ANTIVIRUS: Jurnal Ilmiah Teknik Informatika (p – ISSN: 1978 – 5232; e – ISSN: 2527 – 337X)
Vol. 20 No. 1 Mei 2026, pp. 143 – 156

ANALSIS SENTIMEN TAGAR BUBARKAN DPR DI X.COM MENGGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE* (SVM)

Diterima Redaksi: 21 November 2025; Revisi Akhir: 17 Juli 2026; Diterbitkan Online: 31 Juli 2026

Farrel Julio Akbar[1)], Sahid Maulana[2)], Virdio Samuel Saragih[3)], M. Syamsuddin Wisnubroto[4)], Fajri Farid[5)]
[1, 2, 3,4,5)]Program Studi Sains Data, Fakultas Sains, Institut Teknologi Sumatera
[1, 2,3,4,5)] Jl. Terusan Ryacudu, Way Huwi, Lampung Selatan, Lampung, 35365, Indonesia
e-mail: farrel.122450110@student.itera.ac.id[1)], sahid.122450109@student.itera.ac.id[2)], virdio.122450124@student.itera.ac.id[3)], syamsuddin.wisnubroto@sd.itera.ac.id[4)], fajri.farid@sd.itera.ac.id[3)]

***Abstrak*:** Isu terkait besaran gaji anggota Dewan Perwakilan Rakyat (DPR) kembali menjadi sorotan publik karena dinilai tidak sebanding dengan kinerja yang ditunjukkan. Ketegangan ini mencapai puncaknya pada aksi demonstrasi nasional tanggal 25–28 Agustus 2025, yang diikuti oleh meningkatnya percakapan publik dengan tagar #BubarkanDPR di platform X. Fenomena tersebut menunjukkan perlunya pemetaan sentimen yang akurat untuk memahami persepsi digital masyarakat terhadap isu politik yang berkembang. Penelitian ini mengembangkan model klasifikasi sentimen publik menggunakan algoritma *Support Vector Machine* (SVM) dengan penerapan *Random Oversampling* untuk mengatasi ketidakseimbangan distribusi kelas pada dataset. Sebanyak 555 tweet dikumpulkan melalui proses *web scraping*, kemudian diproses menggunakan tahapan *Natural Language Processing* (NLP) dan diberi label melalui pendekatan semi-otomatis dengan bantuan *Large Language Model* (LLM) yang selanjutnya diverifikasi secara manual untuk memastikan konsistensi anotasi. SVM dipilih karena kinerjanya yang konsisten pada data teks berdimensi tinggi dan stabil dibandingkan beberapa model klasifikasi klasik. Hasil eksperimen menunjukkan bahwa model SVM dengan *Random Oversampling* mencapai akurasi sebesar 92% dengan performa klasifikasi yang stabil pada seluruh kelas sentimen. Hasil studi ini menyajikan bukti empiris terkait reaksi publik terhadap isu sosial-politik yang bersifat krusial dan sensitif, serta menunjukkan potensi analisis sentimen sebagai dasar ataupun landasan strategis dalam pengambilan keputusan berbasis data dalam merespons dinamika opini publik.

Kata Kunci—Analisis Sentimen, Media Sosial X, Support Vector Machine, Random Oversampling, Opini Publik.

***Abstract*:** *The issue of salaries received by members of the Indonesian House of Representatives (DPR) has resurfaced in public discourse, as many consider the compensation disproportionate to their performance. This tension peaked during the nationwide demonstrations held on 25–28 August 2025, which were accompanied by a surge in discussions on X using the hashtag #BubarkanDPR ("Dissolve the DPR"). This increase in online conversations highlights the need for accurate sentiment mapping to better understand the public's digital perception of an evolving political issue. This study develops a public sentiment classification model using the Support Vector Machine (SVM) algorithm, combined with Random Oversampling to address class imbalance within the dataset. A total of 555 tweets were collected through web scraping, processed using Natural Language Processing (NLP) techniques, and labeled using a semi-automated approach with the assistance of a Large Language Model (LLM), followed by manual verification to ensure annotation consistency. SVM was selected due to its consistent performance on high-dimensional text data and its stability compared to classical algorithms such as Naïve Bayes and Decision Tree. Experimental results indicate that the SVM model with Random Oversampling achieved an overall accuracy of 92% and demonstrated stable classification performance across sentiment classes. The findings of this study provide empirical evidence regarding public reactions to critical and sensitive socio-political issues, while highlighting the potential of sentiment analysis as a strategic foundation for data-driven decision making in responding to the dynamics of public opinion.*

Keywords—Sentiment Analysis, Social Media X, Support Vector Machine, Random Oversampling, Public Opinion

Penulis Korespondensi:
M. Syamsuddin Wisnubroto
Program Studi Sains Data
Institut Teknologi Sumatera
Email: syamsuddin.wisnubroto@sd.itera.ac.id
Orchid ID: https://orcid.org/0000-0002-8836-9608

DOI: https://doi.org/10.35457/znctvk95

I. PENDAHULUAN

SENTIMEN publik di media sosial kini menjadi salah satu indikator penting dalam memahami bagaimana masyarakat memandang isu-isu sosial dan politik kontemporer[1]. Pada platform X.com, ekspresi publik sering kali terwujud melalui penggunaan tagar yang viral dan berkembang secara cepat sebagai sarana pengelompokan wacana dan opini kolektif[2]. Tagar yang menuai banyak perhatian dan simpati masyarakat luas beberapa waktu belakangan ini ialah tagar #BubarkanDPR. Tagar ini muncul sebagai bentuk respons masyarakat terhadap pertanyaan mengenai legitimasi dan kinerja lembaga legislatif di Indonesia[3]. Fenomena ini menunjukkan bahwa ruang digital semakin memainkan peran strategis sebagai arena kontrol sosial, di mana masyarakat dapat menyuarakan kritik maupun dukungan terhadap kebijakan dan institusi pemerintah[4].

Analisis sentimen menjadi relevan untuk memetakan pola dukungan, penolakan, maupun sikap netral masyarakat terhadap isu tertentu secara sistematis dan berbasis data [5]. Namun, bahasa yang digunakan pada media sosial cenderung bersifat informal, dipenuhi singkatan, slang, emoji, serta ungkapan sarkastik yang tidak selalu mudah diinterpretasikan secara langsung oleh mesin [6], [7]. Tantangan lain muncul dari ketidakseimbangan distribusi data, karena wacana publik di media sosial umumnya didominasi oleh kelompok pengguna yang lebih vokal dan aktif [8]. Oleh sebab itu, diperlukan pendekatan komputasional yang mampu menangani kompleksitas bahasa sekaligus karakteristik data yang tidak seimbang secara lebih adaptif.

Fenomena *hashtag activism* turut memperlihatkan bagaimana tagar dapat berfungsi sebagai penggerak solidaritas digital dan mempercepat penyebaran opini publik dalam konteks isu politik[4]. Kajian sebelumnya menunjukkan bahwa tagar bernuansa politik sering kali berpusat pada akun-akun berpengaruh, sehingga berpotensi membentuk narasi dominan dan memperkuat polarisasi di tengah Masyarakat[2], [3]. Dengan demikian, analisis sentimen tidak hanya berperan dalam memetakan kecenderungan emosi publik, tetapi juga membantu memahami dinamika perdebatan sosial dan distribusi pengaruh dalam ruang digital secara real time.

Dalam konteks klasifikasi teks media sosial, algoritma *Support Vector Machine* (SVM) menjadi salah satu metode yang dianggap tepat karena kemampuannya dalam memisahkan kelas secara optimal pada ruang berdimensi tinggi[9]. Stabilitas SVM dalam menangani data yang tidak seimbang menjadikannya unggul dibandingkan metode klasik seperti *Naive Bayes* atau *K-Nearest Neighbor*, khususnya pada analisis sentimen berbahasa Indonesia[10],[11]. Berbagai penelitian terdahulu juga menunjukkan bahwa SVM mampu memberikan performa yang konsisten dalam mengklasifikasikan opini publik pada isu politik yang memiliki pola bahasa kompleks[12].

Penelitian ini bertujuan untuk mengidentifikasi dan memetakan kecenderungan sentimen publik terhadap tagar #BubarkanDPR pada media sosial X.com selama periode puncak perbincangan isu tersebut pada tahun 2025. Secara substantif, penelitian ini diarahkan untuk mengetahui persepsi pengguna X.com terhadap gerakan #BubarkanDPR melalui klasifikasi sentimen positif, negatif, dan netral. Secara teknis, penelitian ini juga mengevaluasi kinerja algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan sentimen publik pada data teks media sosial yang memiliki distribusi kelas tidak seimbang. Data dikumpulkan melalui proses *web scraping*, kemudian diolah menggunakan tahapan *Natural Language Processing* (NLP) untuk meningkatkan kualitas teks [6], Selanjutnya, teknik *Random Oversampling* diterapkan untuk menangani ketidakseimbangan distribusi kelas [8], sebelum data diklasifikasikan menggunakan algoritma SVM. Dengan demikian, hasil penelitian ini diharapkan tidak hanya menunjukkan kinerja model klasifikasi, tetapi juga memberikan gambaran empiris mengenai persebaran sentimen publik sebagai dasar pertimbangan berbasis data bagi pembuat kebijakan, media, dan masyarakat dalam merespons dinamika isu sosial-politik secara lebih adaptif dan konstruktif.

II. TINJAUAN PUSTAKA

Kerangka teori pada penelitian ini disusun sebagai dasar konseptual yang menjelaskan teori, penelitian terdahulu, serta pendekatan yang berkaitan dengan analisis sentimen publik pada media sosial X.com terhadap tagar #BubarkanDPR. Penyusunan kajian teori ini bertujuan untuk memberikan landasan ilmiah yang kuat dalam menganalisis arah, intensitas, dan persebaran sentimen publik terhadap isu sosial-politik di ruang digital[5], [1]. Kajian teori mencakup empat aspek utama, yaitu konsep dasar analisis

DOI: https://doi.org/10.35457/znctvk95

sentimen, peran media sosial dan tagar dalam pembentukan opini publik, penerapan algoritma *Support Vector Machine* (SVM) dalam klasifikasi teks berbahasa Indonesia, serta teknik *resampling* untuk mengatasi permasalahan data tidak seimbang (*imbalanced data*)[9], [8].

A. Media Sosial dan Tagar

Media sosial telah menjadi platform utama dalam pertukaran informasi dan opini publik secara cepat dan luas. X.com menjadi salah satu contoh platform media sosial dengan basis pengguna aktif paling banyak digunakan di Indonesia dan seringkali digunakan untuk mendiskusikan berbagai topik, termasuk isu sosial dan politik, sehingga menghasilkan data opini publik dalam jumlah besar dan waktu singkat [2]. Beberapa sifat dan karakteristik tersebut mengukuhkan posisi media sosial sebagai basis data yang relevan dan tepat guna dalam studi analisis sentimen berbasis teks [5].

Tagar berfungsi sebagai mekanisme pengelompokan percakapan berdasarkan topik tertentu dan memudahkan pengguna dalam menelusuri isu yang sedang berkembang. Dalam konteks politik, tagar sering dimanfaatkan untuk menyuarakan tuntutan, membangun solidaritas, serta mengoordinasikan aksi kolektif di ruang digital[4]. Fenomena *hashtag activism* menunjukkan bahwa tagar memiliki peran strategis dalam pembentukan narasi politik dan mobilisasi opini publik secara daring[2].

Penelitian terdahulu mengungkapkan bahwa tagar populer tidak selalu berkembang secara organik, melainkan sering dipengaruhi oleh akun-akun dengan tingkat pengaruh yang tinggi dan narasi yang terorganisir. Kondisi ini mempercepat penyebaran pesan sekaligus berpotensi memperkuat polarisasi opini di kalangan masyarakat[3]. Oleh karena itu, pemahaman terhadap dinamika penggunaan tagar menjadi aspek penting dalam analisis sentimen dan interpretasi perubahan opini publik secara kontekstual [4].

B. Analisis Sentimen pada Media Sosial

Analisis sentimen adalah proses untuk melakukan identifikasi dan mengklasifikasikan opini dan perasaan individu terhadap suatu topik ke dalam kategori positif, negatif, atau netral[5]. Pemanfaatan analisis sentimen dalam media sosial digunakan untuk memahami kecenderungan dan respons publik terhadap isu yang sedang berkembang, terutama ketika opini disampaikan secara spontan dan masif oleh pengguna [1].

Umumnya, metode analisis sentimen terbagi kepada dua pendekatan utama, yaitu *lexicon-based* dan *machine learning*. Pada pendekatan *lexicon-based* menggunakan daftar kata yang telah diberi bobot sentimen tertentu, sedangkan pendekatan *machine learning* mengandalkan algoritma yang dilatih menggunakan data berlabel untuk mengenali pola bahasa dan melakukan prediksi sentimen secara otomatis [9]. Dalam penelitian analisis sentimen berbahasa Indonesia, pendekatan *machine learning* lebih banyak digunakan karena dinilai mampu menyesuaikan diri dengan ragam bahasa alami yang kompleks, termasuk variasi struktur kalimat dan penggunaan istilah tidak baku[12], [11].

Tahapan penting dalam analisis sentimen meliputi pra-pemrosesan data teks, seperti *cleansing*, *case folding*, tokenisasi, *stopword removal*, normalisasi, dan juga *stemming*, yang bertujuan untuk meningkatkan kualitas data sebelum dilakukan klasifikasi[6]. Proses pra-pemrosesan ini berperan penting dalam mengurangi noise dan ambiguitas bahasa sehingga algoritma dapat bekerja secara lebih optimal[7]. Selanjutnya, evaluasi performa model klasifikasi biasanya dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk menilai kemampuan model dalam mengklasifikasikan sentimen secara akurat dan seimbang[8].

C. Support Vector Machine

Kalimat dalam media sosial umumnya bersifat heterogen dan tidak selalu dapat diklasifikasikan secara tegas pada kategori positif, negatif, atau netral. Opini yang telah disampaikan pengguna sering kali ambigu dan mengandung lebih dari satu nuansa makna dalam satu kalimat. Kondisi ini menuntut penggunaan algoritma klasifikasi yang mampu menangani data berdimensi tinggi serta memiliki kemampuan generalisasi yang baik. Dalam ranah mengklasifikasikan teks, algoritma *Support Vector Machine* (SVM) seringkali menjadi pilihan utama karena kemampuannya dalam memisahkan kelas secara optimal, bahkan pada ruang fitur yang memiliki dimensi tinggi [9].

Secara konseptual, SVM bekerja dengan menemukan *hyperplane* terbaik yang akan memisahkan data ke dalam dua atau lebih kelas dengan memaksimalkan jarak antar kelas tersebut[11]. Setiap data

DOI: https://doi.org/10.35457/znctvk95

direpresentasikan sebagai sebuah titik dalam ruang vektor berdimensi banyak, kemudian SVM menentukan garis batas (atau bidang batas) yang memberikan jarak maksimum antara data dari kelas yang berbeda. Pendekatan ini menjadikan SVM relatif stabil terhadap data yang kompleks dan memiliki jumlah fitur yang besar, seperti data teks hasil ekstraksi TF-IDF[13].

Pada kasus klasifikasi multikelas, SVM dapat diimplementasikan menggunakan strategi *one-vs-one* atau *one-vs-all* agar tetap mampu mengklasifikasikan lebih dari dua kelas sentimen secara efektif [9]. Secara matematis, SVM bertujuan memaksimalkan margin antara *hyperplane* pemisah dan titik data terdekat dari masing-masing kelas yang ada, atau sering disebut *support vectors*. Jika diberikan data pelatihan bertanda (x_i, y_i), dengan $x_i$sebagai vektor fitur dan $y_i \in \{+1, -1\}$sebagai label kelas, maka fungsi keputusan SVM dapat dituliskan sebagai berikut:

$$Wi \cdot Xi + b = 0 \quad (1)$$

Keterangan:
W_i : Vektor Bobot / Weight Vector
X_i : Vektor Fitur / Feature Vector
b : Bias/Intercept

$$Wi \cdot Xi + b \leq -1 \quad (2)$$
$$Wi \cdot Xi + b \geq +1 \quad (3)$$

$$min\ \tau(w) = \frac{1}{2}\left|\left|w_i\right|\right|^2 \quad (4)$$

Keterangan:
$\tau(w)$: Fungsi objektif
$\left|\left|w_i\right|\right|$: Panjang vektor

dengan syarat:

$$y_i(xi \cdot w + b) - 1 \geq 0, \forall_i$$
$$\sum_{i=1}^{l} ai(yi((xi \cdot w + b) - 1)) \ \ (i = 1,2,3,\ldots,l) \quad (5)$$

Keterangan:
α : Vektor pengganda Lagrange
b : Bias/Intercept
w : Vektor Bobot
l : Jumlah total sampel data latih
ai : Pengganda Lagrange
yi : Label Kelas
xi : Vektor Fitur

Untuk menangani data yang tidak dapat dipisahkan secara linear (*linearly inseparable*), SVM memperkenalkan parameter *slack* (ξ)serta pendekatan *kernel trick*. Pendekatan ini memungkinkan data dipetakan ke ruang fitur berdimensi tinggi sehingga pola non-linear tetap dapat dipisahkan secara optimal [9]. Penggunaan kernel seperti linear, polinomial, dan *Radial Basis Function* (RBF) memberikan fleksibilitas bagi SVM dalam menangkap hubungan pola data yang kompleks, termasuk ambiguitas bahasa pada teks media sosial [6].

Sejumlah literatur membuktikan bahwa algoritma SVM memiliki performa yang andal dengan memberikan tingkat akurasi tinggi dalam mengklasifikasi sentimen teks, termasuk pada data berbahasa Indonesia yang cenderung berdimensi tinggi dan tidak seimbang. Studi oleh Arsi dan Waluyo[10] serta Putranti dan Winarko[11]membuktikan bahwa SVM mampu mengklasifikasikan sentimen positif, negatif, dan netral pada data media sosial dengan tingkat kesalahan yang relatif rendah. Selain itu, penelitian lain juga menegaskan bahwa kombinasi SVM dengan teknik penanganan data tidak seimbang

DOI: https://doi.org/10.35457/znctvk95

dapat meningkatkan stabilitas dan performa model secara signifikan[13], [8]. Oleh karena itu, SVM dipilih sebagai algoritma utama dalam penelitian ini sebagai alat yang bertujuan untuk menganalisis sentimen publik terhadap isu #BubarkanDPR.

D. Resampling

Dalam analisis sentimen, permasalahan distribusi data yang tidak seimbang (*imbalanced data*) sering dijumpai ketika terdapat ketidakseimbangan distribusi antar kelas yang berbeda secara signifikan, dimana jumlah sampel pada satu kelas jauh lebih rendah dibandingkan dengan kelas lainnya, seperti kelas sentimen negatif kalah dominan dibandingkan sentimen netral atau positif [8]. Kondisi ini dapat menyebabkan algoritma klasifikasi termasuk *Support Vector Machine* (SVM), menjadi bias yakni model memiliki kecenderungan untuk memihak kelas mayoritas dan berdampak langsung secara signifikan pada penurunan performa model dalam mendeteksi kelas minoritas [9].

Salah satu pendekatan yang umum digunakan untuk mengatasi permasalahan tersebut adalah teknik *resampling*, yaitu proses penyesuaian jumlah data pada masing-masing kelas agar distribusinya menjadi lebih seimbang[14]. Teknik resampling secara umum terbagi menjadi dua pendekatan, yaitu *oversampling* dan *undersampling*. Pendekatan *oversampling* dilakukan dengan menambah jumlah sampel pada kelas minoritas, salah satunya melalui metode *Synthetic Minority Oversampling Technique* (SMOTE) yang menghasilkan data sintetis berdasarkan karakteristik data minoritas. Sebaliknya, *undersampling* dilakukan dengan mengurangi jumlah data pada kelas mayoritas melalui pemilihan subset data secara acak atau sistematis[13].

Berbagai penelitian menunjukkan bahwa penerapan teknik resampling mampu meningkatkan performa model *machine learning* dalam analisis sentimen, khususnya dalam meningkatkan nilai *recall* dan *F1-score* pada kelas minoritas[14], [8]. Dengan distribusi data yang lebih seimbang, model dapat mengenali pola sentimen pada seluruh kelas secara lebih adil dan mengurangi risiko bias terhadap kelas mayoritas [9]. Oleh karena itu, dalam penelitian ini teknik resampling diterapkan untuk memastikan distribusi data sentimen pada dataset X.com mendekati proporsional sehingga hasil klasifikasi menjadi lebih representatif.

E. Uji Performa Model

Dalam pengembangan model *machine learning*, tahap evaluasi diperlukan untuk mengukur sejauh mana model mampu melakukan prediksi atau klasifikasi secara akurat dan efisien. Evaluasi performa dilakukan menggunakan metrik evaluasi, yaitu ukuran kuantitatif yang menggambarkan kualitas hasil prediksi model terhadap data uji. Pada penelitian klasifikasi sentimen dengan algoritma *Support Vector Machine* (SVM), metrik evaluasi umumnya meliputi *accuracy*, *precision*, *recall*, dan *F1-score*, karena cocok dan relevan dalam memberikan gambaran performa model klasifikasi secara komprehensif, terutama pada keadaan data yang tidak seimbang[9], [8].

a) Accuracy

Accuracy mengukur proporsi prediksi yang tepat dibandingkan dengan total keseluruhan prediksi model. Metrik ini menunjukkan tingkat ketepatan model secara umum dalam mengklasifikasikan data. Secara matematis, *accuracy* didefinisikan sebagai:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

Keterangan:

TP : Jumlah prediksi positif yang benar
TN : Jumlah prediksi negatif yang benar
FP : Prediksi positif yang salah
FN : Prediksi negatif yang salah

b) Precision

Precision digunakan untuk mengukur seberapa banyak prediksi positif yang benar dari seluruh data yang diprediksi sebagai positif oleh model. Metrik ini penting untuk menilai keandalan model dalam menghindari kesalahan klasifikasi positif. *Precision* dirumuskan sebagai

DOI: https://doi.org/10.35457/znctvk95

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

Keterangan:
TP : Jumlah prediksi positif yang benar
FP : Jumlah prediksi negatif yang benar

c) Recall

Recall digunakan untuk mengukur kemampuan model dalam menemukan kembali semua sampel positif yang relevan dalam dataset. Recall akan menghitung berapa persen yang berhasil diprediksi dengan benar oleh model, dari semua data yang seharusnya positif. Secara matematis, *recall* dapat dihitung dengan:

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

Keterangan:
TP : Jumlah prediksi positif yang benar
FN : Prediksi negatif yang salah

d) F1-Score

F1-Score merupakan rata-rata dari *precision* dan *recall* yang digunakan untuk memperoleh keseimbangan antara kedua metrik tersebut. F1-Score sangat relevan digunakan ketika distribusi kelas tidak seimbang karena mampu merepresentasikan performa model secara lebih adil. Rumus F1-Score dinyatakan sebagai:

$$F1Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (10)$$

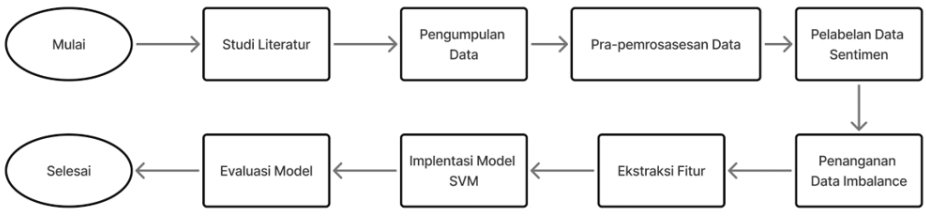
F. Model Alternatif dan Pengembangan Analisis Sentimen

Beberapa penelitian terbaru mengembangkan analisis sentimen menggunakan model berbasis *deep learning* seperti *Long Short-Term Memory* (LSTM) *dan* , yang mampu menangkap konteks semantik secara lebih mendalam dibandingkan metode klasik[15], [16]. Selain itu, pendekatan *word embedding* seperti Word2Vec dan GloVe juga terbukti meningkatkan kualitas representasi fitur teks dalam analisis sentimen politik[17]. Pada aspek penanganan data tidak seimbang, metode lanjutan seperti ADASYN dan SMOTE-Tomek telah dieksplorasi untuk meningkatkan sensitivitas model terhadap kelas minoritas[14]. Meskipun demikian, pendekatan tersebut umumnya memerlukan sumber daya komputasi yang lebih besar, sehingga dalam penelitian ini dipilih metode SVM dengan teknik resampling yang lebih sederhana namun stabil.

III. METODE PENELITIAN

Metodologi penelitian ini disusun sebagai gambaran sistematis mengenai langkah-langkah yang akan dilakukan dalam menganalisis sentimen publik terhadap tagar #BubarkanDPR di platform media sosial X.com. Pendekatan yang digunakan adalah kuantitatif dengan desain eksploratif, menggunakan data cuitan berbahasa Indonesia dari X.com.

A. Flowchart Penelitian



Gambar 1. Diagram Alir Tahapan Analisis Data

DOI: https://doi.org/10.35457/znctvk95

Penelitian dilakukan dengan pendekatan kuantitatif melalui model eksploratif untuk mengenali pola sentimen publik terhadap tagar #BubarkanDPR di media sosial X.com periode 20 Agustus hingga 12 September 2025. Kerangka penelitian disusun mengikuti tahapan sistematis sebagai berikut:

a) Studi Literatur
Menetapkan tujuan penelitian untuk mengidentifikasi dan memetakan pola sentimen, opini positif, negatif, maupun netral dalam sentimen masyarakat Indonesia terhadap isu #BubarkanDPR di platform media sosial X.com selama periode pengambilan data berlangsung.

b) Data Understanding
Mengumpulkan data cuitan berbahasa Indonesia yang memuat tagar #BubarkanDPR tanpa filter lokasi, memeriksa karakteristik data, distribusi frekuensi, ragam ekspresi bahasa, dan keberagaman opini.

c) Data Preparation
Melakukan tahapan pra-pemrosesan berupa *cleansing*, tokenisasi, *case folding*, *stopword removal*, normalisasi bahasa gaul, dan *stemming*; dilanjutkan pelabelan sentimen dan pembentukan dataset berlabel.

d) Modelling
Menerapkan teknik *resampling* (*random oversampling*) untuk mengatasi ketidakseimbangan kelas; melakukan ekstraksi fitur dengan TF-IDF; serta membangun model klasifikasi menggunakan algoritma *Support Vector Machine* (SVM).

e) Evaluation
Mengukur efektifitas dan kinerja model dalam mengidentifikasi sentiman untuk setiap kelas menggunakan beberapa metrik evaluasi yang relevan yakni *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*.

f) Deployment
Menginterpretasikan dan menyajikan hasil analisis sentimen sebagai gambaran opini digital masyarakat; hasil penelitian dapat dimanfaatkan sebagai masukan bagi pembuat kebijakan, media, maupun masyarakat luas dalam merespons dinamika isu politik secara lebih tepat dan berbasis data.

B. Sumber dan Teknik Pengumpulan Data

Pengumpulan data dilakukan dengan menjalankan skrip Python untuk melakukan *scraping data* yang menghasilkan data sekunder dari plaform media sosial X.com (sebelumnya Twitter). Data terdiri dari kumpulan cuitan yang memuat tagar #BubarkanDPR berbahasa Indonesia, diambil dalam periode 20 Agustus hingga 12 September 2025. Cuitan yang dikumpulkan tidak menggunakan filter lokasi, sehingga mewakili opini publik dari seluruh Indonesia selama puncak perbincangan isu tersebut. Berikut merupakan variabel utama yang diambil dari data cuitan.

Tabel 1. Nama Kolom dan Definisi

Nama Kolom	Definisi
Waktu Unggah	Tanggal dan jam saat cuitan dipublikasikan
Teks Cuitan	Isi pesan/opini yang diunggah pengguna
Label Sentimen	Kategori sentimen hasil pelabelan, yaitu positif, negatif, atau netral.
Username	Nama atau identitas akun pengguna.

Teknik pengumpulan data dilakukan dengan menjalankan script berbasis Python yang otomatis mengakses, mengambil, dan menyimpan data cuitan secara konsisten sesuai parameter yang ditentukan. Data yang diperoleh dapat direplikasi dan diolah pada format CSV agar siap untuk analisis lanjutan seperti pra-pemrosesan dan klasifikasi sentimen.

Untuk memperjelas proses pelabelan sentimen, penelitian ini juga menyajikan contoh data cuitan yang telah diklasifikasikan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Kategori sentimen dalam penelitian ini ditentukan berdasarkan posisi cuitan terhadap gerakan #BubarkanDPR. Sentimen positif menunjukkan dukungan terhadap gerakan #BubarkanDPR, sentimen negatif menunjukkan penolakan terhadap gerakan tersebut atau pembelaan terhadap DPR/pemerintah, sedangkan sentimen netral menunjukkan cuitan yang tidak memuat sikap secara jelas.

DOI: https://doi.org/10.35457/znctvk95

Tabel 2. Sampel Data Cuitan dan Label Sentimen

Contoh Cuitan	Label Sentimen	Dasar Pelabelan
"Keinginan rakyat untuk #BubarkanDPR alasannya begitu kuat..."	Positif	Cuitan menunjukkan dukungan terhadap gerakan #BubarkanDPR dan memuat kritik terhadap DPR.
"#BubarkanDPR"	Negatif	Cuitan hanya memuat tagar tanpa opini atau sikap yang jelas.
"DPR masih diperlukan, jangan semua masalah diselesaikan dengan pembubaran."	Netral	Cuitan menunjukkan penolakan terhadap gerakan #BubarkanDPR atau pembelaan terhadap keberadaan DPR.

C. Pra-pemrosesan Data

Pra-pemrosesan data menjadi tahap yang penting guna memastikan kualitas dataset dapat dikatakan layak dan cukup baik, sebelum dilakukan analisis sentimen. Pra-pemrosesan data (*data preprocessing*) bertujuan membersihkan dan mempersiapkan data teks agar dapat diolah secara optimal oleh algoritma machine learning. Tahapan pra-pemrosesan yang dilakukan meliputi:

a) Cleansing
Menghilangkan karakter tidak penting seperti tanda baca, simbol, url/link, mention, hashtag tambahan yang tidak relevan, serta emoji dan karakter non-alfabetik.

b) Case Folding
Mentransformasikan seluruh huruf pada teks cuitan menjadi huruf kecil, agar konsisten dan tidak menimbulkan redudansi akibat perbedaan huruf kecil dan huruf kapital.

c) Tokenisasi
Memecah teks cuitan menjadi kata-kata atau token secara terpisah, sehingga analisis lanjutan dapat dilakukan per kata.

d) Stopword Removal
Menghilangkan kata-kata umum yang tidak memiliki bobot sentimen penting seperti "dan", "di", "ke", "adalah", dsb, sesuai daftar stopword Bahasa Indonesia.

e) Normalisasi dan Pembersihan Slang
Mengonversi kata-kata tidak baku dan slang ke bentuk baku atau padanan yang sesuai dalam Bahasa Indonesia.

f) Stemming
Mengembalikan setiap kata ke bentuk dasarnya, misalnya "membubarkan" menjadi "bubar".

Seluruh proses dilakukan secara otomatis menggunakan library pemrosesan teks di Python. Hasil pra-pemrosesan disimpan dan menjadi dataset siap analisis untuk tahap berikutnya, yaitu pelabelan sentimen dan ekstraksi fitur.

D. Pelabelan Data Sentimen

Pelabelan data sentimen pada penelitian ini dilakukan menggunakan pendekatan semi-otomatis. *Large Language Model* (LLM) dimanfaatkan untuk menghasilkan label awal dengan mempertimbangkan konteks bahasa, slang, serta sarkasme khas media sosial. Selanjutnya, hasil pelabelan tersebut diverifikasi secara manual oleh peneliti untuk memastikan kesesuaian konteks dan mengurangi potensi kesalahan anotasi.Klasifikasi sentimen menggunakan tiga kategori utama, yaitu:

a) Positif, yaitu cuitan yang mengandung dukungan, pujian, atau pernyataan positif terhadap isu yang dibahas.
b) Negatif, yaitu cuitan yang mengandung kritikan, keluhan, atau pernyataan tidak setuju terhadap isu.
c) Netral, yaitu cuitan yang tidak mengekspresikan sikap positif atau negatif secara jelas, atau bersifat informatif dan objektif.

Pelabelan dilakukan dengan memeriksa konteks, slang, dan ekspresi khas media sosial secara cermat agar hasilnya representatif. Setelah proses pelabelan otomatis, dilakukan validasi antar anotator untuk mengukur konsistensi dan mengurangi subjektivitas. Label kategorikal ini kemudian diubah

DOI: https://doi.org/10.35457/znctvk95

menjadi format numerik (contoh: Positif=2, Netral=0, Negatif=1) sehingga dapat diolah oleh algoritma machine learning pada tahap pelatihan model.

E. Penanganan Data Imbalanced

Data imbalanced (ketidakseimbangan data) merupakan masalah yang sering terjadi dalam analisis sentimen media sosial, dimana jumlah cuitan dengan sentimen tertentu (biasanya positif atau netral) jauh lebih sedikit dibanding sentimen lainnya (negatif). Ketidakseimbangan antar kelas dalam data ini dapat menyebabkan model klasifikasi yang dikembangkan menjadi bias terhadap kelas mayoritas dan mengurangi akurasi prediksi kelas minoritas.

Untuk mengatasi hal tersebut, penelitian ini menerapkan teknik *resampling* yang bertujuan menyeimbangkan distribusi kelas sentimen. Teknik yang digunakan meliputi teknik *Random Oversampling*, yaitu metode penambahan data secara acak dari kelas minoritas untuk menyeimbangkan jumlah sampel kelas mayoritas dengan kelas minoritas.

Penggunaan *random oversampling* terbukti meningkatkan performa model klasifikasi dalam berbagai studi analisis sentimen, khususnya dengan algoritma *Support Vector Machine* (SVM). Teknik ini membantu model lebih sensitif terhadap kelas minoritas, sehingga meningkatkan *recall* dan *F1-score* tanpa mengorbankan akurasi keseluruhan. Evaluasi metode *resampling* dilakukan dengan membandingkan metrik performa sebelum dan sesudah penerapan *resampling*. Hasil terbaik adalah *random oversampling* pada kelas minoritas.

F. Ekstraksi Fitur

Ekstraksi fitur merupakan proses penting dalam analisis sentimen untuk mengubah data teks hasil pra-pemrosesan menjadi representasi numerik yang dapat diolah oleh algoritma *machine learning*. Pada penelitian ini, teknik yang digunakan adalah *Term Frequency-Inverse Document Frequency* (TF-IDF). Langkah-langkah ekstraksi fitur meliputi:

a) Perhitungan Term Frequency (TF)
Menghitung frekuensi kemunculan setiap kata pada tiap cuitan agar diketahui kata-kata yang sering muncul dan kemungkinan berperan dalam menentukan sentimen.

b) Perhitungan Inverse Document Frequency (IDF)
Menentukan bobot unik kata berdasarkan seberapa jarang kata tersebut muncul di seluruh cuitan, sehingga kata yang sering muncul di banyak dokumen mendapatkan bobot lebih rendah.

c) Pembangunan Matriks TF-IDF
Mengalikan skor TF dan IDF untuk setiap kata, sehingga terbentuk matriks fitur TF-IDF yang mewakili setiap cuitan dalam bentuk vektor numerik berdimensi sesuai jumlah kata unik pada dataset.

TF-IDF dipilih karena terbukti efektif dalam menyeleksi kata-kata yang benar-benar berpengaruh terhadap sentimen, serta mampu mengurangi pengaruh kata-kata umum yang tidak informatif pada data media sosial X.com. Hasil dari tahap digunakan sebagai input bagi proses pelatihan dan pengujian model klasifikasi sentimen.

G. Implementasi Model SVM

Dalam penelitian analisis sentimen media sosial X, penerapan teknik *resampling* dilakukan sebelum proses pelatihan model *Support Vector Machine* (SVM) untuk mengatasi ketidakseimbangan distribusi kelas. Setelah proses resampling, dataset dibagi menjadi 2 yakni *train data* (data latih) dan *train data* (data ujiI menggunakan skema *hold-out validation* yang representatif, sehingga setiap kelas sentimen memiliki proporsi yang seimbang pada tahap pelatihan dan pengujian.

Pada penelitian ini, *cross-validation* tidak diterapkan karena proses resampling berpotensi menimbulkan *data leakage* apabila dilakukan di dalam setiap lipatan validasi. Oleh karena itu, evaluasi performa model dilakukan menggunakan satu pembagian data latih dan data uji setelah resampling agar hasil pengujian lebih terkontrol dan konsisten.

Fokus penerapan SVM diarahkan pada pelatihan model menggunakan data hasil resampling dengan parameter yang ditentukan melalui eksperimen awal. Pendekatan ini memungkinkan proses pelatihan yang lebih efisien sekaligus tetap menghasilkan performa klasifikasi yang representatif dalam mengenali pola sentimen publik pada data media sosial.

DOI: https://doi.org/10.35457/znctvk95

H. Evaluasi Model

Evaluasi performa model klasifikasi merupakan tahap penting untuk mengukur efektivitas algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan sentimen pada data cuitan media sosial X. Dalam penelitian ini, pengukuran performa dilakukan dengan menggunakan beberapa metrik evaluasi yang umum diterapkan dalam klasifikasi teks, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*.

a) *Accuracy* mengukur persentase prediksi yang benar dari seluruh data uji.
b) *Precision* menunjukkan perbandingan prediksi positif yang benar-benar positif, berguna untuk mengukur keandalan model dalam mengidentifikasi sentimen positif, negatif, atau netral.
c) *Recall* mencerminkan kemampuan model dalam menemukan semua data positif sebenarnya di kelas tertentu, menilai sensitivitas model terhadap kelas minoritas.
d) *F1-score* merupakan *harmonic mean* dari *precision* dan *recall* sehingga memberikan nilai yang seimbang sebagai indikator performa keseluruhan.

Selain metrik tersebut, analisis *confusion matrix* juga dilakukan untuk melihat distribusi kesalahan klasifikasi di antara kelas sentimen. Evaluasi ini dilakukan terhadap data uji yang sebelumnya dipisahkan dari data pelatihan dan telah melewati proses *resampling* agar distribusi kelas menjadi seimbang. Hasil evaluasi digunakan untuk menentukan efektivitas SVM sebagai metode klasifikasi sentimen dalam penelitian ini dan menjadi dasar untuk interpretasi lebih lanjut mengenai pola dan distribusi opini publik mengenai isu #BubarkanDPR di platform X.com.

IV. HASIL DAN PEMBAHASAN

A. Deskripsi Data Penelitian

Pengambilan data dilakukan pada thread X.com dari tanggal 30 Agustus 2025 hingga 12 September 2025 (periode 15 hari) dengan keyword "*#bubarkandpr*" menggunakan library Tweepy, yang berhasil mengumpulkan sebanyak 555 data thread aplikasi X.com. Hasil tersebut memberikan gambar mengenai tren pembicaraan yang terjadi sehubungan dengan topik tersebut dalam rentang waktu yang ditentukan.

Tabel 2. Sampel Data Media Sosial X yang Diperoleh

	date	rawContent
274	2025-09-03 04:06:50+00:00	#BubarkanDPR
121	2025-09-05 05:00:07+00:00	Di sini ada hukum. Kalau hukum tidak ditaati lagi, mari, mari kita panggil hakim. - Pramoedya Ananta Toer 🔧🔨 #BubarkanDPR #StopBayarPajak
363	2025-09-02 11:16:16+00:00	Mereka anggota DPR RI memang agak lain.. #BubarkanDPR
456	2025-09-01 13:30:16+00:00	Keinginan rakyat untuk #BubarkanDPR alasannya begitu kuat. Karena @DPR_RI tidak ada yang mewakili penderitaan rakyat yang begitu sangat tapi tatkala berjuang datang predator #PolisiPembunuhRakyat Pejabat Indonesia dikuasai roh antikris. Pemberontakan sudah mendarah daging
188	2025-09-04 03:00:06+00:00	Dalam hidup kita, cuma satu yang kita punya, yaitu keberanian. Kalau tidak punya itu, lantas apa harga hidup kita ini? - Pramoedya Ananta Toer 😳 #BubarkanDPR #StopBayarPajak

Data yang sudah diambil lalu dilakukan *labelling* atau proses pemberian label pada setiap thread, dengan mengikuti panduan pelabelan menggunakan bantuan layanan *Large Language Model* OpenAI. Layanan ini digunakan untuk menangkap *thread* yang berisi sarkasme dan ambiguitas dari thread-nya. Pemberian label tersebut diberlakukan berdasarkan ketetapan pada Tabel 3.

Tabel 3. Kelas Sentimen dan Interpretasi

Kelas Sentimen	Interpretasi	Sampel Data
Positif	komentar mengekspresikan persetujuan atau dukungan terhadap gerakan #BubarkanDPR. Biasanya ditandai dengan kritik, kekecewaan, kemarahan, atau sentimen negatif terhadap DPR/pemerintah.	"Keinginan rakyat untuk #BubarkanDPR alasannya begitu kuat..."

DOI: https://doi.org/10.35457/znctvk95

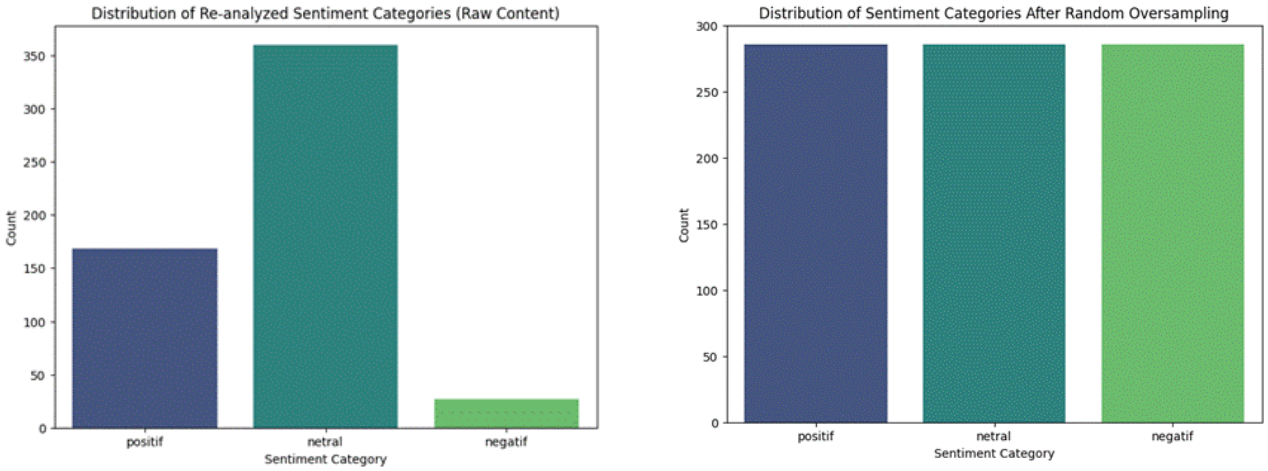
Kelas Sentimen	Interpretasi	Sampel Data
Negatif	komentar mengekspresikan dukungan terhadap DPR atau pemerintah. Biasanya ditandai dengan pujian, pembelaan, atau sentimen positif terhadap DPR/pemerintah.	"DPR masih diperlukan, jangan semua masalah diselesaikan dengan pembubaran."
Netral	komentar tidak jelas menunjukkan dukungan atau penolakan terhadap #BubarkanDPR maupun DPR/pemerintah. Termasuk komentar informatif, bercanda tanpa arah politik, atau hanya menyertakan tagar/emoji/mention tanpa opini.	"#BubarkanDPR"

B. Hasil Pra-Processing Data

Setelah dilakukan pelabelan, langkah selanjutnya adalah melakukan visualisasi untuk melihat persebaran sentimen positif, netral, dan negatif dari setiap *thread*. Visualisasi ini bertujuan untuk memberikan gambaran yang jelas mengenai distribusi sentimen dalam dataset, memudahkan analisis, dan membantu dalam pengambilan keputusan terkait pemrosesan data lebih lanjut, seperti perbaikan *imbalanced data* atau peningkatan akurasi model prediksi. Distribusi *thread* pada setiap kelas sentimen (*rawContent*) atau dengan kondisi *imbalanced data* dapat dilihat pada Gambar 2 (a).

Terdapat sebanyak 360 kalimat netral, 168 kalimat positif, dan 27 kalimat negatif hasil dari pelabelan data otomatis menggunakan layanan LLM OpenAI, yang menunjukkan bahwa data tersebut tidak seimbang (*imbalanced*), sehingga perlu dilakukan *oversampling* untuk menyeimbangkan jumlah masing-masing kelas. Setelah dilakukan proses *oversampling*, data yang telah seimbang dapat diproses kembali untuk analisis lebih lanjut.

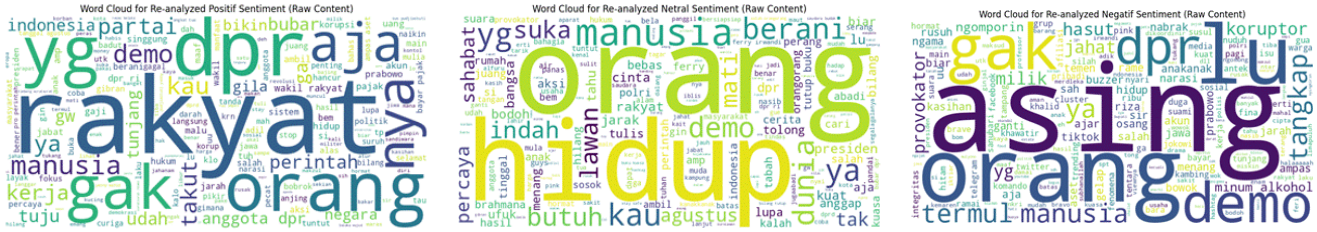
Hasil visualisasi pada *Gambar 3* menegaskan temuan penting mengenai pola bahasa yang digunakan publik di media sosial. Analisis *wordcloud* membantu peneliti mengidentifikasi kata yang paling sering muncul pada *thread* atau *postingan* dalam setiap kategori sentimen. Pada *Gambar 3(a)*, terlihat bahwa kata "rakyat", "dpr", dan "orang" mendominasi kalimat positif, menunjukkan kecenderungan dukungan publik terhadap isu pembubaran DPR. Sementara itu, *Gambar 3(b)* memperlihatkan bahwa kata "orang" dan "hidup" paling sering muncul dalam komentar netral, yang cenderung informatif dan tidak menunjukkan posisi terhadap isu *#BubarkanDPR*. Di sisi lain, *Gambar 3(c)* mengungkap dominasi kata "asing", "orang", dan "demo", diikuti oleh "lu", "gak", dan "dpr", yang mencerminkan ekspresi kritik dan ketidakpuasan terhadap DPR maupun pemerintah. Secara keseluruhan, visualisasi ini menegaskan bahwa kata "rakyat" dan "dpr" menjadi pusat percakapan lintas kategori sentimen, menandakan kuatnya perhatian publik terhadap isu politik ini dan relevansinya sebagai cerminan opini digital masyarakat.



(a) Proses sebelum Random Oversampling *(b) Proses setelah Random resampling (oversampling)*

Gambar 2. Distribusi Kelas Sentimen

DOI: https://doi.org/10.35457/znctvk95



(a) Sentimen Positif *(b) Sentimen Netral* *(c) Sentimen Negatif*

Gambar 3. Wordcloud

C. Teknik Resampling dengan Random Oversampling

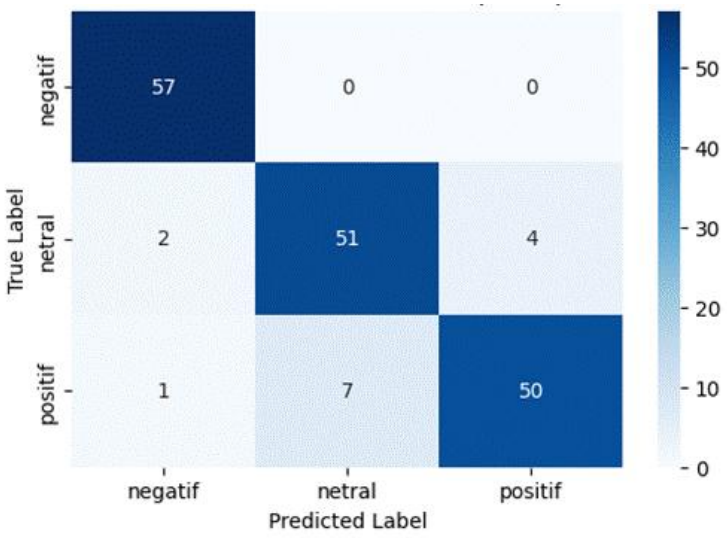
Berdasarkan distribusi kelas sentimen pada Gambar 2(a), data yang digunakan dalam penelitian ini memiliki ketidakseimbangan jumlah data pada setiap kelas. Hasil pelabelan menunjukkan bahwa kelas netral memiliki jumlah data paling dominan, diikuti oleh kelas positif, sedangkan kelas negatif memiliki jumlah data paling sedikit. Ketidakseimbangan tersebut dapat menyebabkan model klasifikasi cenderung lebih mengenali kelas mayoritas dan kurang optimal dalam mengidentifikasi kelas minoritas. Oleh karena itu, penelitian ini menerapkan teknik *Random Oversampling* untuk menyeimbangkan distribusi data sebelum dilakukan proses klasifikasi menggunakan algoritma *Support Vector Machine* (SVM).

Random Oversampling merupakan teknik penanganan data tidak seimbang dengan cara menambahkan sampel pada kelas minoritas secara acak melalui duplikasi data yang sudah ada. Pada penelitian ini, proses *oversampling* dilakukan terhadap kelas yang memiliki jumlah data lebih sedikit agar distribusi setiap kelas menjadi seimbang. Sebagaimana ditunjukkan pada Gambar 2(b), setelah dilakukan proses *Random Oversampling*, jumlah data pada masing-masing kelas sentimen menjadi 285 data. Dengan distribusi yang lebih seimbang, model SVM diharapkan dapat mempelajari pola dari setiap kelas sentimen secara lebih proporsional.

Setelah proses *Random Oversampling*, data kemudian digunakan dalam tahap ekstraksi fitur dan klasifikasi menggunakan algoritma SVM. Evaluasi model dilakukan dengan perhitungan *confusion matrix* yang disajikan pada gambar 4, lalu dihitung kembali menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil evaluasi performa model SVM setelah penerapan *Random Oversampling* disajikan pada Tabel 4.

Tabel 4. Hasil Evaluasi Model SVM dengan Random Oversampling

Kelas Sentimen	F1-Score	Recall (%)	Precision (%)
Negatif	0.97	1.00	0.95
Netral	0.89	0.89	0.88
Positif	0.89	0.86	0.93



Gambar 4. Confusion Matrix Algoritma SVM

DOI: https://doi.org/10.35457/znctvk95

Nilai evaluasi pada Tabel 4 diperoleh dari hasil perbandingan antara label aktual dan label prediksi pada data uji. Berdasarkan *confusion matrix* pada Gambar 4, jumlah prediksi benar diperoleh dari penjumlahan nilai diagonal utama, yaitu 57 data negatif, 51 data netral, dan 50 data positif yang berhasil diklasifikasikan sesuai dengan kelas sebenarnya. Dengan demikian, terdapat 158 prediksi benar dari total 172 data uji, sehingga nilai *accuracy* model diperoleh sebesar 158/172 = 0,9186 atau dibulatkan menjadi 0,92. Nilai *precision*, *recall*, dan *F1-score* pada setiap kelas dihitung berdasarkan jumlah prediksi benar dan kesalahan klasifikasi pada masing-masing kelas sentimen.

V. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, analisis sentimen terhadap 555 data cuitan dengan tagar #BubarkanDPR menunjukkan bahwa opini publik di platform X.com memiliki kecenderungan yang beragam, dengan distribusi awal yang didominasi oleh sentimen netral, kemudian diikuti oleh sentimen positif dan negatif. Dominasi sentimen netral menunjukkan bahwa sebagian besar cuitan yang dikumpulkan bersifat informatif, berupa kutipan, penyebaran tagar, atau tidak menunjukkan sikap yang jelas terhadap gerakan #BubarkanDPR. Sementara itu, keberadaan sentimen positif menunjukkan adanya dukungan terhadap gerakan #BubarkanDPR atau kritik terhadap DPR, sedangkan sentimen negatif merepresentasikan penolakan terhadap gerakan tersebut atau pembelaan terhadap DPR/pemerintah.

Dari sisi pemodelan, implementasi algoritma *Support Vector Machine* (SVM) yang dipadukan dengan teknik *Random Oversampling* mampu menghasilkan kinerja klasifikasi yang baik pada data sentimen yang memiliki distribusi kelas tidak seimbang. Model memperoleh nilai *accuracy* sebesar 0,92, dengan nilai *recall* pada kelas negatif sebesar 1,00, kelas netral sebesar 0,89, dan kelas positif sebesar 0,86. Hasil ini menunjukkan bahwa SVM mampu mengklasifikasikan sentimen publik secara cukup konsisten setelah dilakukan penyeimbangan data. Meskipun demikian, nilai *recall* yang sangat tinggi pada kelas negatif perlu diinterpretasikan secara cermat karena dapat dipengaruhi oleh penerapan teknik *oversampling* pada data minoritas.

Secara keseluruhan, penelitian ini menunjukkan bahwa analisis sentimen dapat digunakan untuk memetakan persepsi digital masyarakat terhadap isu sosial-politik yang berkembang di media sosial. Hasil klasifikasi tidak hanya memberikan gambaran mengenai performa model, tetapi juga dapat menjadi dasar awal dalam memahami dinamika opini publik terhadap isu #BubarkanDPR secara berbasis data.

Untuk penelitian selanjutnya, pengembangan dapat dilakukan dengan memperluas jumlah data dan rentang waktu pengambilan data agar pola perubahan opini publik dapat diamati secara lebih komprehensif. Selain itu, penelitian berikutnya dapat membandingkan kinerja SVM dengan model lain seperti *Bidirectional Encoder Representations from Transformers* (BERT), *Long Short-Term Memory* (LSTM), atau pendekatan representasi teks berbasis *word embedding*. Evaluasi terhadap teknik penyeimbangan data lain, seperti SMOTE, ADASYN, atau *undersampling*, juga dapat dilakukan untuk memperoleh hasil klasifikasi yang lebih stabil dan representatif.

UCAPAN TERIMA KASIH

Penulis menyampaikan apresiasi kepada Institut Teknologi Sumatera atas dukungan riset dan lingkungan akademik yang kondusif selama proses penelitian ini berlangsung. Ucapan terima kasih juga diberikan kepada para dosen Program Studi Sains Data atas arahan, masukan, dan bimbingan yang sangat berarti dalam penyusunan penelitian ini.

Ucapan terima kasih juga disampaikan kepada rekan-rekan sejawat yang telah memberikan dukungan, berdiskusi, dan berbagi pengetahuan selama proses penelitian berlangsung, serta kepada berbagai pihak lain yang turut berkontribusi, baik secara langsung maupun tidak langsung, hingga penelitian ini dapat diselesaikan dengan sebaik-baiknya.

DOI: https://doi.org/10.35457/znctvk95

DAFTAR RUJUKAN

[1] H. Winda dan P. Ricky, "Pemanfaatan Algoritma Machine Learning dalam Mendeteksi Pola dan Dinamika Opini Publik di Media Sosial," *Jurnal Teknologi Informasi dan Data*, vol. 15, no. 1, hlm. 45–58, 2024.

[2] D. Anwar, "Dinamika Tagar dan Kontrol Sosial: Kajian Jangka Panjang Opini Publik di Platform X.com," *Jurnal Kajian Media dan Komunikasi*, vol. 8, no. 1, hlm. 30–45, 2025.

[3] R. Nadjima, "Dinamika Legitimasi Publik dan Kritik Terhadap Lembaga Legislatif Melalui Tagar Politik di Media Sosial Indonesia," *Jurnal Sosial Politik Nusantara*, vol. 10, no. 3, hlm. 211–225, 2024.

[4] F. Nancy, W. Yuliana, dan S. Hadi, "Hashtag Activism pada Isu Politik di Media Sosial," *Jurnal Ilmu Komunikasi*, vol. 12, no. 2, hlm. 100–115, 2024.

[5] W. Andriyani, T. Utama, dan L. Rahmat, "Analisis Sentimen pada Ulasan Produk Menggunakan Metode Pembelajaran Mesin," *JIKO*, vol. 8, no. 1, hlm. 173–185, 2024.

[6] G. R. Smith, *Introduction to Natural Language Processing for Social Media Data*. New York: Tech Press, 2022.

[7] Y. Yunitasari, A. Musdholifah, dan A. K. Sari, "Sarcasm Detection for Sentiment Analysis in Indonesian Language Tweets," *Indonesian Journal of Computing and Cybernetics Systems*, 2019.

[8] A. D. Syafutra, R. Kadir, dan N. R. Semara, "Analisis Sentimen Omnibus Law di Twitter dengan Machine Learning dan Teknik Resampling," *Jambura Journal of Informatics*, vol. 7, no. 1, hlm. 22–35, 2025.

[9] M. Rahman dan R. Fadhil, "Support Vector Machine for Text Classification: A Comprehensive Review," *Journal of Artificial Intelligence Research*, vol. 45, hlm. 112–130, 2021.

[10] A. Arsi dan I. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Support Vector Machine," dalam *Prosiding Seminar Nasional Teknologi Informasi*, 2021.

[11] N. D. Putranti dan E. Winarko, "Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine," *Indonesian Journal of Computing and Cybernetics Systems*, vol. 8, no. 2, hlm. 187–197, 2014.

[12] A. M. Yolanda dan R. T. Mulya, "Implementasi Metode Support Vector Machine untuk Analisis Sentimen pada Ulasan Aplikasi Sayurbox di Google Play Store," *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, vol. 6, no. 2, hlm. 76–83, 2024, doi: 10.35580/variansiunm258.

[13] N. Fajriyah, "Implementasi Support Vector Machine dan SMOTE pada Analisis Sentimen," *JIPI: Journal of Informatics & Programming Innovation*, vol. 9, no. 3, hlm. xx–xx, 2025, [Daring]. Tersedia pada: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi/article/view/6246

[14] B. P. Nugroho dan S. A. Dewi, "Evaluasi Metode Resampling ADASYN dan SMOTE-Tomek untuk Mengatasi Data Imbalanced pada Analisis Sentimen," *Jurnal Informatika*, vol. 11, no. 2, hlm. 88–102, 2024.

[15] J. Ahmad dan S. Siregar, "Efektivitas Model BERT dalam Klasifikasi Sentimen Multibahasa pada Data Twitter (X)," *Jurnal Ilmu Komputer dan Informasi*, vol. 18, no. 1, hlm. 1–15, 2025.

[16] K. L. Wijaya, "Perbandingan Kinerja LSTM dan SVM untuk Analisis Sentimen Teks Pendek Bahasa Indonesia," dalam *Prosiding Konferensi Nasional Kecerdasan Buatan*, 2024.

[17] T. Susanto, M. Hadi, dan L. Utami, "Penerapan Word2Vec dan GloVe untuk Peningkatan Akurasi Analisis Sentimen Politik," *Jurnal Data Sains*, vol. 5, no. 3, hlm. 200–215, 2023.

DOI: https://doi.org/10.35457/znctvk95