampilkan kata-kata seperti "rakyat", "Indonesia", "negara", "TNI", dan "presiden", yang mengindikasikan adanya dukungan publik terhadap penguatan peran TNI dalam menjaga stabilitas negara. Gambar 3

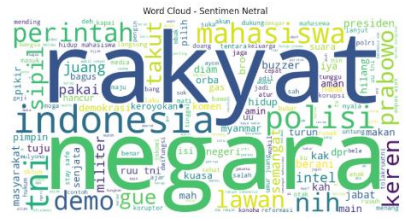
menampilkan *wordcloud* untuk sentimen negatif, dengan kata-kata dominan seperti “demo”, “negara”, “militer”, “korupsi”, dan “salah”, yang merefleksikan kritik atau kekhawatiran publik terhadap potensi penyalahgunaan kekuasaan serta ancaman terhadap demokrasi. Sementara itu, Gambar 4 memperlihatkan *wordcloud* untuk sentimen netral dengan kata-kata seperti “rakyat”, “negara”, “mahasiswa”, “pemerintah”, dan “politik”, yang mencerminkan narasi informatif atau netral dalam menyampaikan opini mengenai RUU TNI tanpa muatan emosi yang kuat. Visualisasi ini memberikan gambaran umum mengenai distribusi kosakata yang mewakili sikap publik terhadap topik yang sedang dibahas.



Gambar 2. Visualisasi *wordcloud* sentimen positif



Gambar 3. Visualisasi *wordcloud* sentimen negatif



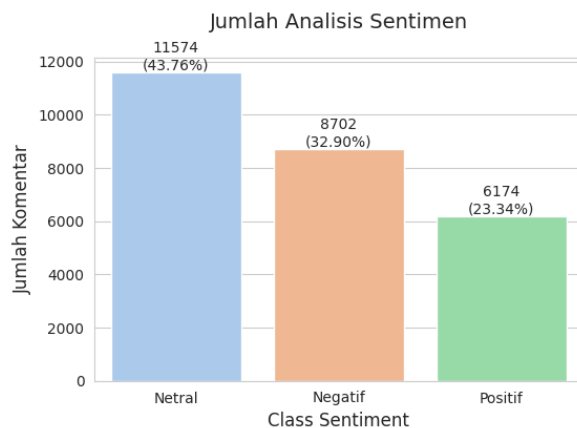
Gambar 4. Visualisasi *wordcloud* sentimen netral

2.5 Ekstraksi Fitur TF-IDF

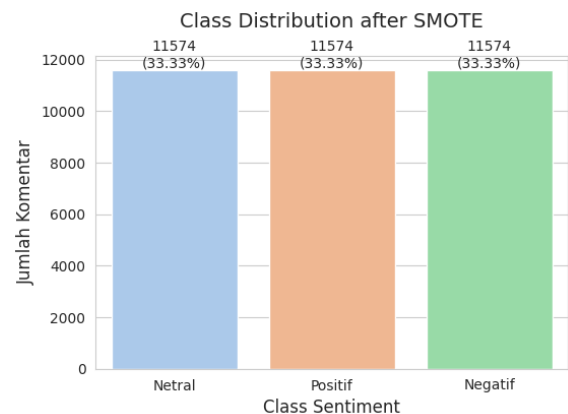
TF-IDF merupakan metode yang efektif dalam memberikan bobot terhadap kata-kata dalam dokumen teks berdasarkan frekuensi kemunculannya secara lokal (dalam satu dokumen) dan secara global (di seluruh dokumen). Dengan menggabungkan dua komponen utama, yaitu *Term Frequency* dan *Inverse Document Frequency*, TF-IDF mampu menyoroti kata-kata yang memiliki makna penting dan relevan dalam konteks tertentu, serta mengurangi pengaruh kata-kata umum yang sering muncul namun kurang informatif. Oleh karena itu, TF-IDF menjadi salah satu teknik dasar yang sangat penting dan banyak digunakan dalam berbagai aplikasi analisis teks, seperti klasifikasi dokumen, pencarian informasi, dan analisis sentimen [18].

2.6 Synthetic Minority Over-sampling Technique (SMOTE)

SMOTE (*Synthetic Minority Over-sampling Technique*) adalah suatu teknik pemrosesan data yang digunakan untuk menangani permasalahan ketidakseimbangan kelas (*class imbalance*) dalam dataset, terutama pada konteks klasifikasi. Metode ini bekerja dengan cara menciptakan data sintesis baru pada kelas minoritas, bukan dengan melakukan duplikasi data secara langsung, melainkan dengan membangun sampel baru berdasarkan interpolasi antara sampel minoritas yang ada dan tetangga terdekatnya. Dengan menerapkan SMOTE, distribusi data menjadi lebih seimbang, sehingga algoritma klasifikasi dapat belajar secara lebih adil dan tidak bias terhadap kelas mayoritas. Teknik ini umum digunakan dalam skenario *supervised learning*, terutama ketika model menunjukkan performa rendah dalam mendeteksi kelas dengan jumlah data yang sedikit [19]. Berikut merupakan visualisasi perbedaan data penelitian sebelum SMOTE pada Gambar 5 dan setelah dilakukan proses SMOTE pada Gambar 6.



Gambar 5. Perbandingan jumlah data sebelum SMOTE



Gambar 6. Perbandingan jumlah data sesudah SMOTE

Berdasarkan hasil diatas, sebelum penerapan SMOTE distribusi data sentimen tidak seimbang, dengan kelas netral yang mendominasi 43,78%, sedangkan kelas positif dan negatif jauh lebih sedikit. Ketidakseimbangan data ini dapat memengaruhi akurasi model karena bias terhadap kelas mayoritas. Setelah diterapkan SMOTE semua data menjadi seimbang dengan pembagian 33,33%. Dengan distribusi seimbang, model klasifikasi diharapkan menjadi lebih adil dalam mengenali jenis sentimen dan meningkatkan performa keseluruhan analisis.

2.7 Penerapan Algoritma

Untuk keperluan klasifikasi sentimen teks, penelitian ini menerapkan tiga algoritma, yaitu *Support Vector Machine* (SVM), *Random Forest*, dan *Naive Bayes*, yang dikenal memiliki performa baik dalam menangani data teks dengan karakteristik yang kompleks. Algoritma SVM digunakan untuk membedakan kelas sentimen dengan mencari batas pemisah terbaik antar kelas [20]. Sementara itu, *Random Forest* membangun sejumlah pohon keputusan yang digabungkan untuk menghasilkan prediksi akhir yang lebih akurat dan tahan terhadap *overfitting* [21]. Sebaliknya, *Naive Bayes* melakukan klasifikasi teks dengan menghitung probabilitas munculnya kata-kata, sambil mengasumsikan bahwa setiap fitur bersifat independen [22]. Proses klasifikasi dimulai dari teks yang telah dibersihkan melalui tahapan pra-pemrosesan, lalu diubah menjadi bentuk vektor numerik menggunakan metode TF-IDF. Setelah itu, dataset dibagi menjadi dua bagian, yakni data pelatihan 80% (X_{train} , y_{train}) dan 20% data pengujian (X_{test} , y_{test}), yang selanjutnya digunakan untuk melatih dan menguji kinerja model dalam mengenali sentimen publik terhadap revisi UU TNI.

2.8 Evaluasi Model

Pada tahap ini, penelitian memaparkan hasil implementasi dari keseluruhan proses yang telah dilakukan, khususnya penerapan algoritma *Support Vector Machine* (SVM), *Random Forest*, dan *Naive Bayes* dalam analisis sentimen terhadap opini masyarakat mengenai Revisi Undang-Undang TNI. Tujuan dari analisis ini adalah untuk mengidentifikasi pola sentimen serta memahami kecenderungan pandangan publik terhadap isu tersebut. Evaluasi kinerja masing-masing model dilakukan dengan mengacu pada empat indikator utama, yaitu akurasi, presisi, *recall*, dan *F1-score*. Berikut rumus untuk menentukan evaluasi kinerja model pada penelitian ini.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - \text{Score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (4)$$

True Positive (TP) merujuk pada jumlah data yang memang positif dan berhasil diprediksi dengan benar oleh model, sedangkan *True Negative* (TN) menunjukkan jumlah data yang memang negatif dan diklasifikasikan dengan tepat. Di sisi lain, *False Positive* (FP) menggambarkan data yang sebenarnya negatif tetapi keliru diprediksi sebagai positif, sementara *False Negative* (FN) merujuk pada data yang seharusnya positif namun salah diklasifikasikan sebagai negatif. Hasil dari ketiga algoritma tersebut kemudian dibandingkan untuk menentukan model yang paling efektif dalam melakukan klasifikasi sentimen secara akurat, sehingga dapat memberikan gambaran yang objektif mengenai persepsi publik yang berkembang di media sosial.

3. HASIL DAN PEMBAHASAN

Setelah melalui serangkaian tahapan pengolahan data agar dapat digunakan dalam proses klasifikasi berbasis machine learning, data mentah terlebih dahulu diproses melalui pembersihan teks (*cleaning*), kemudian dilakukan pembagian data (*splitting*) dengan proporsi 80% sebagai data pelatihan (*training data*) dan 20% sebagai data pengujian (*testing data*), serta pelabelan data ke dalam tiga kategori sentimen utama: positif, negatif, dan netral. Proses ini bertujuan memastikan bahwa data yang digunakan sudah bersih, terstruktur, dan siap untuk dianalisis oleh model pembelajaran mesin.

Setelah data disiapkan, dilakukan implementasi tiga algoritma klasifikasi, yaitu *Support Vector Machine* (SVM), *Random Forest*, dan *Naive Bayes*, yang masing-masing memiliki pendekatan berbeda dalam memetakan pola dan memisahkan kelas sentimen. Evaluasi terhadap performa ketiga model dilakukan dengan menggunakan *confusion matrix*, yang berfungsi untuk menunjukkan jumlah prediksi yang benar dan salah untuk setiap kategori sentimen. *Confusion matrix* memberikan gambaran menyeluruh terkait seberapa baik model dalam mengenali kelas-kelas sentimen berdasarkan data pengujian.

Untuk mengevaluasi performa dari masing-masing model, digunakan *confusion matrix* yang dapat dilihat pada gambar 8,9 dan 10, yang memberikan gambaran kuantitatif mengenai prediksi benar dan salah pada setiap kategori sentimen. *Confusion matrix* juga memungkinkan peneliti untuk mengukur kinerja model secara lebih komprehensif melalui metrik-metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score*. Hasil dari evaluasi ini kemudian dijadikan dasar untuk membandingkan efektivitas masing-masing algoritma dalam mengklasifikasikan opini masyarakat terhadap Revisi Undang-Undang TNI. Dengan demikian, hasil dari pengolahan data dan penerapan model tidak hanya menunjukkan seberapa baik klasifikasi dilakukan, tetapi juga memberikan gambaran objektif tentang algoritma mana yang paling sesuai digunakan dalam konteks analisis sentimen berbasis teks. Untuk melihat perbandingan nilai akurasi, presisi, *recall* dan *F1-Score* dapat dilihat pada Tabel 7 dan Tabel 8 berikut ini.

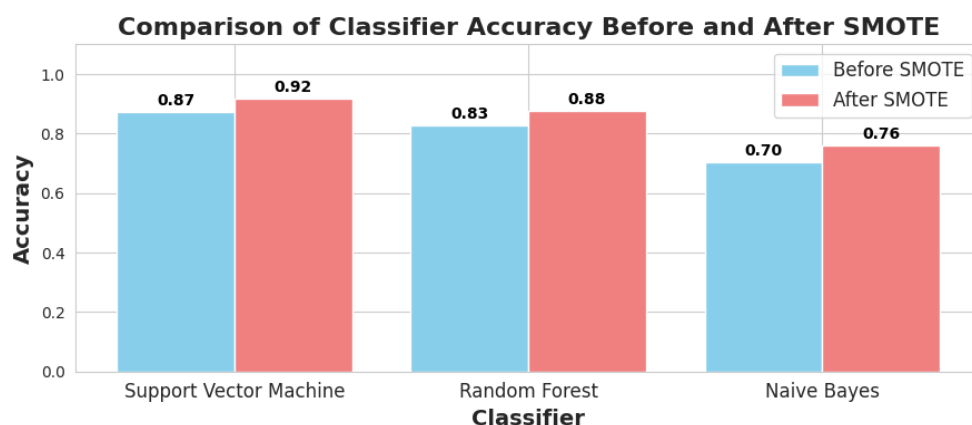
Tabel 7. Komparasi Hasil Evaluasi Model Algoritma Support Vector Machine, Random Forest, dan Naive Bayes Sebelum Dilakukan SMOTE

Model	Kategori	Akurasi	Presisi	Recall	F1-Score
Support Vector Machine	Negatif	87%	94%	85%	89%
	Netral		82%	91%	87%
	Positif		90%	84%	87%
Random Forest	Negatif	83%	82%	86%	84%
	Netral		71%	82%	83%
	Positif		82%	80%	81%
Naive Bayes	Negatif	70%	72%	78%	75%
	Netral		65%	81%	72%
	Positif		90%	39%	55%

Tabel 8. Komparasi Hasil Evaluasi Model Algoritma Support Vector Machine, Random Forest, dan Naive Bayes Setelah Dilakukan SMOTE

Model	Kategori	Akurasi	Presisi	Recall	F1-Score
Support Vector Machine	Negatif	92%	94%	91%	93%
	Netral		87%	89%	88%
	Positif		94%	95%	95%
Random Forest	Negatif	88%	87%	90%	88%
	Netral		87%	80%	83%
	Positif		89%	93%	91%
Naive Bayes	Negatif	76%	86%	77%	82%
	Netral		74%	84%	79%
	Positif		92%	89%	91%

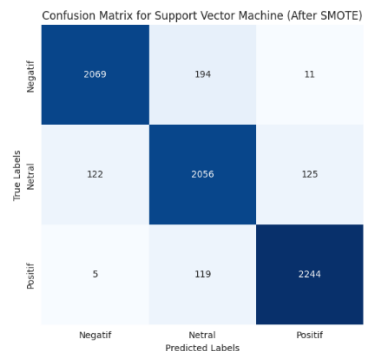
Pada Gambar 7 berikut ini menyajikan perbandingan visual mengenai kinerja akurasi dari tiga algoritma klasifikasi *Support Vector Machine* (SVM), *Random Forest*, dan *Naive Bayes*. Perbandingan ini dilakukan pada dua skenario, yaitu sebelum dan sesudah penerapan metode *oversampling* SMOTE (*Synthetic Minority Over-sampling Technique*) pada dataset.



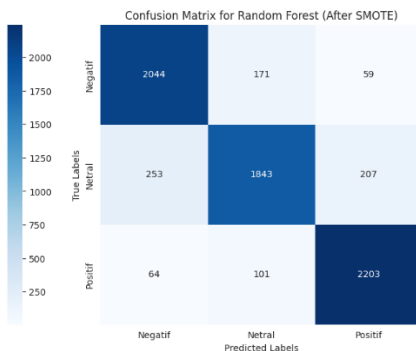
Gambar 7. Visualisasi Perbandingan Algoritma *Support Vector Machine*, *Random Forest*, dan *Naive Bayes* sebelum dan setelah dilakukan metode SMOTE.

Berdasarkan hasil evaluasi yang disajikan pada Tabel 7 dan Tabel 8, terlihat bahwa penerapan teknik SMOTE (*Synthetic Minority Over-sampling Technique*) memberikan dampak positif terhadap kinerja ketiga algoritma klasifikasi yang diuji, yaitu *Support Vector Machine* (SVM), *Random Forest*, dan *Naive Bayes*. Sebelum dilakukan SMOTE, algoritma SVM menunjukkan performa yang paling unggul dengan akurasi mencapai 87% dan nilai *F1-Score* tertinggi sebesar 89% pada kategori negatif, sementara algoritma *Naive Bayes* memiliki performa paling rendah dengan akurasi 70% dan *F1-Score* hanya 55% pada kategori positif. Ketidakseimbangan data yang belum ditangani tampak memengaruhi ketepatan klasifikasi, khususnya pada kelas minoritas. Setelah dilakukan SMOTE, seluruh algoritma mengalami peningkatan performa yang cukup signifikan. Algoritma SVM mencatat peningkatan akurasi menjadi 92% dengan *F1-Score* tertinggi mencapai 95%, mengindikasikan konsistensinya sebagai model terbaik.

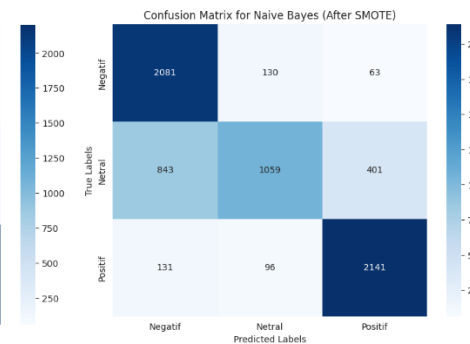
Random Forest juga mengalami peningkatan akurasi menjadi 88% dengan *F1-Score* yang merata pada semua kategori. Sementara itu, *Naive Bayes* yang sebelumnya menunjukkan performa terendah, turut mengalami peningkatan akurasi hingga 76% dan peningkatan nilai *F1-Score* di atas 80% pada semua kategori. Secara keseluruhan, hasil ini menunjukkan bahwa SMOTE efektif dalam mengatasi ketidakseimbangan kelas pada data latih, sehingga dapat meningkatkan kemampuan model dalam mengklasifikasikan sentimen secara lebih akurat dan seimbang.



Gambar 8. Confusion Matrix SVM Setelah SMOTE



Gambar 9. Confusion Matrix Random Forest Setelah SMOTE



Gambar 10. Confusion Matrix Naive Bayes Setelah SMOTE

Berdasarkan Gambar 8, Gambar 9, dan Gambar 10 yang menampilkan *confusion matrix* dari masing-masing algoritma setelah penerapan SMOTE, dapat disimpulkan bahwa algoritma Support Vector Machine (SVM) memiliki performa klasifikasi terbaik dibandingkan dengan *Random Forest* dan *Naive Bayes*. Pada Gambar 5, SVM menunjukkan jumlah prediksi yang paling akurat dengan nilai *True Positive* dan *True Negative* yang tinggi serta nilai *False Positive* dan *False Negative* yang rendah, terutama pada kategori positif dengan 2.244 prediksi benar dari total data yang ada. Sementara itu, pada Gambar 6, *Random Forest* menghasilkan prediksi yang juga cukup baik, namun masih terdapat kesalahan klasifikasi yang relatif lebih tinggi, seperti 253 data positif yang salah diklasifikasikan sebagai negatif. Di sisi lain, Gambar 7 menunjukkan bahwa *Naive Bayes* memiliki tingkat kesalahan prediksi yang paling tinggi, terlihat dari banyaknya data positif dan netral yang salah diklasifikasikan, misalnya 843 data netral yang diklasifikasikan sebagai negatif dan 401 data netral yang diklasifikasikan sebagai positif. Hal ini menunjukkan bahwa meskipun SMOTE berhasil memperbaiki distribusi data, efektivitas algoritma dalam mengklasifikasikan data tetap bergantung pada kompleksitas model, di mana SVM terbukti paling stabil dan andal dalam mengidentifikasi kelas sentimen secara tepat setelah proses penyeimbangan data dilakukan. Secara keseluruhan, SVM dinilai sebagai algoritma yang paling akurat, stabil, dan konsisten, sehingga paling direkomendasikan untuk digunakan dalam klasifikasi sentimen berbasis teks terhadap opini masyarakat di media sosial terkait RUU TNI.

4. KESIMPULAN

Penelitian ini berhasil mengevaluasi dan membandingkan kinerja algoritma Naive Bayes, Random Forest, dan Support Vector Machine (SVM) dalam menganalisis sentimen publik terhadap Revisi Undang-Undang TNI dari komentar Instagram. Setelah melalui tahapan pra-pemrosesan, transformasi TF-IDF, dan penyeimbangan data dengan teknik SMOTE, hasil evaluasi menunjukkan bahwa SVM mencapai performa terbaik dengan akurasi 92%, diikuti oleh Random Forest (88%) dan Naive Bayes (76%). Dengan demikian, SVM terbukti menjadi model yang paling efektif dan direkomendasikan untuk klasifikasi sentimen pada isu kebijakan publik yang kompleks di media sosial. Meskipun demikian, penelitian ini memiliki beberapa keterbatasan. Pertama, data hanya bersumber dari platform Instagram, sehingga belum mencerminkan opini dari platform media sosial lainnya. Kedua, analisis sentimen yang dilakukan belum mampu mendeteksi nuansa bahasa yang kompleks seperti sarkasme atau ironi secara mendalam. Keterbatasan ini dapat menjadi acuan untuk penelitian selanjutnya guna menghasilkan analisis opini publik yang lebih komprehensif. Penelitian ini diharapkan memberikan kontribusi dalam pemanfaatan machine learning untuk evaluasi kebijakan pertahanan nasional berbasis data digital.

REFERENCES

- [1] H. K. Muhid, “Fakta-fakta RUU TNI Disahkan Walau Tuai Penolakan,” TEMPO. Accessed: May 10, 2025. [Online]. Available: <https://www.tempo.co/politik/fakta-fakta-ruu-tni-disahkan-walau-tuai-penolakan--1222777>
- [2] A. T. DA, “DPR Akhirnya Setujui RUU TNI Jadi UU, Ini 3 Substansi yang Diatur!,” HukumOnline.Com. Accessed: May 10, 2025. [Online]. Available: <https://www.hukumonline.com/berita/a/dpr-akhirnya-setujui-ruu-tni-jadi-uu--ini-3-substansi-yang-diatur-lt67dbbddd68c6f/>
- [3] A. T. DA, “RUU TNI Berpeluang Tentara Masuk Sistem Penegakan Hukum,” Hukumonline.Com. Accessed: Aug. 11, 2025. [Online]. Available: <https://www.hukumonline.com/berita/a/ruu-tni-berpeluang-tentara-masuk-sistem-penegakan-hukum-lt67db7ce6cb026/?page=3>
- [4] S. Fatimah, “Transformasi Ruang Publik Digital : Tantangan Sosial dan Konstitusional dalam Demokrasi Era Media Baru,” Jurnal Litbang Kebijakan, vol. 19, no. 1, pp. 67–86, 2025, doi: 10.32781/cakrawala.v19i1.785



- [5] M. S. Alam, M. S. H. Mrida, and M. A. Rahman, “Sentiment Analysis in Social Media: How Data Science Impacts Public Opinion Knowledge Integrates Natural Language Processing (Nlp) With Artificial Intelligence (Ai),” *American Journal of Scholarly Research and Innovation*, vol. 4, no. 1, pp. 63–100, 2025, doi: 10.63125/r3sq6p80.
- [6] A. A. Nurhasananda and M. Akbar, “Analisis Sentimen Masyarakat Terhadap Kebijakan Undang-Undang Tentara Nasional Indonesia (UU TNI) Menggunakan Support Vector Machine,” *Jurnal Komputer, Informasi dan Teknologi*, vol. 5, no. 1, pp. 1–14, 2025.
- [7] A. Haris, K. Sandi, E. Haerani, L. Oktavia, and F. Kurnia, “Klasifikasi Sentimen Masyarakat Terhadap Revisi Undang-Undang Tentara Nasional Indonesia Menggunakan Naïve Bayes Classifier,” *Bulletincsr*, vol. 5, no. 4, pp. 594–603, 2025, doi: 10.47065/bulletincsr.v5i4.615.
- [8] H. Santoso, R. A. Putri, and S. Sahbandi, “Deteksi Komentar Cyberbullying pada Media Sosial Instagram Menggunakan Algoritma Random Forest,” *Jurnal Manajemen Informatika (JAMIKA)*, vol. 13, no. 1, pp. 62–72, 2023, doi: 10.34010/jamika.v13i1.9303.
- [9] M. I. Ghozali, W. H. Sugiharto, and A. F. Iskandar, “Analisis Sentimen Pinjaman Online Di Media Sosial Twitter Menggunakan Metode Naive Bayes,” *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 3, no. 6, pp. 1340–1348, 2023, doi: 10.30865/klik.v3i6.936.
- [10] I. H. Kusuma and N. Cahyono, “Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor,” *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 8, no. 3, pp. 302–307, 2023, doi: 10.30591/jpit.v8i3.5734.
- [11] A. Agung, A. Daniswara, I. Kadek, and D. Nuryana, “Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru,” *Journal of Informatics and Computer Science*, vol. 05, pp. 97–100, 2023.
- [12] A. Sentimen et al., “Analisis Sentimen Masyarakat Di Media Sosial X Terhadap Kemenkes Dengan Naive Bayes dan SVM,” *Jurnal Sains dan Teknologi*, vol. 7, no. 1, pp. 1–6, 2025.
- [13] F. N. Hidayat and S. Sugiyono, “Analisis Sentimen Masyarakat Terhadap Perekrutan Pppk Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine,” *Jurnal Sains dan Teknologi*, vol. 5, no. 2, pp. 665–672, 2023, doi: 10.55338/saintek.v5i2.1359.
- [14] L. Hermawati, V. Berland, A. Rahmadiyah, E. Hutabarat, and D. D. Saputra, “Komparasi Metode Text Mining Terhadap Masalah Pengklasifikasian Narasi Informatif & Non Informatif Pada twitter @ PLN _ 123,” *Jurnal Sistim Informasi dan Teknologi*, vol. 5, no. 1, pp. 109–120, 2023, doi: 10.37034/jsisfotek.v4i2.191.
- [15] R. Rinandyaswara, Y. A. Sari, and M. T. Furqon, “Pembentukan Daftar Stopword Menggunakan Term Based Random Sampling Pada Analisis Sentimen Dengan Metode Naïve Bayes (Studi Kasus: Kuliah Daring Di Masa Pandemi),” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 4, p. 717, 2022, doi: 10.25126/jtiik.2022934707.
- [16] T. Ridwansyah, “Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier,” *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 2, no. 5, pp. 178–185, 2022, doi: 10.30865/klik.v2i5.362.
- [17] S. F. N. Hadju and R. Jayadi, “Sentiment analysis of Indonesian e-commerce product reviews using support vector machine based term frequency inverse document frequency,” *J Theor Appl Inf Technol*, vol. 99, no. 17, pp. 4316–4325, 2021.
- [18] O. I. Gifari, Muh. Adha, F. Freddy, and F. F. S. Durrand, “Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine,” *Journal of Information Technology*, vol. 2, no. 1, pp. 36–40, 2022, doi: 10.46229/jifotech.v2i1.330.
- [19] A. Yulianti, Fadhilah Fitri, Nonong Amalita, and Dodi Vionanda, “The SMOTE Application of CART Methods for Coping Imbalanced Data in Classifying Status Work on Labor Force in the City of Padang,” *UNP Journal of Statistics and Data Science*, vol. 1, no. 3, pp. 172–179, 2023, doi: 10.24036/ujsds/vol1-iss3/12.
- [20] H. Eldo, A. Ayuliana, D. Suryadi, G. Chrisnawati, and L. Judijanto, “Penggunaan Algoritma Support Vector Machine (SVM) Untuk Deteksi Penipuan pada Transaksi Online,” *Jurnal Minfo Polgan*, vol. 13, no. 2, pp. 1627–1632, 2024, doi: 10.33395/jmp.v13i2.14186.
- [21] E. P. Febtiawan, L. A. S. I. Akbar, and A. S. Rachman, “Forecasting Produksi Energi Photovoltaic Menggunakan Algoritma Random Forest Classification,” *Journal of Information System Research (JOSH)*, vol. 5, no. 4, pp. 1053–1062, 2024, doi: 10.47065/josh.v5i4.5514.
- [22] F. Septianingrum and A. S. Y. Irawan, “Metode Seleksi Fitur Untuk Klasifikasi Sentimen Menggunakan Algoritma Naive Bayes: Sebuah Literature Review,” *Jurnal Media Informatika Budidarma*, vol. 5, no. 3, p. 799, 2021, doi: 10.30865/mib.v5i3.2983.