



Analisis Sentiment Terhadap Barak Militer pada Media Sosial X Menggunakan Metode Support Vector Machine (SVM)

Alfian Wicaksana, Maulida Ayu Fitriani*, Elindra Ambar Pambudi, Muhammad Hamka

Fakultas Teknik dan Sains, Prodi Teknik Informatika, Universitas Muhammadiyah Purwokerto, Purwokerto, Indonesia

Email: ¹alfiankamang33@gmail.com, ^{2,*}maulidaayuf@ump.ac.id, ³elindraambarpambudi@ump.ac.id,

⁴muhammadhamka@ump.ac.id,

Email Penulis Korespondensi: maulidaayuf@ump.ac.id

Abstrak—Pemerintah Provinsi Jawa Barat menginisiasi program barak militer dengan tujuan membangun karakter dan memulihkan jati diri remaja bermasalah melalui pendekatan semi-militer. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap program tersebut melalui media sosial X dengan menerapkan metode *Support Vector Machine* (SVM) yang dioptimalkan menggunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) untuk mengatasi ketidakseimbangan data. Data penelitian diperoleh melalui proses *crawling* dan menghasilkan 7.039 tweet setelah dilakukan penyaringan terhadap data duplikat dan penghapusan data netral. Tahapan *preprocessing* meliputi *cleaning*, *case folding*, *tokenizing*, *normalisasi*, *stopword removal*, dan *stemming*. Pelabelan sentimen dilakukan menggunakan pendekatan *lexicon-based* dan menghasilkan 6.596 tweet dengan sentimen positif serta 443 tweet dengan sentimen negatif. Model SVM dilatih menggunakan 80% data dan diuji menggunakan 20% data. Hasil evaluasi menunjukkan bahwa model memperoleh akurasi sebesar 96%, dengan *precision* negatif 45%, *precision* positif 98%, *recall* negatif 48%, *recall* positif 98%, *F1-score* negatif 46%, dan *F1-score* positif 98%. Visualisasi menggunakan *word cloud* memperkuat temuan bahwa opini publik dalam dataset cenderung positif, terutama terkait persepsi terhadap peningkatan disiplin dan pembentukan karakter remaja. Kontribusi penelitian ini terletak pada pengembangan analisis sentimen terhadap program barak militer melalui penerapan SVM yang dikombinasikan dengan SMOTE serta interpretasi pola opini publik secara kontekstual. Hasil penelitian tidak hanya memberikan gambaran kinerja klasifikasi, tetapi juga mengidentifikasi kecenderungan dukungan masyarakat terhadap aspek kedisiplinan dan pembentukan karakter serta kritik terhadap potensi stigmatisasi dan pendekatan pembinaan yang bersifat memaksa.

Kata Kunci: Program Kerja Gubernur Jawa Barat; Barak Militer Remaja; Pembentukan Karakter Remaja; Jati Diri Remaja; SVM

Abstract—The Government of West Java Province initiated a military barracks program aimed at building character and restoring the identity of troubled adolescents through a semi-military approach. This study aims to analyze public sentiment toward the program on social media X by applying the Support Vector Machine (SVM) method, optimized using the Synthetic Minority Over-sampling Technique (SMOTE) to address data imbalance. The data were collected through a crawling process, resulting in 7,039 tweets after duplicate filtering and removal of neutral data. The preprocessing stages included cleaning, case folding, tokenizing, normalization, stopwords removal, and stemming. Sentiment labeling was performed using a lexicon-based approach, resulting in 6,596 positive tweets and 443 negative tweets. The SVM model was trained using 80% of the data and tested using the remaining 20%. The evaluation results show that the model achieved an accuracy of 96%, with 45% negative precision, 98% positive precision, 48% negative recall, 98% positive recall, 46% negative F1-score, and 98% positive F1-score. Visualization using a word cloud further supports the finding that public opinion in the dataset tends to be positive, particularly regarding perceptions of improved discipline and adolescent character development. The contribution of this study lies in developing sentiment analysis of the military barracks program through the application of SVM combined with SMOTE and contextual interpretation of public opinion patterns. The findings provide not only an overview of classification performance but also identify public support for discipline and character development alongside criticism of potential stigmatization and coercive approaches to adolescent development.

Keywords: Jawa Barat Governor's Program; Military Barracks Program for Adolescents; Adolescent Character Development; Adolescent Identity; Support Vector Machine (SVM)

1. PENDAHULUAN

Masa remaja merupakan salah satu tahap perkembangan yang penting dalam kehidupan individu karena pada periode ini terjadi proses pembentukan identitas, moral, dan kepribadian. Pada tahap tersebut, remaja dihadapkan pada berbagai tugas perkembangan (*developmental tasks*) yang berkaitan dengan proses memahami dan menginternalisasi nilai, norma, serta pola perilaku yang akan menjadi dasar dalam menjalani kehidupan pada masa dewasa. Apabila tugas perkembangan tersebut tidak dapat dilalui dengan baik, remaja berpotensi mengalami berbagai permasalahan dalam perilaku, termasuk munculnya perilaku menyimpang. Sebaliknya, kemampuan remaja dalam menjalani proses perkembangan secara positif dapat mendukung terbentuknya individu yang produktif, bertanggung jawab, dan memiliki karakter yang baik [1].

Pemerintah Provinsi Jawa Barat menginisiasi program Barak Militer, yaitu pelatihan dengan pendekatan semi-militer dan berfokus pada pembentukan disiplin bagi remaja yang terlibat dalam perilaku menyimpang. Program ini bertujuan untuk membentuk karakter, menanamkan kedisiplinan, serta membantu remaja dalam proses pencarian dan pemulihan jati diri. Kebijakan ini memunculkan kekhawatiran terkait kemungkinan munculnya tindakan represif yang bertentangan dengan prinsip-prinsip pendidikan yang bersifat humanistic. Pendekatan yang dinilai terlalu militeristik ini dianggap tidak menyentuh inti persoalan psikososial yang dialami remaja dan bahkan berpotensi menimbulkan dampak traumatis[2].

Media sosial X (yang sebelumnya dikenal sebagai Twitter) telah berkembang menjadi salah satu ruang diskusi publik yang memungkinkan masyarakat menyampaikan berbagai bentuk respons, baik berupa dukungan, kritik, maupun aspirasi, terhadap kebijakan pemerintah secara langsung dan cepat [3]. Media sosial memiliki peran strategis dalam berbagai tahapan proses kebijakan publik, mulai dari perumusan dan pelaksanaan hingga evaluasi kebijakan. Berbagai



opini dan respons masyarakat yang muncul di media sosial dapat dimanfaatkan sebagai gambaran awal mengenai tingkat penerimaan publik (*public acceptance*) terhadap suatu kebijakan, sebelum pemerintah melaksanakan proses evaluasi secara formal.

Aktualitas kebijakan barak militer juga telah mendorong munculnya sejumlah penelitian analisis sentimen yang secara khusus mengkaji respons masyarakat terhadap kebijakan tersebut. Meskipun demikian, penelitian-penelitian tersebut memiliki perbedaan dalam hal cakupan data, sumber data, maupun metode klasifikasi yang digunakan. Delayana dkk., misalnya, mengkaji opini publik mengenai kebijakan barak militer dengan menerapkan algoritma Naive Bayes melalui platform RapidMiner, menggunakan data yang bersumber dari media sosial X. Pemilihan topik tersebut didasarkan pada pertimbangan bahwa opini publik terhadap kebijakan barak militer dapat menjadi salah satu gambaran awal mengenai tingkat penerimaan masyarakat terhadap konsep edukasi yang ditawarkan oleh pemerintah[4]. Sementara itu, Syalsabina dkk. melakukan analisis sentimen terhadap respons masyarakat mengenai kebijakan yang sama dengan memanfaatkan data komentar pada platform YouTube dan algoritma XGBoost. Hasil penelitian tersebut menunjukkan tingkat akurasi rata-rata sebesar 83%, dengan komposisi sentimen berupa 44,7% positif, 12,1% negatif, dan 43,2% netral[5].

Berdasarkan penelitian terdahulu, analisis sentimen terhadap kebijakan barak militer telah dilakukan dengan menggunakan berbagai sumber data dan algoritma klasifikasi. Delayana dkk. menggunakan *Naive Bayes* pada data media sosial X untuk menganalisis opini publik terhadap kebijakan barak militer[4]. Syalsabina dkk. menggunakan *XGBoost* terhadap komentar YouTube dan memperoleh akurasi rata-rata sebesar 83%, dengan distribusi sentimen sebesar 44,7% positif, 12,1% negatif, dan 43,2% netral[5]. Sementara itu, Tantri menerapkan pendekatan *Natural Language Processing* (NLP) untuk mengidentifikasi kecenderungan sentimen pengguna Twitter terhadap kebijakan Gubernur Jawa Barat terkait barak militer[6]. Penelitian lain oleh Setiawan menggunakan algoritma *Long Short-Term Memory* (LSTM) untuk menganalisis sentimen masyarakat terhadap kebijakan Gubernur Jawa Barat secara lebih umum dan menunjukkan bahwa hasil klasifikasi dapat dimanfaatkan sebagai pendukung pemantauan opini publik serta *data-driven policy making*. Perbedaan sumber data, cakupan objek, dan algoritma tersebut menunjukkan bahwa respons masyarakat terhadap kebijakan barak militer telah dikaji melalui berbagai pendekatan komputasional[7].

Adapun penelitian yang memiliki keterkaitan paling dekat dengan penelitian ini dilakukan oleh Mansyur dkk. yang menerapkan *Support Vector Machine* (SVM) untuk mengklasifikasikan sentimen terhadap program barak militer menggunakan data Twitter. Penelitian tersebut menggunakan 1.826 tweet hasil *crawling* yang kemudian disaring menjadi 1.000 tweet, dengan pelabelan sentimen menggunakan model BERT yang selanjutnya divalidasi melalui anotasi manual terhadap 200 sampel data[8]. Selain itu, Fitrah dan Anggreini menerapkan SVM untuk menganalisis opini warganet mengenai pendidikan militer bagi pelajar dan memperoleh *precision* sebesar 0,98 pada kelas positif, meskipun masih ditemukan kesalahan klasifikasi pada sebagian data[9]. Temuan tersebut menunjukkan bahwa SVM memiliki kemampuan yang cukup baik dalam melakukan klasifikasi sentimen, tetapi performanya tetap dipengaruhi oleh karakteristik dan distribusi dataset yang digunakan.

Berdasarkan pembahasan pada penelitian sebelumnya, dapat diidentifikasi tiga kesenjangan utama yang menjadi dasar pentingnya penelitian ini dilakukan. Pertama, dari aspek substansi, kajian mengenai pembentukan karakter remaja selama ini lebih banyak berfokus pada lingkungan pendidikan formal melalui penerapan pembiasaan, keteladanan, dan pendidikan karakter. Sementara itu, penelitian yang secara khusus mengkaji respons dan penerimaan masyarakat terhadap pendekatan pembentukan karakter yang bercorak disipliner melalui kebijakan barak militer masih relatif terbatas. Penelitian terdahulu mengenai barak militer juga lebih banyak menitikberatkan pada klasifikasi sentimen atau pembahasan kebijakan secara umum, sehingga keterkaitan antara kecenderungan opini masyarakat dengan isu pembentukan karakter remaja belum banyak dikaji secara lebih mendalam.

Kedua, dari aspek metodologis, penelitian analisis sentimen mengenai kebijakan barak militer telah menggunakan berbagai algoritma, seperti *Naive Bayes*, *XGBoost*, *Long Short-Term Memory* (LSTM), dan *Support Vector Machine* (SVM), dengan karakteristik dataset dan hasil klasifikasi yang berbeda-beda. Meskipun SVM telah digunakan dalam penelitian terdahulu untuk menganalisis sentimen terhadap program barak militer, penelitian tersebut menggunakan jumlah data yang relatif terbatas dibandingkan dengan dataset yang digunakan dalam penelitian ini. Selain itu, karakteristik distribusi sentimen yang tidak seimbang juga menjadi persoalan yang perlu diperhatikan karena dapat memengaruhi kemampuan model dalam mengenali kelas minoritas. Dengan demikian, masih diperlukan pengujian penerapan SVM pada dataset media sosial X yang lebih aktual dan lebih luas dengan memperhatikan ketidakseimbangan kelas melalui tahapan pengolahan data yang sistematis.

Ketiga, dari aspek integrasi kajian, penelitian yang menghubungkan pembentukan karakter remaja, kebijakan barak militer, dan analisis sentimen berbasis *machine learning* pada media sosial X masih relatif terbatas. Sebagian besar penelitian terdahulu cenderung mengkaji aspek-aspek tersebut secara terpisah, yaitu analisis sentimen berfokus pada performa klasifikasi, sedangkan kajian mengenai pembentukan karakter lebih banyak membahas aspek pendidikan atau pendekatan pembinaan. Kondisi tersebut menunjukkan adanya ruang penelitian untuk menginterpretasikan kecenderungan sentimen masyarakat tidak hanya sebagai hasil klasifikasi komputasional, tetapi juga sebagai representasi respons publik terhadap pendekatan pembentukan karakter remaja melalui kebijakan barak militer.

Berdasarkan ketiga kesenjangan tersebut, penelitian ini diarahkan untuk memperluas kajian analisis sentimen terhadap kebijakan barak militer melalui penerapan *Support Vector Machine* (SVM) pada data media sosial X yang lebih aktual dan lebih luas. Penelitian ini juga memperhatikan permasalahan ketidakseimbangan kelas melalui penerapan *Synthetic Minority Over-sampling Technique* (SMOTE) sebelum proses klasifikasi. Dengan demikian, penelitian tidak



hanya berorientasi pada pencapaian nilai akurasi model, tetapi juga menginterpretasikan kecenderungan sentimen positif dan negatif sebagai gambaran respons masyarakat terhadap kebijakan barak militer dalam konteks pembentukan karakter remaja.

Penelitian ini memberikan kontribusi pada dua aspek, yaitu metodologis dan substantif. Secara metodologis, penelitian ini memperluas penerapan Support Vector Machine (SVM) dalam analisis sentimen terhadap kebijakan barak militer dengan memanfaatkan dataset media sosial X yang lebih aktual dan lebih luas, disertai tahapan *preprocessing* serta penanganan ketidakseimbangan kelas menggunakan Synthetic Minority Over-sampling Technique (SMOTE). Pendekatan tersebut diharapkan dapat memberikan gambaran yang lebih terukur mengenai kemampuan SVM dalam mengklasifikasikan opini masyarakat terhadap isu kebijakan yang bersifat aktual dan kontroversial. Secara substantif, penelitian ini tidak hanya menghasilkan ukuran kinerja model berupa akurasi, *precision*, *recall*, dan *F1-score*, tetapi juga menginterpretasikan kecenderungan sentimen positif dan negatif sebagai representasi respons masyarakat terhadap kebijakan barak militer, khususnya dalam kaitannya dengan pembentukan karakter remaja. Dengan demikian, penelitian ini diharapkan dapat memperkaya kajian analisis sentimen terhadap kebijakan publik sekaligus menyediakan informasi berbasis data mengenai kecenderungan penerimaan dan penolakan masyarakat yang dapat menjadi bahan pertimbangan dalam evaluasi kebijakan.

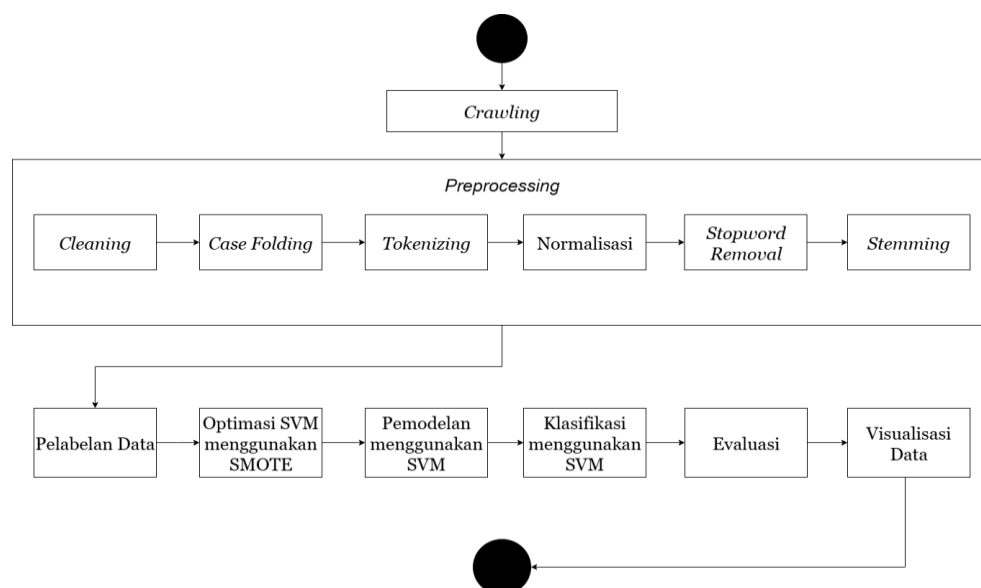
Kelemahan SVM mencakup kesulitan dalam penerapan pada dataset yang memiliki jumlah sampel besar dan dimensi tinggi, serta menghadapi masalah ketika dataset tidak seimbang. SVM dirancang untuk menyelesaikan masalah klasifikasi dua kelas. Teknik *Synthetic Minority Over-sampling Technique* (SMOTE) dapat diterapkan untuk menyeimbangkan dataset dengan cara membuat instance baru dari minority class. Alasan pemilihan metode SVM untuk program Barak Militer pada media sosial X adalah karena keunggulan yang dimilikinya serta popularitasnya sebagai teknik dalam klasifikasi teks. SVM diterapkan untuk prediksi, baik dalam kasus klasifikasi maupun regresi, dengan prinsip dasar menentukan hyperplane terbaik untuk memaksimalkan jarak antara dua kelas[10].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Kajian analisis sentimen dalam penelitian ini dilakukan melalui beberapa tahapan yang disusun secara sistematis. Tahap pertama adalah pengumpulan data opini masyarakat dari media sosial X menggunakan teknik *crawling*. Data yang diperoleh selanjutnya melalui tahap *preprocessing* untuk mengurangi *noise* dan menghasilkan representasi teks yang lebih terstruktur sebelum digunakan dalam proses analisis. Tahapan *preprocessing* meliputi *cleaning*, *case folding*, *tokenizing*, normalisasi, *stopword removal*, dan *stemming*. Setelah proses tersebut, data diberikan label sentimen menggunakan pendekatan *Lexicon-Based* untuk menentukan kategori sentimen yang menjadi dasar dalam pembentukan data untuk proses klasifikasi.

Tahap selanjutnya adalah menangani ketidakseimbangan distribusi kelas menggunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE). Data yang telah diproses kemudian dibagi menjadi data latih dan data uji untuk membangun model klasifikasi menggunakan *Support Vector Machine* (SVM). Model yang dihasilkan dievaluasi menggunakan *confusion matrix* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengukur kemampuan model dalam melakukan klasifikasi sentimen. Tahap akhir dilakukan melalui visualisasi data untuk menggambarkan distribusi serta pola kecenderungan sentimen masyarakat terhadap program barak militer. Keseluruhan tahapan penelitian tersebut disajikan secara sistematis pada Gambar 1.



Gambar 1. Alur Penelitian



2.2 Pengumpulan Data

Strategi crawling yang memanfaatkan Python dan Google Colab digunakan untuk mengekstrak data opini dari media sosial X [11]. Pengumpulan data dilakukan dengan menggunakan beberapa kata kunci yang relevan dengan fokus penelitian, meliputi *Program Barak Militer Remaja*, *Pembentukan Karakter*, *Program Kerja Gubernur Jawa Barat*, *Program Dedi Mulyadi*, dan *Jati Diri Remaja*. Pemilihan kata kunci tersebut dimaksudkan untuk memperoleh data yang merepresentasikan berbagai bentuk respons masyarakat terhadap program barak militer, baik yang secara langsung membahas program maupun yang mengaitkannya dengan isu pembentukan karakter dan perilaku remaja. Data hasil *crawling* selanjutnya digunakan sebagai dataset awal yang akan melalui tahapan preprocessing sebelum dilakukan pelabelan dan klasifikasi sentimen. Preprocessing

Tahap preprocessing merupakan proses pembersihan dan penyempurnaan data, karena data yang diperoleh umumnya bersifat tidak terstruktur dan mengandung banyak elemen yang tidak relevan. Tujuan utama dari tahap ini adalah untuk mengurangi noise dalam data [12]. Salah satu langkah dalam preprocessing adalah:

- Cleaning merupakan tahap pembersihan teks dengan menghapus berbagai elemen yang tidak diperlukan dalam proses analisis. Elemen yang dihilangkan meliputi tanda baca, username, URL, mention, hashtag, retweet, simbol, serta angka yang tidak memiliki relevansi terhadap analisis sentimen
- Case Folding merupakan tahap normalisasi teks dengan mengubah seluruh huruf menjadi huruf kecil (lowercase). Tahap ini bertujuan menyeragamkan bentuk penulisan kata sehingga kata yang memiliki makna sama tetapi ditulis dengan kapitalisasi berbeda dapat diperlakukan sebagai token yang sama.
- Tokenizing merupakan tahap pemecahan teks atau kalimat menjadi unit-unit kata yang disebut token. Proses ini dilakukan untuk memisahkan setiap kata dalam teks sehingga dapat diproses dan dianalisis pada tahapan berikutnya.
- Normalisasi merupakan proses mengubah kata tidak baku, kata tidak valid, singkatan, maupun istilah informal yang umum digunakan pada media sosial menjadi bentuk kata yang baku. Tahap ini diperlukan untuk mengurangi variasi penulisan dan meningkatkan kesesuaian kata dengan kamus yang digunakan dalam proses pelabelan sentimen.
- Stopword Removal merupakan tahap menghilangkan kata-kata yang memiliki frekuensi kemunculan tinggi tetapi dianggap tidak memberikan kontribusi signifikan terhadap informasi atau makna sentimen dalam teks. Pada penelitian ini, proses tersebut dilakukan menggunakan Stopwords dari Natural Language Toolkit (NLTK) yang mendukung daftar kata umum dalam bahasa Indonesia.
- Stemming merupakan tahap menghilangkan imbuhan pada kata untuk mengembalikannya ke bentuk dasar (root word). Proses ini mencakup penghilangan awalan, akhiran, maupun kombinasi imbuhan sehingga kata-kata yang memiliki bentuk turunan dapat direpresentasikan dalam bentuk dasar yang sama [13].

2.3 Pelabelan Data

Pelabelan data dalam penelitian ini dilakukan untuk membentuk ground truth yang kemudian digunakan sebagai dasar dalam penyusunan data latih (training data) dan data uji pada tahap klasifikasi menggunakan Machine Learning. Penting untuk dipahami bahwa proses pelabelan (labeling) dan klasifikasi (classification) merupakan dua tahapan yang berbeda serta memiliki fungsi masing-masing. Pada tahap pelabelan, pendekatan Lexicon-Based digunakan untuk memberikan label sentimen awal terhadap data mentah yang belum memiliki kategori sentimen. Selanjutnya, pada tahap klasifikasi, algoritma Support Vector Machine (SVM) mempelajari hubungan antara representasi fitur teks dan label yang telah terbentuk untuk menemukan pola yang dapat digeneralisasikan pada data baru yang belum pernah digunakan sebelumnya. Oleh karena itu, penerapan pendekatan Lexicon-Based dalam proses pelabelan tidak dapat dianggap sebagai pengulangan atau duplikasi terhadap metode klasifikasi Machine Learning. Pendekatan tersebut merupakan tahap awal (pre-labeling) yang diperlukan dalam skema supervised learning, khususnya ketika dataset yang tersedia belum dilengkapi dengan label sentimen [14].

Proses pelabelan data dalam penelitian ini menggunakan pendekatan Lexicon-Based dengan memanfaatkan InSet Lexicon, yaitu kamus sentimen berbahasa Indonesia yang dikembangkan oleh Koto dan Rahmanningtyas dan telah digunakan dalam berbagai penelitian analisis sentimen pada teks berbahasa Indonesia. InSet Lexicon memuat 3.609 kata dengan polaritas positif dan 6.609 kata dengan polaritas negatif [15]. Setiap kata dalam kamus tersebut memiliki bobot polaritas pada rentang -5 hingga +5 yang diperoleh melalui proses anotasi dan pengujian oleh pengembangnya. Mengingat data penelitian berupa tweet yang cenderung menggunakan bahasa informal, dilakukan proses adaptasi kamus (domain adaptation) agar lexicon dapat menyesuaikan dengan karakteristik data yang dianalisis. Proses adaptasi dilakukan melalui beberapa tahapan. Pertama, kata-kata tidak baku atau istilah alay dinormalisasi menggunakan kamus normalisasi kata sebelum dilakukan proses pencocokan dengan skor lexicon. Kedua, dilakukan penambahan entri berupa kata maupun frasa yang relevan dengan konteks penelitian, khususnya istilah yang berkaitan dengan disiplin, kekerasan, dan pendidikan karakter yang belum tersedia dalam InSet Lexicon. Ketiga, seluruh entri yang ditambahkan melalui proses perluasan kamus ditinjau secara manual oleh peneliti untuk memastikan bahwa bobot polaritas yang diberikan sesuai dengan konteks penggunaannya dalam domain penelitian. Tahapan tersebut dilakukan untuk mengurangi kemungkinan ketidaktepatan pemberian skor yang dapat muncul akibat penggunaan kamus sentimen yang bersifat umum pada data dengan karakteristik dan konteks yang lebih spesifik.

$$S_{\{positive\}} = \sum_{i \in \{positive\}} score_i \quad (1)$$

$$S_{\{negative\}} = \sum_{i \in \{negative\}} score_i \quad (2)$$



(*Spositive*) merupakan ukuran kalimat yang diperoleh dari total nilai polaritas kata opini positive, sedangkan (*Snegative*) merupakan ukuran kalimat yang didapat dari nilai polaritas kata opini negatif. Rumus ini dipakai untuk mengetahui arah sentimen dengan membandingkan nilai positive, negative, dan netral [15].

$$\begin{aligned} \text{Sentence}_{\{\text{sentiment}\}} \begin{cases} \text{positive if } S_{\{\text{positive}\}} > S_{\{\text{negative}\}} \\ \text{neutral if } S_{\{\text{positive}\}} = S_{\{\text{negative}\}} \\ \text{negative if } S_{\{\text{positive}\}} < S_{\{\text{negative}\}} \end{cases} \end{aligned} \quad (3)$$

Teks akan dilabeli sentimen positif ketika kata positif kuantitasnya lebih banyak daripada kata negatif. Sebaliknya, ketika kata negatif kuantitasnya lebih banyak daripada kata positif, teks tersebut akan dilabeli sentimen negatif. Teks tersebut diberi label sentimen netral jika jumlah kata positif dan negatif sama. Karena pendekatan Lexicon-Based bersifat generik dan memiliki keterbatasan dalam menangani struktur kalimat kompleks, seperti sarkasme dan negasi berganda, label yang dihasilkan tidak langsung ditetapkan sebagai ground truth, tetapi divalidasi secara manual sebelum digunakan dalam klasifikasi Machine Learning.

Validasi dilakukan oleh tiga anotator secara independen terhadap sampel data yang dipilih secara acak terstratifikasi tanpa mengetahui label dari Lexicon-Based untuk meminimalkan bias. Tingkat kesepakatan anotator diukur menggunakan Fleiss' Kappa karena melibatkan lebih dari dua anotator. Berdasarkan kriteria Landis dan Koch[16], nilai kappa 0,61–0,80 menunjukkan kesepakatan substansial, sedangkan 0,81–1,00 menunjukkan kesepakatan hampir sempurna. Perbedaan label diselesaikan melalui adjudication atau majority voting, sehingga ground truth yang digunakan dalam pelatihan model dapat lebih andal [17].

2.4 Optimasi SVM menggunakan SMOTE

Ketidakseimbangan jumlah data antara kelas opini positif dan negatif berpotensi menimbulkan bias pada model klasifikasi, khususnya terhadap kelas yang memiliki jumlah sampel lebih dominan. Oleh karena itu, digunakan metode Synthetic Minority Over-sampling Technique (SMOTE) untuk menangani permasalahan ketidakseimbangan tersebut. Metode SMOTE dilakukan dengan menghasilkan sampel sintesis baru pada kelas minoritas melalui pendekatan Nearest Neighbors, sehingga proporsi data pada kelas minoritas dan mayoritas menjadi lebih seimbang sebelum data digunakan dalam proses pemodelan klasifikasi. Ketidakseimbangan jumlah data antara kelas opini positif dan negatif berpotensi menimbulkan bias pada model klasifikasi, khususnya terhadap kelas yang memiliki jumlah sampel lebih sedikit. Kondisi tersebut dapat menyebabkan model cenderung mempelajari pola dari kelas mayoritas dan menghasilkan kemampuan prediksi yang kurang optimal terhadap kelas minoritas. Oleh karena itu, penelitian ini menerapkan Synthetic Minority Over-sampling Technique (SMOTE) sebagai upaya untuk mengatasi ketidakseimbangan distribusi kelas sebelum proses pemodelan menggunakan *Support Vector Machine* (SVM).

SMOTE bekerja dengan membentuk sampel sintesis baru pada kelas minoritas berdasarkan kedekatan antar-*instance* menggunakan pendekatan *Nearest Neighbors*. Sampel sintesis tersebut dihasilkan melalui interpolasi antara data minoritas yang berdekatan sehingga tidak sekadar melakukan penggandaan data yang telah tersedia. Dalam penelitian ini, penerapan SMOTE dilakukan setelah proses pembagian data dan hanya diterapkan pada data latih untuk menghindari terjadinya *data leakage* terhadap data pengujian. Hasil penerapan SMOTE selanjutnya digunakan sebagai data pelatihan yang lebih seimbang dalam proses pembentukan model SVM. Dengan demikian, model diharapkan memiliki kesempatan yang lebih proporsional untuk mempelajari karakteristik kedua kelas sentimen dan meningkatkan kemampuan dalam mengenali kelas minoritas.[18].

2.5 Pemodelan Menggunakan SVM

Tahap selanjutnya adalah pemodelan klasifikasi menggunakan algoritma Support Vector Machine (SVM). Klasifikasi bertujuan untuk membangun model yang dapat mempelajari pola hubungan antara fitur teks dan label sentimen, kemudian menggunakan pola tersebut untuk memprediksi kelas pada data yang belum diketahui labelnya. Sebelum proses pemodelan, dataset dibagi menjadi dua kelompok, yaitu data latih (*training data*) dan data uji (*testing data*). Data latih digunakan untuk membangun dan mempelajari pola klasifikasi, sedangkan data uji digunakan untuk mengevaluasi kemampuan model dalam mengklasifikasikan data yang tidak digunakan selama proses pelatihan. Pembagian dataset dilakukan dengan rasio 80:20, yaitu 80% data digunakan sebagai data latih dan 20% sebagai data uji.

Pemodelan SVM dalam penelitian ini menggunakan kernel linear (*linear kernel*) yang sesuai untuk klasifikasi teks dengan representasi fitur berdimensi tinggi. Kernel linear digunakan untuk menentukan batas pemisah (*decision boundary*) yang optimal antara kelas sentimen positif dan negatif berdasarkan fitur yang telah dibentuk dari data teks. Implementasi model dilakukan menggunakan *library* Scikit-learn, sedangkan *library* Natural Language Toolkit (NLTK) digunakan pada tahap *preprocessing*, khususnya dalam proses *stopword removal*. Pengaturan tersebut memungkinkan proses pengolahan teks dan pembentukan model klasifikasi dilakukan secara terpisah sesuai dengan fungsi masing-masing *library* [19].

2.6 Klasifikasi Menggunakan SVM

Support Vector Machine (SVM) merupakan algoritma klasifikasi *supervised learning* yang bekerja dengan menentukan *hyperplane* optimal untuk memisahkan data ke dalam kelas yang berbeda. Hyperplane dipilih dengan mempertimbangkan jarak maksimum (*margin*) antara batas pemisah dan data terdekat dari masing-masing kelas yang disebut *support vectors*.



Dengan memaksimalkan *margin*, SVM berupaya memperoleh batas keputusan yang mampu memberikan pemisahan kelas secara optimal serta memiliki kemampuan generalisasi terhadap data yang belum pernah digunakan dalam proses pelatihan [20]. Hyperplane sendiri ialah fungsi yang digunakan untuk membedakan berbagai kelas. Pada klasifikasi linier menggunakan SVM, hyperplane biasanya dinyatakan dalam bentuk persamaan 4

$$f(x) = w^T x + b \quad (4)$$

Keterangan: f = nilai target, x = vector input, w = vector bobot, b = bias atau intercept.

2.7 Evaluasi Model

Evaluasi dilakukan untuk mengukur performa model dalam mengklasifikasikan sentimen berdasarkan tingkat ketepatan prediksi terhadap data aktual. Penelitian ini menggunakan *confusion matrix* sebagai metode evaluasi untuk mengidentifikasi jumlah data yang diklasifikasikan secara benar maupun salah pada setiap kelas sentimen. Berdasarkan nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) yang dihasilkan, selanjutnya dilakukan perhitungan *accuracy*, *precision*, *recall*, dan *F1-score*. Rumus perhitungan masing-masing metrik evaluasi ditunjukkan pada Persamaan (5) sampai dengan Persamaan (8).

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1 - Score = \frac{2(Precision \times Recall)}{Precision + Recall} \quad (8)$$

Komponen True Positive (TP) menunjukkan jumlah data yang secara aktual termasuk dalam kelas positif dan berhasil diklasifikasikan oleh model sebagai kelas positif. False Positive (FP) menunjukkan jumlah data yang secara aktual termasuk dalam kelas negatif, tetapi diklasifikasikan oleh model sebagai kelas positif. False Negative (FN) menunjukkan jumlah data yang secara aktual termasuk dalam kelas positif, tetapi diklasifikasikan oleh model sebagai kelas negatif. Sementara itu, True Negative (TN) menunjukkan jumlah data yang secara aktual termasuk dalam kelas negatif dan berhasil diklasifikasikan oleh model sebagai kelas negatif.

2.8 Visualisasi Data

Visualisasi data dilakukan terhadap teks opini yang telah melalui tahapan *preprocessing* dan pelabelan sentimen. Sebelum divisualisasikan, data terlebih dahulu difilter melalui proses *stopword removal* untuk menghilangkan kata-kata yang memiliki frekuensi kemunculan tinggi tetapi relatif rendah dalam memberikan informasi substantif terhadap interpretasi opini. Dengan demikian, kata-kata yang dipertahankan diharapkan lebih merepresentasikan topik, karakteristik, dan kecenderungan sentimen yang terdapat dalam data penelitian.

Selanjutnya, visualisasi dilakukan menggunakan *Word Cloud* untuk mengidentifikasi kosakata yang dominan pada masing-masing kelompok sentimen. Frekuensi kemunculan suatu kata direpresentasikan melalui ukuran teks, sehingga kata dengan frekuensi yang lebih tinggi ditampilkan dalam ukuran yang lebih besar. Hasil visualisasi tidak digunakan sebagai dasar tunggal dalam menentukan polaritas sentimen, melainkan sebagai instrumen pendukung untuk menginterpretasikan pola opini yang telah diperoleh melalui proses klasifikasi. Melalui visualisasi tersebut, kata atau istilah yang dominan pada sentimen positif dan negatif dapat diidentifikasi sehingga memberikan gambaran yang lebih kontekstual mengenai isu, persepsi, dan kecenderungan masyarakat terhadap program barak militer [21].

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data penelitian berupa opini masyarakat yang diperoleh dari media sosial X melalui teknik *crawling* menggunakan bahasa pemrograman Python pada platform Google Colab. Proses pengumpulan data dilakukan berdasarkan sejumlah kata kunci yang disesuaikan dengan fokus penelitian, yaitu program barak militer, pembentukan karakter remaja, serta kebijakan Gubernur Jawa Barat. Media sosial X dipilih sebagai sumber data karena memungkinkan pengguna menyampaikan opini, tanggapan, dan persepsi secara terbuka terhadap isu kebijakan yang sedang berkembang. Penggunaan data dari platform tersebut memungkinkan penelitian menangkap respons publik secara langsung dalam bentuk teks yang dapat dianalisis secara komputasional. Berdasarkan proses *crawling*, diperoleh sebanyak 7.092 data awal yang selanjutnya menjadi bahan dalam tahapan pengolahan data.

Data hasil *crawling* selanjutnya dilakukan penyaringan untuk menjaga kualitas dan konsistensi dataset. Penyaringan dilakukan dengan mengidentifikasi dan menghapus data duplikat agar setiap opini tidak direpresentasikan lebih dari satu kali dan tidak menyebabkan pengulangan informasi dalam proses analisis. Setelah proses penyaringan, diperoleh sebanyak 7.063 data yang ditetapkan sebagai dataset penelitian. Dataset tersebut kemudian digunakan pada tahap *preprocessing* untuk mengurangi *noise* dan mengatasi variasi penulisan yang umum ditemukan pada data media



sosial, sehingga teks dapat direpresentasikan secara lebih terstruktur sebelum memasuki tahap pelabelan dan klasifikasi sentimen. Sebagian data hasil pengumpulan dan penyaringan ditampilkan pada Tabel 1.

Tabel 1. Hasil Pengumpulan Data

No	Hasil Pengumpulan Data
1	Antara Barak Militer dan Pesantren: Bagaimana Cara Efektif Mengatasi Kenakalan Remaja? Menurut Pengamat Pendidikan Unika Soegijapranata Semarang Tukiman Taruna idiomnya jgn diberikan cap anak nakal. Tentu mencari akar masalah dari keluarga tapi kita tdk menyalahkan keluarga
2	@triawdd19 @mardigu024 @mardigu024 Jika program barak militer berhasil manfaat bisa dirasakan oleh: 1. Remaja bermasalah dengan disiplin dan prospek hidup lebih baik. 2. Masyarakat karena potensi penurunan kriminalitas. 3. Keluarga dengan perubahan posisi
...	...
7063	Melihat kenakalan remaja yg sering melakukan tawuran dijalanan bahkan mereka dijadikan 'begal' yg diperalat oleh miras dan narkoba sampai mengganggu keamanan publik gagasan membawa mrk ke barak militer oleh gub Jabar KDM ckp cemerlang demi memperbaiki generasi muda kedepan

3.2 Preprocessing

Data opini yang diperoleh dari media sosial X memiliki karakteristik berupa penggunaan bahasa informal, singkatan, kesalahan penulisan, istilah slang, penggunaan tanda baca, serta variasi bentuk kata. Karakteristik tersebut dapat menimbulkan *noise* dan meningkatkan keragaman fitur apabila data digunakan secara langsung dalam proses klasifikasi. Oleh karena itu, diperlukan tahap *preprocessing* untuk membersihkan dan menyeragamkan data teks sehingga diperoleh representasi yang lebih terstruktur serta sesuai untuk digunakan pada tahap analisis sentimen.

Pemilihan tahapan *preprocessing* dalam penelitian ini disesuaikan dengan karakteristik data dan kebutuhan analisis. Putra dan Fitriani dalam penelitian mengenai analisis sentimen Program Kampus Mengajar menggunakan algoritma *Support Vector Machine* (SVM) menerapkan tahapan *cleaning*, *case folding*, *tokenizing*, *normalisasi*, *filtering*, dan *stemming* sebelum proses klasifikasi. Meskipun memiliki alur pengolahan teks yang serupa, penelitian ini tidak menerapkan seluruh tahapan tersebut secara identik. Perbedaan karakteristik opini mengenai program barak militer, khususnya penggunaan bahasa informal, singkatan, istilah slang, dan variasi penulisan pada media sosial X, menjadi pertimbangan dalam melakukan penyesuaian tahapan *preprocessing*. Penelitian ini menambahkan *stopword removal* dan menggunakan beberapa kamus normalisasi secara terintegrasi untuk mengurangi variasi bentuk kata tanpa menghilangkan informasi yang relevan terhadap polaritas sentimen. Penyesuaian tersebut diharapkan dapat mengurangi *noise* serta menghasilkan representasi teks yang lebih sesuai dengan kebutuhan klasifikasi menggunakan SVM [22].

Berdasarkan pertimbangan tersebut, tahapan *preprocessing* dalam penelitian ini meliputi *cleaning*, *case folding*, *tokenizing*, *normalisasi*, *stopword removal*, dan *stemming*. *Cleaning* digunakan untuk menghilangkan elemen yang tidak relevan, *case folding* untuk menyeragamkan bentuk huruf, dan *tokenizing* untuk memecah teks menjadi unit kata. Selanjutnya, *normalisasi* dilakukan untuk mengubah kata tidak baku atau istilah informal ke bentuk yang lebih baku, *stopword removal* untuk menghilangkan kata yang kurang memberikan informasi terhadap klasifikasi, dan *stemming* untuk mengembalikan kata ke bentuk dasarnya. Rangkaian tahapan tersebut dilakukan secara berurutan agar data teks yang semula memiliki variasi bentuk dapat menjadi lebih konsisten dan siap digunakan pada tahap pelabelan serta klasifikasi.

- Cleaning* dilakukan guna menghapus serta membersihkan beberapa elemen yang tidak diinginkan dalam kalimat ataupun tidak relevan. Hasil *cleaning* ditunjukkan pada Tabel 2.
- Case Folding* dilakukan untuk mengganti format huruf besar menjadi huruf kecil. Hasil *case folding* ditunjukkan pada Tabel 2.
- Tokenizing* dilakukan untuk mengubah kalimat menjadi perkata. Hasil *Tokenizing* ditunjukkan pada Tabel 2

Tabel 2. Hasil *Cleaning*, *Case Folding*, *Tokenizing*

No	Hasil <i>Cleaning</i>	Hasil <i>Case Folding</i>	Hasil <i>Tokenizing</i>
1	antara barak militer dan pesantren bagaimana cara efektif mengatasi kenakalan remaja menurut pengamat pendidikan unika soegijapranata semarang tukiman taruna idiomnya jgn diberikan cap anak nakal tentu mencari akar masalah dari keluarga tapi kita tdk menyalahkan keluarga	antara barak militer dan pesantren bagaimana cara efektif mengatasi kenakalan remaja menurut pengamat pendidikan unik a soegijapranata semarang tukiman taruna idiomnya jgn diberikan cap anak nakal tentu mencari akar masalah dari keluarga tapi kita tdk menyalahkan keluarga	['antara', 'barak', 'militer', 'dan', 'pesantren', 'bagaimana', 'cara', 'efektif', 'mengatasi', 'kenakalan', 'remaja', 'menurut', 'pengamat', 'pendidikan', 'unika', 'soegijapranata', 'semarang', 'tukiman', 'taruna', 'idiomnya', 'jangan', 'diberikan', 'cap', 'anak', 'nakal', 'tentu', 'mencari', 'akar', 'masalah', 'dari', 'keluarga', 'tetapi', 'kita', 'tidak', 'menyalahkan', 'keluarga']



No	Hasil Cleaning	Hasil Case Folding	Hasil Tokenizing
2	jika program barak militer berhasil manfaat bisa dirasakan oleh remaja bermasalah dengan disiplin dan prospek hidup lebih baik masyarakat karena potensi penurunan kriminalitas keluarga dengan perubahan positif pada anak sistem pendidikan dengan	jika program barak militer berhasil manfaat bisa dirasakan oleh remaja bermasalah dengan disiplin dan prospek hidup lebih baik masyarakat karena potensi penurunan kriminalitas keluarga dengan perubahan positif pada anak sistem pendidikan dengan	['jika', 'program', 'barak', 'militer', 'berhasil', 'manfaat', 'bisa', 'dirasakan', 'oleh', 'remaja', 'bermasalah', 'dengan', 'disiplin', 'dan', 'prospek', 'hidup', 'lebih', 'baik', 'masyarakat', 'karena', 'potensi', 'penurunan', 'kriminalitas', 'keluarga', 'dengan', 'perubahan', 'positif', 'pada', 'anak', 'sistem', 'pendidikan', 'dengan']
...			
7063	melihat kenakalan remaja yg sering melakukan tawuran dijalanan bahkan mereka dijadikan begal yg diperalat oleh miras dan narkoba sampai mengganggu keamanan publik gagasan membawa mrk ke barak	melihat kenakalan remaja yg sering melakukan tawuran dijalanan bahkan mereka dijadikan begal yg diperalat oleh miras dan narkoba sampai mengganggu keamanan publik gagasan membawa mrk ke barak	['melihat', 'kenakalan', 'remaja', 'yang', 'sering', 'melakukan', 'tawuran', 'dijalanan', 'bahkan', 'mereka', 'dijadikan', 'begal', 'yang', 'diperalat', 'oleh', 'minuman', 'narkoba', 'sampai', 'mengganggu', 'keamanan', 'membawa', 'mrk', 'ke', 'barak']

- d. Normalisasi dilakukan dengan memanfaatkan library pandas. Proses ini melibatkan pengubahan menjadi bentuk yang baku dari kata-kata yang tidak terstruktur serta tidak baku. Untuk mencapai hal tersebut, kamus ejaan, kamus colloquial, kamus alay, kamus singkatan, serta kamus slang semuanya digunakan bersama dalam satu basis data besar. Pada Tabel 3 ditampilkan hasil dari proses normalisasi tersebut.
- e. Stopword Removal menggunakan library NLTK guna menghilangkan beberapa kata yang termasuk dalam daftar stopwords. Proses ini bertujuan menghilangkan beberapa kata yang tidak mempunyai pengaruh signifikan dalam analisis sentimen. Pada Tabel 3 ditampilkan hasil dari proses stopwords removal
- f. Stemming dilakukan dalam menghilangkan imbuhan pada kata dan mengembalikannya ke bentuk dasarnya. Hasil stemming ditunjukkan pada Tabel 3

Tabel 3. Hasil Normalisasi, Stopword Removal, Stemming

No	Hasil Normalisasi	Hasil Stopword Removal	Hasil Stemming
1	antara barak militer dan pesantren bagaimana cara efektif mengatasi kenakalan remaja menurut pengamat pendidikan unika soegijapranata semarang tukiman taruna idiomnya jangan diberikan cap anak nakal tentu mencari akar masalah dari keluarga tetapi kita tidak menyalahkan keluarga	['barak', 'militer', 'pesantren', 'efektif', 'mengatasi', 'kenakalan', 'remaja', 'pengamat', 'pendidikan', 'unika', 'soegijapranata', 'semarang', 'tukiman', 'taruna', 'idiomnya', 'cap', 'anak', 'nakal', 'mencari', 'akar', 'keluarga', 'menyalahkan', 'keluarga']	barak militer pesantren efektif atas nakal remaja amat didik unika soegijapranata semarang tukiman taruna idiom cap anak nakal cari akar keluarga salah keluarga
2	jika program barak militer berhasil manfaat bisa dirasakan oleh remaja bermasalah dengan disiplin dan prospek hidup lebih baik masyarakat karena potensi penurunan kriminalitas keluarga dengan perubahan positif pada anak sistem pendidikan dengan	['program', 'barak', 'militer', 'berhasil', 'manfaat', 'dirasakan', 'remaja', 'bermasalah', 'disiplin', 'prospek', 'hidup', 'masyarakat', 'potensi', 'penurunan', 'kriminalitas', 'keluarga', 'perubahan', 'positif', 'anak', 'sistem', 'pendidikan']	program barak militer hasil manfaat rasa remaja masalah disiplin prospek hidup masyarakat potensi turun kriminalitas keluarga ubah positif anak sistem didik
...			
7063	melihat kenakalan remaja yang sering melakukan tawuran dijalanan bahkan mereka dijadikan begal yang diperalat oleh minuman keras dan narkoba sampai mengganggu keamanan publik gagasan membawa mereka ke barak militer oleh gubernur jabar kdm cukup	['kenakalan', 'remaja', 'tawuran', 'dijalanan', 'dijadikan', 'begal', 'diperalat', 'minuman', 'keras', 'narkoba', 'mengganggu', 'keamanan', 'publik', 'gagasan', 'membawa', 'barak', 'militer', 'gubernur', 'jabar', 'kdm', 'cemerlang', 'memperbaiki']	nakal remaja tawur jalan jadi begal alat minum keras narkoba mengganggu aman publik gagasan bawa barak militer gubernur jabar kdm cemerlang baik generasi muda

No	Hasil Normalisasi	Hasil Stopword Removal	Hasil Stemming
	cemerlang demi memperbaiki generasi muda.	'generasi', 'muda']	

3.3 Pelabelan Data

Setelah melalui tahapan *preprocessing*, data diberikan label berdasarkan polaritas sentimennya. Penelitian ini menggunakan pendekatan berbasis *lexicon* dengan kamus SenticNet 7 untuk menentukan orientasi sentimen pada setiap data. Nilai polaritas diperoleh berdasarkan bobot kata yang terdapat dalam kamus, kemudian digunakan untuk menentukan kategori sentimen. Data diklasifikasikan ke dalam tiga kategori, yaitu positif, negatif, dan netral. Sentimen dengan nilai polaritas lebih besar dari nol dikategorikan sebagai positif, sedangkan nilai polaritas kurang dari nol dikategorikan sebagai negatif. Data dengan nilai polaritas sama dengan nol dikategorikan sebagai netral. Karena penelitian difokuskan pada klasifikasi sentimen positif dan negatif, data dengan kategori netral selanjutnya dikeluarkan dari dataset yang digunakan dalam pemodelan. Hasil pelabelan terhadap 7.063 data menunjukkan sebanyak 6.596 data positif, 443 data negatif, dan 24 data netral. Dengan demikian, setelah data netral dikeluarkan, diperoleh 7.039 data yang terdiri atas 6.596 data positif dan 443 data negatif.

Tabel 4. Hasil Pelabelan Data

Positif	Negatif	Netral
6596	443	24

Selanjutnya, data dengan kategori netral dihapus sehingga dataset yang digunakan pada tahap pemodelan berjumlah 7.039 data.

Tabel 5. Hasil Pelabelan tanpa Data Netral

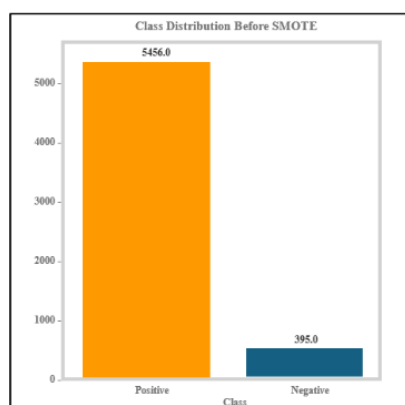
Positif	Negatif
6596	443

Komposisi tersebut menunjukkan adanya ketidakseimbangan kelas yang cukup besar, dengan jumlah data positif jauh lebih banyak dibandingkan data negatif. Ketidakseimbangan tersebut berpotensi menyebabkan model lebih cenderung mengenali kelas mayoritas dan kurang optimal dalam mengidentifikasi kelas minoritas. Oleh karena itu, penanganan *class imbalance* dilakukan pada tahap pemodelan menggunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE).

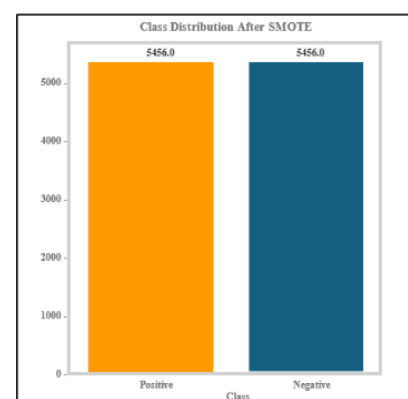
3.4 Optimasi SVM Menggunakan SMOTE

Ketidakseimbangan distribusi kelas pada dataset penelitian menjadi salah satu pertimbangan dalam proses pemodelan. Dari 7.039 data setelah penghapusan sentimen netral, sebanyak 6.596 data termasuk kelas positif dan 443 data termasuk kelas negatif. Kondisi tersebut menunjukkan adanya dominasi kelas positif yang cukup besar dibandingkan kelas negatif. Ketidakseimbangan ini berpotensi menyebabkan model lebih banyak mempelajari karakteristik kelas mayoritas sehingga kemampuan dalam mengenali pola pada kelas minoritas menjadi kurang optimal. Oleh karena itu, diperlukan teknik penyeimbangan data sebelum model SVM digunakan dalam proses klasifikasi.

Penelitian ini menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE) untuk mengatasi ketidakseimbangan tersebut. SMOTE bekerja dengan membentuk sampel sintesis baru pada kelas minoritas berdasarkan karakteristik data yang berdekatan, sehingga distribusi antarkelas pada data latih menjadi lebih seimbang. Penerapan SMOTE dilakukan hanya pada data latih setelah proses pembagian dataset, sedangkan data uji tetap dipertahankan dalam distribusi aslinya. Strategi tersebut diterapkan untuk mencegah terjadinya *data leakage*, sehingga data sintesis yang dihasilkan tidak memengaruhi proses pengujian dan performa model dapat dievaluasi secara lebih objektif terhadap data yang tidak mengalami proses *oversampling*.



Gambar 2. SVM Sebelum SMOTE



Gambar 3. SVM Setelah SMOTE



Berdasarkan hasil pembagian data sebelum penerapan SMOTE, terdapat 5.456 data sentimen positif dan 395 data sentimen negatif, sehingga terlihat adanya ketidakseimbangan distribusi antarkelas. Kondisi tersebut menunjukkan bahwa kelas positif memiliki jumlah data yang jauh lebih dominan dibandingkan kelas negatif. Ketidakseimbangan ini berpotensi menyebabkan model SVM lebih cenderung mengenali kelas mayoritas dan kurang optimal dalam mengidentifikasi data pada kelas minoritas.

Penerapan SMOTE dilakukan dengan membentuk data sintetis pada kelas minoritas berdasarkan karakteristik data yang berdekatan (*Nearest Neighbors*). Setelah proses oversampling, jumlah data pada kelas negatif ditingkatkan dari 395 menjadi 5.456 data, sehingga jumlah data pada kedua kelas menjadi seimbang, masing-masing sebanyak 5.456 data. Dengan demikian, diperlukan penambahan sebanyak 5.061 data sintetis pada kelas negatif. Hasil tersebut menunjukkan bahwa SMOTE berhasil menyeimbangkan distribusi kelas sebelum data digunakan dalam proses pemodelan SVM, sehingga model memiliki proporsi data yang lebih setara dalam mempelajari karakteristik kedua kelas sentimen.

3.5 Pemodelan Support Vector Machine (SVM)

Pemilihan Support Vector Machine (SVM) sebagai algoritma klasifikasi didasarkan pada kemampuannya dalam menangani data teks berdimensi tinggi yang direpresentasikan dalam bentuk fitur numerik. Putra dan Fitriani menunjukkan bahwa SVM dapat digunakan untuk mengklasifikasikan sentimen berbasis data Twitter dan memperoleh akurasi sebesar 92,73% pada penelitian mengenai Program Kampus Mengajar. Namun, kinerja SVM tidak hanya ditentukan oleh algoritma klasifikasi, melainkan juga oleh kualitas representasi data dan distribusi kelas yang digunakan dalam proses pelatihan. Hal tersebut menjadi pertimbangan dalam penelitian ini karena dataset yang diperoleh menunjukkan ketidakseimbangan yang cukup besar antara sentimen positif dan negatif. Oleh karena itu, sebelum proses pemodelan dilakukan, penelitian ini menerapkan teknik Synthetic Minority Over-sampling Technique (SMOTE) untuk meningkatkan representasi kelas minoritas sehingga model tidak cenderung memprediksi kelas mayoritas. Setelah mengurutkan 7.039 data menjadi dua kategori 20% untuk pengujian serta 80% untuk pelatihan proses pemodelan dapat dimulai. Data pelatihan dipergunakan untuk mengembangkan model, sedangkan data pengujian berperan dalam menilai tingkat akurasi model. Pada Tabel 6 pembagian data ini ditampilkan.

Tabel 6. Hasil Pemodelan SVM

	Data Uji	Data Latih
Presentase	20%	80%
Jumlah	1407	5631

3.6 Klasifikasi Menggunakan Support Vector Machine (SVM)

Memisahkan data ke dalam set pelatihan dan pengujian merupakan langkah dalam penggunaan SVM untuk klasifikasi. Gambar 4 menampilkan hasil klasifikasi teks.

accuracy			0.96	1407
macro avg	0.83	0.82	0.82	1407
weighted avg	0.96	0.96	0.96	1407

Gambar 4. Hasil Klasifikasi Teks

Gambar 4 memperlihatkan hasil klasifikasi teks menggunakan *machine learning* dengan nilai akurasi sebesar 0,96 atau 96%. Hasil pelabelan menunjukkan bahwa dari 7.063 data, terdapat 6.596 data sentimen positif, 443 data sentimen negatif, dan 24 data sentimen netral. Distribusi tersebut menunjukkan adanya dominasi sentimen positif dalam dataset, dengan proporsi yang jauh lebih besar dibandingkan sentimen negatif dan netral.

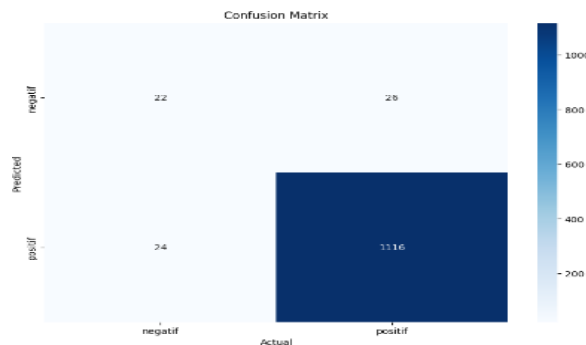
Selanjutnya, data dengan sentimen netral dihilangkan karena penelitian difokuskan pada klasifikasi dua kelas, yaitu sentimen positif dan negatif. Setelah penghapusan tersebut, diperoleh 7.039 data yang digunakan dalam tahapan analisis selanjutnya. Pengurangan data netral dilakukan untuk memastikan bahwa model SVM hanya mempelajari pola yang berkaitan dengan dua polaritas utama, sehingga proses klasifikasi dapat diarahkan secara lebih spesifik pada perbedaan antara opini positif dan negatif.

3.7 Evaluasi

Tahap evaluasi dilakukan untuk mengukur kinerja model *Support Vector Machine* (SVM) dalam mengklasifikasikan sentimen berdasarkan hasil prediksi terhadap data uji. Pengukuran dilakukan menggunakan *confusion matrix* yang digunakan untuk membandingkan hasil prediksi model dengan kelas aktual. Matriks tersebut menghasilkan empat komponen utama, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Keempat komponen tersebut menjadi dasar dalam mengidentifikasi jumlah data yang berhasil maupun yang mengalami kesalahan klasifikasi pada masing-masing kelas sentimen.

Nilai yang diperoleh dari *confusion matrix* selanjutnya digunakan untuk menghitung metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Penggunaan beberapa metrik tersebut memungkinkan kinerja model dinilai secara lebih komprehensif, tidak hanya berdasarkan jumlah prediksi yang benar secara keseluruhan, tetapi juga berdasarkan

kemampuan model dalam mengenali masing-masing kelas sentimen. Hasil *confusion matrix* dan *classification report* yang memuat nilai metrik evaluasi tersebut disajikan pada Gambar 5 dan Gambar 6.



Gambar 5. Hasil Evaluasi Confusion Matrix

Laporan Klasifikasi:				
	precision	recall	f1-score	support
negatif	0.45	0.48	0.46	48
positif	0.98	0.98	0.98	1407
accuracy			0.96	1407
macro avg	0.83	0.82	0.82	1407
weighted avg	0.96	0.96	0.96	1407

Gambar 6. Hasil Laporan Klasifikasi SVM

Gambar 5 menampilkan hasil evaluasi model menggunakan *confusion matrix* dengan nilai TP = 22, TN = 1.335, FP = 26, dan FN = 24. Berdasarkan hasil tersebut, sebanyak 1.357 data berhasil diklasifikasikan dengan benar, yang terdiri atas 22 data *true positive* dan 1.335 data *true negative*. Sementara itu, terdapat 50 data yang mengalami kesalahan klasifikasi, yaitu 26 data *false positive* dan 24 data *false negative*. Hasil tersebut menunjukkan bahwa secara keseluruhan model mampu mengklasifikasikan sebagian besar data uji dengan tingkat kesalahan yang relatif rendah.

Evaluasi lebih lanjut melalui *classification report* pada Gambar 6 menunjukkan bahwa model memperoleh akurasi sebesar 96%. Pada kelas negatif, diperoleh nilai *precision* sebesar 45%, *recall* sebesar 48%, dan *F1-score* sebesar 46%. Sementara itu, kelas positif menghasilkan *precision* sebesar 98%, *recall* sebesar 98%, dan *F1-score* sebesar 98%. Perbedaan capaian antarkelas tersebut menunjukkan bahwa performa model belum sepenuhnya merata, meskipun akurasi keseluruhan yang diperoleh tergolong tinggi. Dengan demikian, akurasi tidak dapat dijadikan satu-satunya indikator untuk menilai keberhasilan model, karena metrik pada masing-masing kelas memberikan informasi yang lebih spesifik mengenai kemampuan model dalam mengenali setiap kategori sentimen.

Tingginya akurasi dan performa pada kelas positif perlu dipahami dalam kaitannya dengan ketidakseimbangan distribusi data antara sentimen positif dan negatif. Dominasi kelas positif menyebabkan model memiliki lebih banyak pola dari kelas mayoritas untuk dipelajari, sehingga kemampuan dalam mengenali sentimen negatif menjadi relatif lebih rendah. Kondisi tersebut tercermin dari nilai *recall* dan *F1-score* kelas negatif yang masing-masing hanya mencapai 48% dan 46%. Meskipun penerapan SMOTE telah dilakukan untuk mengurangi ketidakseimbangan data pada tahap pemodelan, hasil evaluasi menunjukkan bahwa penyeimbangan jumlah sampel belum sepenuhnya menghilangkan kesulitan model dalam membedakan karakteristik kelas negatif. Dengan demikian, model SVM memiliki kinerja klasifikasi yang baik secara keseluruhan, tetapi masih memerlukan pengembangan lebih lanjut untuk meningkatkan kemampuan identifikasi terhadap sentimen negatif.

3.8 Visualisasi Data

Visualisasi data dilakukan untuk memberikan gambaran yang lebih mudah dipahami mengenai distribusi dan kecenderungan sentimen dalam dataset penelitian. Visualisasi yang digunakan meliputi *wordcloud* dan diagram lingkaran (*pie chart*). Diagram lingkaran digunakan untuk memperlihatkan proporsi masing-masing kategori sentimen berdasarkan hasil pelabelan, sedangkan *wordcloud* digunakan untuk mengidentifikasi kosakata yang memiliki frekuensi kemunculan tinggi pada setiap kategori sentimen. Penggunaan kedua bentuk visualisasi tersebut berfungsi sebagai pelengkap hasil klasifikasi sehingga pola umum dalam data dapat diamati secara lebih representatif.

Gambar 7 menunjukkan distribusi polaritas sentimen dari keseluruhan 7.063 tweet sebelum dilakukan penghapusan terhadap data dengan kategori netral. Sementara itu, Gambar 8 memperlihatkan distribusi sentimen setelah data netral dihilangkan sehingga diperoleh 7.039 tweet yang terdiri atas kelas positif dan negatif. Perbandingan kedua visualisasi tersebut menunjukkan adanya perubahan proporsi data setelah kategori netral dikeluarkan dari dataset. Tahapan ini dilakukan karena penelitian difokuskan pada kemampuan model dalam membedakan sentimen positif dan negatif terhadap program barak militer.

Polaritas Sentimen Program Barak Militer
Total Tweets = 5963

Sentiment	Percentage
positive	93.8%
negative	5.8%
neutral	0.4%

Polaritas Sentimen Program Barak Militer
Total Tweets = 7039



Sentiment	Percentage
positive	94.2%
negative	5.8%

[illegible]

This Journal is licensed under a Creative Commons Attribution 4.0 International License



geng motor. Kelompok masyarakat yang mendukung kebijakan tersebut umumnya memandang bahwa institusi formal, termasuk sekolah, belum mampu secara optimal menanamkan nilai kedisiplinan kepada remaja. Dalam konteks tersebut, pendekatan berbasis militer dipersepsikan sebagai solusi yang cepat dan praktis, sekaligus sebagai bentuk technocratic remediation atau upaya pemulihan melalui pendekatan yang bersifat teknokratis. Hal tersebut tercermin dalam kutipan data uji, yaitu “minggu di barak militer remaja bermasalah itu kini cium tangan ibu dan meminta maaf”. Pernyataan tersebut menunjukkan adanya konstruksi moral yang menghubungkan kedisiplinan yang diperoleh melalui pembinaan fisik dengan perubahan perilaku dan kesadaran moral remaja secara langsung.

Sebaliknya, hasil wordcloud pada kelas negatif menunjukkan dominasi sejumlah kata, seperti “salah”, “bandel”, “anak masalah”, “siswa masalah”, dan “anak nakal”. Berdasarkan analisis yang lebih mendalam, terdapat dua faktor utama yang menjadi dasar munculnya penolakan masyarakat terhadap fenomena tersebut. Penggunaan istilah seperti “anak nakal” dan “siswa bermasalah” memperoleh kritik dari masyarakat karena dianggap memberikan label negatif terhadap anak secara terus-menerus. Pemberian label tersebut berpotensi membentuk stigma yang dapat memengaruhi kondisi psikologis, identitas, serta perkembangan sosial anak. Salah satu data pada kategori negatif, yaitu “topik diskusi keluarga... paksa ikut orang hasil bagus”, menunjukkan adanya kritik terhadap proses pelaksanaan pembinaan meskipun sebagian pihak mengakui bahwa hasil yang diperoleh dapat dinilai positif. Pemindahan anak ke barak militer melalui mekanisme yang bersifat memaksa dipandang berpotensi mengabaikan hak-hak anak serta tidak menyentuh faktor penyebab utama munculnya perilaku bermasalah. Dalam perspektif ini, persoalan tersebut lebih banyak dikaitkan dengan disfungsi keluarga dan keterbatasan sistem pendidikan formal dalam menangani permasalahan perilaku remaja.

Kemunculan sejumlah kata yang sama, seperti “barak militer”, “anak”, dan “orang tua”, baik pada kelas positif maupun negatif menunjukkan adanya kemiripan kosakata (vocabulary overlap) antarkelas sentimen. Kondisi tersebut dapat menjadi salah satu faktor yang menyebabkan kinerja model Support Vector Machine (SVM) kurang optimal. Representasi teks berbasis TF-IDF maupun CountVectorizer terutama menangkap frekuensi dan bobot kemunculan kata, tetapi belum sepenuhnya memahami makna kata berdasarkan konteks penggunaannya. Akibatnya, kata atau frasa yang sama dapat muncul dalam teks dengan polaritas sentimen yang berbeda. Hal ini menunjukkan bahwa pendekatan berbasis bag-of-words memiliki keterbatasan dalam membedakan sentimen ketika kedua kelas memiliki kosakata yang relatif serupa, terutama tanpa dukungan representasi semantik yang lebih mendalam seperti semantic embedding.

3.9 Pembahasan

Hasil penelitian menunjukkan bahwa opini masyarakat terhadap program barak militer pada media sosial X didominasi oleh sentimen positif. Dari 7.063 data yang berhasil dikumpulkan setelah proses penyaringan, sebanyak 6.596 data dikategorikan positif dan 443 data dikategorikan negatif, sedangkan 24 data termasuk kategori netral. Setelah data netral dihilangkan, penelitian berfokus pada 7.039 data dengan dua kategori sentimen. Dominasi sentimen positif tersebut menunjukkan bahwa sebagian besar opini yang teridentifikasi dalam dataset penelitian cenderung menerima atau memberikan penilaian positif terhadap program barak militer. Namun, dominasi tersebut tidak serta-merta menunjukkan bahwa kebijakan telah memperoleh penerimaan publik secara menyeluruh karena distribusi sentimen dapat dipengaruhi oleh karakteristik data, proses pelabelan, serta ketidakseimbangan jumlah antarkelas.

Temuan tersebut memiliki kesamaan dengan penelitian Delayana dkk. yang juga menganalisis sentimen publik terhadap program barak militer menggunakan data dari media sosial X. Penelitian Delayana dkk. menggunakan algoritma *Naive Bayes* melalui RapidMiner, sedangkan penelitian ini menggunakan SVM. Kesamaan penggunaan media sosial X menunjukkan bahwa kedua penelitian memanfaatkan opini publik pada platform yang sama sebagai sumber data untuk menggambarkan respons masyarakat terhadap kebijakan barak militer. Perbedaan terletak pada algoritma klasifikasi yang digunakan, sehingga penelitian ini memberikan perspektif metodologis yang berbeda dalam mengidentifikasi pola sentimen pada topik yang sama. Penggunaan SVM dalam penelitian ini juga memungkinkan pengujian terhadap kemampuan algoritma berbasis *margin* dalam mengklasifikasikan teks opini yang memiliki karakteristik bahasa informal dan variasi kosakata media sosial.

Jika dibandingkan dengan penelitian Syalsabina dan Legiawan, terdapat perbedaan yang lebih jelas dari sisi sumber data dan distribusi sentimen. Penelitian tersebut menggunakan komentar YouTube dan algoritma XGBoost dengan akurasi rata-rata sebesar 83%, serta menghasilkan komposisi 44,7% sentimen positif, 12,1% negatif, dan 43,2% netral. Sementara itu, penelitian ini menggunakan data dari media sosial X dan menghasilkan dominasi sentimen positif yang jauh lebih besar, dengan data netral yang relatif sangat sedikit. Perbedaan tersebut menunjukkan bahwa respons masyarakat terhadap isu yang sama dapat menghasilkan distribusi sentimen yang berbeda bergantung pada karakteristik platform dan dataset yang digunakan. Komentar YouTube dan unggahan pada X memiliki pola interaksi, karakteristik pengguna, serta konteks penyampaian opini yang berbeda sehingga tidak tepat apabila distribusi sentimen dari kedua penelitian dibandingkan secara langsung sebagai ukuran penerimaan publik secara keseluruhan.

Hasil penelitian ini juga relevan untuk dibandingkan dengan penelitian Mansyur dkk. yang telah menerapkan SVM dalam mengklasifikasikan sentimen terhadap program barak militer menggunakan data Twitter. Kesamaan algoritma dan topik menunjukkan bahwa penelitian ini berada dalam jalur penelitian yang sama, tetapi tetap memiliki perbedaan pada tahapan pengolahan dan pengelolaan data. Penelitian ini menerapkan *preprocessing* yang mencakup *cleaning*, *case folding*, *tokenizing*, *normalisasi*, *stopword removal*, dan *stemming*, serta menerapkan SMOTE untuk mengatasi ketidakseimbangan kelas pada data latih. Selain itu, penelitian ini melakukan interpretasi terhadap kosakata dominan pada masing-masing kelas melalui *wordcloud*. Hasil evaluasi menunjukkan akurasi sebesar 96%, dengan *precision*, *recall*, dan



F1-score kelas positif sebesar 98%, sedangkan performa kelas negatif masih lebih rendah. Temuan tersebut menunjukkan bahwa tingginya akurasi keseluruhan perlu dibaca secara kritis karena belum sepenuhnya mencerminkan kemampuan model dalam mengenali kelas minoritas.

Perbedaan performa antarkelas tersebut menjadi temuan penting dalam penelitian ini sekaligus memperluas penelitian terdahulu. Hasil *wordcloud* menunjukkan bahwa opini positif banyak berkaitan dengan kata seperti “barak militer”, “disiplin”, “bentuk karakter”, dan “solusi”, yang mengindikasikan adanya pandangan bahwa program tersebut dipersepsikan sebagai sarana pembentukan kedisiplinan dan karakter remaja. Sebaliknya, sentimen negatif memperlihatkan kemunculan kata seperti “anak nakal”, “anak masalah”, “salah”, dan “siswa masalah”, yang menunjukkan adanya kritik terhadap pemberian label negatif serta pendekatan pembinaan yang dianggap berpotensi mengabaikan faktor keluarga, pendidikan, dan kondisi sosial remaja. Dengan demikian, kontribusi penelitian ini tidak hanya terletak pada penerapan SVM untuk mengklasifikasikan sentimen, tetapi juga pada pengaitan hasil klasifikasi dengan konteks substantif kebijakan. Temuan tersebut menunjukkan bahwa penerimaan terhadap program barak militer bersifat kompleks: dukungan terutama berkaitan dengan harapan terhadap peningkatan disiplin dan karakter, sedangkan penolakan berkaitan dengan kekhawatiran terhadap stigmatisasi dan pendekatan pembinaan yang bersifat memaksa.

4. KESIMPULAN

Hasil analisis sentimen terhadap wacana program pembinaan remaja di barak militer menunjukkan bahwa mayoritas opini publik pada media sosial X cenderung positif, yaitu sekitar 94%, sedangkan opini negatif mencapai sekitar 6%. Opini positif umumnya memandang program tersebut sebagai solusi yang tegas dalam meningkatkan kedisiplinan dan membentuk karakter remaja, sementara opini negatif menyoroti potensi stigmatisasi terhadap remaja serta risiko pengabaian terhadap akar permasalahan, seperti disfungsi keluarga dan keterbatasan peran pendidikan formal. Model Support Vector Machine (SVM) yang dioptimalkan menggunakan SMOTE mampu memperoleh akurasi sebesar 96%. Namun, tingginya ketidakseimbangan kelas masih menyebabkan performa klasifikasi kelas minoritas relatif rendah, yang tercermin dari nilai *F1-score* kelas negatif sebesar 46%. Kontribusi penelitian ini terletak pada pengembangan kajian analisis sentimen terhadap program barak militer dengan memanfaatkan data aktual dari media sosial X serta mengombinasikan SVM dan SMOTE untuk menangani permasalahan ketidakseimbangan data. Selain menghasilkan pengukuran kinerja klasifikasi, penelitian ini memberikan interpretasi kontekstual terhadap pola opini publik dengan mengidentifikasi dua kecenderungan utama, yaitu dukungan terhadap aspek kedisiplinan dan pembentukan karakter serta kritik terhadap potensi stigmatisasi dan pendekatan pembinaan yang bersifat memaksa. Temuan tersebut memberikan perspektif empiris bahwa tingginya sentimen positif di media sosial tidak serta-merta menunjukkan efektivitas kebijakan, sehingga penerapan program tetap perlu mempertimbangkan perlindungan hak anak, penguatan fungsi keluarga dan sekolah, serta evaluasi terhadap pendekatan pembinaan yang digunakan. Penelitian ini juga menunjukkan keterbatasan model berbasis pembobotan frekuensi kata dalam menangkap konteks semantik akibat adanya *vocabulary overlap* antarkelas sentimen. Oleh karena itu, penelitian selanjutnya dapat mengembangkan analisis dengan menggunakan sumber data yang lebih heterogen dan model berbasis *contextual embedding*, seperti IndoBERT, untuk memperoleh pemahaman sentimen yang lebih kontekstual.

REFERENCES

- [1] R. Mardining Tyas, “Implementasi Pendidikan Karakter Kedisiplinan (Di SDK Santa Maria Pare),” 2023.
- [2] V. A. Sulistiani and M. Hamka, “Analisis Sentimen Pengguna Media Sosial Terhadap Identitas Kependudukan Digital Menggunakan Metode Support Vector Machine (SVM),” *Journal of Information System Research (JOSH)*, vol. 5, no. 4, pp. 1323–1332, Jul. 2024, doi: 10.47065/josh.v5i4.5614.
- [3] A. F. Nugraha, S. R. Putri, and D. P. Wibowo, “Peran Media Sosial dalam Pembentukan Persepsi Publik terhadap Kebijakan Pemerintah,” *PERSEPTIF: Jurnal Ilmu Sosial dan Humaniora*, vol. 4, no. 1, pp. 36–42, Feb. 2026, doi: 10.70716/perseptif.v4i1.457.
- [4] K. Delayana, S. M. Marpaung, S. N. I. Gulo, and S. Sipayung, “Analisis Sentimen Publik terhadap Barak Militer Menggunakan Naive Bayes pada Rapid MinerJudul,” *remik*, vol. 10, no. 1, pp. 325–334, Jan. 2026, doi: 10.33395/remik.v10i1.15827.
- [5] E. N. Syalsabina and M. K. Legiawan, “Analisis Sentimen Masyarakat terhadap Kebijakan Pembinaan Barak Militer oleh Gubernur Jawa Barat pada Komentar YouTube Menggunakan Metode XGBoost,” *PROSIDING SEMINAR NASIONAL TEKNIK UNIVERSITAS SURYAKANCANA*, Jul. 2025, Accessed: Aug. 15, 2026. [Online]. Available: <https://jurnal.unsur.ac.id/index.php/semnastekunsur/article/view/5383>
- [6] N. L. Tantri, “Analisis Sentimen Pengguna Twitter Terhadap Kebijakan Gubernur Jawa Barat terkait Barak Militer menggunakan Natural Language Processing (NLP).”
- [7] M. F. Setiawan, “Analisis Sentimen Masyarakat terhadap Kebijakan Gubernur Jawa Barat di Sosial Media Menggunakan Algoritma Long Short-Term Memory,” *Journal of Artificial Intelligence and Digital Business (RIGGS)*, vol. 5, no. 2, pp. 15540–15560, 2026, doi: 10.31004/riggs.v5i2.11497.
- [8] N. A. Mansyur, F. Wajidi, and M. R. Rasyid, “Klasifikasi Sentimen terhadap Program Barak Militer Anak Dedi Mulyadi Menggunakan Support Vector Machine,” *Techno.com*, vol. 25, no. 1, p. 153, Feb. 2026, doi: 10.62411/tc.v25i1.15708.
- [9] N. K. Fitrah and N. L. Anggreini, “Analisis Sentimen Opini Warganet Terhadap Pendidikan Militer Bagi Pelajar Menggunakan Algoritma Support Vector Machine,” *ZONasi: Jurnal Sistem Informasi*, vol. 7, no. 3, pp. 862–873, Sep. 2025, doi: 10.31849/4ggj4c78.
- [10] S. Styawati, N. Hendrastuty, and A. R. Isnain, “Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter



- Dengan Metode Support Vector Machine,” *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 6, no. 3, pp. 150–155, Oct. 2021, doi: 10.30591/jpit.v6i3.2870.
- [11] N. Normah, B. Rifai, S. Vambudi, and R. Maulana, “Analisa sentimen perkembangan Vtuber dengan metode support vector machine berbasis SMOTE,” *Jurnal Teknik Komputer*, vol. 8, no. 2, pp. 174–180, 2022.
 - [12] A. Sitanggang, Y. Umaidah, and R. I. Adam, “Analisis sentimen masyarakat terhadap program makan siang gratis pada media sosial X menggunakan algoritma Naïve Bayes,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, pp. 2755–2762, 2024, doi: 10.23960/jitet.v12i3.4902
 - [13] R. Mufidah, H. Triana, and S. Savina, “Analisis Sentimen terhadap Kebijakan Food Estate Menggunakan Algoritma Support Vector Machine,” *Syntax: Jurnal Informatika*, vol. 13, no. 01, pp. 77–90, 2024, url: <https://journal-cs.unsika.ac.id/index.php/syntax/article/view/12155>.
 - [14] E. Apriani and N. Pratiwi, “Analisis Sentimen Game Show Clash of Champions Ruangguru dengan Algoritma KNN dan SVM,” *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 1, pp. 39–51, 2025, doi: 10.52436/1.jpti.556.
 - [15] M. K. K. Insan, U. Hayati, and O. Nurdian, “Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di Google Play Menggunakan Algoritma Naïve Bayes,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, pp. 478–483, 2023, doi: 10.36040/jati.v7i1.6373
 - [16] I. Cero, J. Luo, and J. M. Falligant, “Lexicon-based sentiment analysis in behavioral research,” *Perspectives on behavior science*, vol. 47, no. 1, pp. 283–310, 2024, doi: 10.1007/s40614-023-00394-x.
 - [17] F. Lim and A. T. Ayunda, “Analisis Sentimen Masyarakat Indonesia Terhadap Aplikasi Threads di Twitter Menggunakan Naïve Baiyes,” *JUSIFO (Jurnal Sistem Informasi)*, vol. 9, no. 2, pp. 55–64, 2023, doi: 10.19109/jusifo.v9i2.19965.
 - [18] A. R. Ismail and R. B. F. Hakim, “Implementasi Lexicon Based Untuk Analisis Sentimen Dalam Menentukan Rekomendasi Pantai Di DI Yogyakarta Berdasarkan Data Twitter: Implementasi Lexicon Based,” *Emerging Statistics and Data Science Journal*, vol. 1, no. 1, pp. 37–46, 2023, doi: 10.20885/esds.voll.iss.1.art5.
 - [19] I. T. Julianto and L. Lindawati, “Analisis sentimen terhadap sistem informasi akademik Institut Teknologi Garut,” *Jurnal Algoritma*, vol. 19, no. 1, pp. 458–468, 2022, doi: 10.33364/algoritma/v.19-1.1112.
 - [20] A. S. Firmansyah, A. Aziz, and M. Ahsan, “Optimasi K-nearest neighbor menggunakan algoritma SMOTE untuk mengatasi imbalance class pada klasifikasi analisis sentimen,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 6, pp. 3341–3347, 2023, doi: 10.36040/jati.v7i6.7257.
 - [21] S. Hadiani *et al.*, “Analisis Sentiment Covid-19 Di Twitter Menggunakan Metode Naive Bayes dan SVM,” *Jurnal Teknologi Informasi*, vol. 6, no. 1, 2022, url: <https://jurnal.una.ac.id/index.php/jurti/article/view/2557>.
 - [22] S. Madduru, M. V. Sanjamala, A. Danala, A. Annam, and A. Gorla, “Multilingual sentiment analysis on product reviews and recommendations,” presented at the The 6th International Conference Of Ice-Elinvo 2023: Digital Solutions for Sustainable and Green Development, Yogyakarta, Indonesia, 2025, p. 030038. doi: 10.1063/5.0247941.