

Analisis Sentimen Masyarakat Terhadap Pembentukan Provinsi Luwu Raya Pada Komentar Tiktok Menggunakan Metode TF-IDF Dan Support Vector Machine

Abdillah SAS¹, Muhammad Nur Arbi², Hermila A³

¹Teknologi Informasi, Universitas Bosowa, Makassar;

²Teknik Informatika, Universitas Teknologi Akba, Makassar;

³Pendidikan Teknologi Informasi, Universitas Negeri Gorontalo, Gorontalo
Email:abdillah.sas@universitasbosowa.ac.id

Abstract—The discourse on the establishment of Luwu Raya Province as a new autonomous region in South Sulawesi generated diverse public responses, part of which was reflected in comments on the TikTok platform. The large volume of comments and the use of non-standard language made manual sentiment analysis inefficient. This study aimed to classify public sentiment toward the discourse using the Term Frequency-Inverse Document Frequency (TF-IDF) method for feature extraction and the Support Vector Machine (SVM) algorithm for classification. A total of 3,227 TikTok comments were collected through web scraping, processed through text preprocessing stages, and automatically labeled using a lexicon-based approach into three classes: negative, neutral, and positive. Features were extracted using TF-IDF with a combination of unigrams and bigrams, and the dataset was split into 80% training data and 20% testing data using a stratified scheme. The test results showed that the SVM model achieved an accuracy of 95.51%, with a stable 5-fold cross validation result ranging from 94.12% to 95.82% (mean 95.13%). Further analysis revealed that neutral sentiment was dominant (68.05%), followed by negative (26.89%) and positive (5.05%) sentiment, while the word cloud analysis indicated that negative sentiment was largely driven by the technical impact of the demonstrations rather than rejection of the expansion idea itself. This study demonstrated that the combination of TF-IDF and SVM was effective for analyzing public sentiment toward regional governance issues based on TikTok social media data.

Keywords— *Luwu Raya Province, Sentiment Analysis, Support Vector Machine, TF-IDF, TikTok Comments*

Abstrak— Wacana pembentukan Provinsi Luwu Raya sebagai daerah otonom baru di Sulawesi Selatan menuai beragam respons dari masyarakat, salah satunya tercermin melalui komentar pada platform TikTok. Volume komentar yang besar serta penggunaan bahasa yang tidak baku menyebabkan analisis sentimen secara manual menjadi tidak efisien. Penelitian ini bertujuan untuk mengklasifikasikan sentimen masyarakat terhadap wacana tersebut menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) sebagai teknik ekstraksi fitur dan Support Vector Machine (SVM) sebagai algoritma klasifikasi. Data penelitian berupa 3.227 komentar TikTok dikumpulkan melalui teknik web scraping, kemudian diolah melalui tahapan text preprocessing dan diberi label secara otomatis berbasis kamus (lexicon-based labeling) ke dalam tiga kelas, yaitu negatif, netral, dan positif. Fitur diekstraksi menggunakan TF-IDF dengan kombinasi unigram dan bigram, kemudian data dibagi menjadi 80% data latih dan 20% data uji secara stratifikasi. Hasil pengujian menunjukkan bahwa model SVM memperoleh akurasi sebesar 95,51%, dengan hasil validasi silang (5-fold cross validation) yang stabil pada kisaran 94,12%–95,82% (rata-rata 95,13%). Analisis lebih lanjut menunjukkan sentimen netral mendominasi (68,05%), diikuti negatif (26,89%) dan positif (5,05%), dengan visualisasi word cloud mengungkap bahwa sentimen negatif banyak dipicu oleh dampak teknis aksi demonstrasi, bukan penolakan terhadap gagasan pemekaran itu sendiri. Penelitian ini membuktikan bahwa kombinasi TF-IDF dan SVM efektif digunakan untuk menganalisis sentimen publik terhadap isu pemerintahan daerah berbasis data media sosial TikTok.

Kata Kunci— *Analisis Sentimen, Komentar TikTok, Provinsi Luwu Raya, Support Vector Machine, TF-IDF*

I. PENDAHULUAN

Perkembangan teknologi informasi dan meluasnya penggunaan media sosial telah mengubah cara masyarakat menyampaikan opini terhadap berbagai isu publik, termasuk kebijakan dan wacana pemerintahan. Kemudahan akses internet melalui perangkat seluler mendorong lahirnya berbagai platform media sosial yang berfungsi sebagai ruang komunikasi, penyebaran informasi, sekaligus

pembentukan opini publik secara luas [1]. Volume opini yang dihasilkan pengguna media sosial terus meningkat dan bersifat tidak terstruktur, sehingga analisis sentimen secara manual menjadi tidak efisien dan mendorong kebutuhan terhadap pendekatan otomatis berbasis pembelajaran mesin (*machine learning*) untuk mengklasifikasikan kecenderungan opini masyarakat [1].

Salah satu isu yang belakangan ramai diperbincangkan masyarakat, khususnya warga Sulawesi

Selatan, adalah wacana pembentukan Provinsi Luwu Raya sebagai daerah otonom baru (DOB) yang mencakup Kabupaten Luwu, Kabupaten Luwu Utara, Kabupaten Luwu Timur, dan Kota Palopo. Perjuangan pemekaran ini telah berlangsung sejak tahun 1950-an dan kembali menguat pada awal tahun 2026, ditandai oleh aksi demonstrasi masyarakat yang memblokir Jalan Trans Sulawesi selama beberapa hari sebagai bentuk desakan kepada pemerintah pusat [2]. Salah satu alasan utama tuntutan pemekaran tersebut adalah ketimpangan pembangunan infrastruktur antara wilayah Luwu Raya dan pusat pertumbuhan Sulawesi Selatan lainnya [2]. Di sisi lain, kajian akademik yang dilakukan Institut Pemerintahan Dalam Negeri (IPDN) menilai bahwa kawasan Luwu Raya memiliki kapasitas fiskal yang memadai untuk menopang pemerintahan provinsi secara mandiri apabila pemekaran terwujud [3], sementara Pemerintah Provinsi Sulawesi Selatan menegaskan bahwa keputusan pembentukan daerah otonom baru sepenuhnya berada pada kewenangan pemerintah pusat dan harus melalui tahapan administratif yang berlaku [4]. Dinamika ini menunjukkan bahwa wacana pembentukan Provinsi Luwu Raya bukan sekadar persoalan administratif, tetapi juga melibatkan beragam kepentingan, harapan, serta sikap pro maupun kontra dari masyarakat.

Perdebatan publik mengenai isu ini turut berkembang secara masif di media sosial, salah satunya TikTok, yang telah menjadi salah satu platform berbagi video terpopuler di Indonesia dan banyak digunakan masyarakat untuk mengikuti serta menanggapi pemberitaan terkini, termasuk isu pemekaran Luwu Raya. Kolom komentar pada video-video pemberitaan terkait Luwu Raya di TikTok memuat ribuan opini masyarakat, mulai dari dukungan, kritik, kekecewaan, hingga harapan terhadap proses pembentukan provinsi baru tersebut. Namun demikian, komentar tersebut umumnya ditulis dengan bahasa yang tidak baku, bercampur bahasa daerah, singkatan, dan emotikon, sehingga menyulitkan proses pemahaman kecenderungan sentimen masyarakat secara manual apabila dilakukan pada data dengan volume besar.

Untuk mengatasi permasalahan tersebut, analisis sentimen berbasis machine learning menjadi pendekatan yang relevan untuk mengklasifikasikan opini masyarakat ke dalam kategori sentimen positif, negatif, maupun netral secara otomatis. Beberapa penelitian terdahulu telah membuktikan efektivitas kombinasi metode *Term Frequency-Inverse Document Frequency* (TF-IDF) sebagai teknik ekstraksi fitur dan *Support Vector Machine* (SVM) sebagai algoritma klasifikasi pada berbagai konteks analisis sentimen media sosial. Br Sinulingga dan Sitorus [5] menerapkan TF-IDF dan SVM untuk menganalisis opini masyarakat terhadap film horor Indonesia dan menunjukkan bahwa kombinasi kedua metode tersebut mampu mengklasifikasikan sentimen ulasan dengan baik. Sementara itu, Fernandes dkk. [1] menerapkan pendekatan serupa pada komentar terkait Bitcoin di platform X dan memperoleh kinerja klasifikasi terbaik pada skema pembagian data latih dan data uji 90:10 dengan kernel linear. Kedua penelitian tersebut menunjukkan bahwa SVM dengan pembobotan fitur TF-IDF mampu menangani data

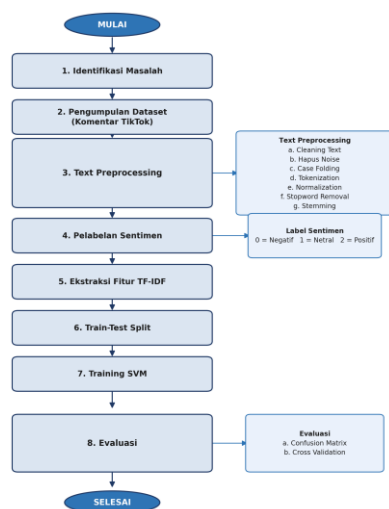
teks berbahasa Indonesia yang tidak terstruktur dan bervariasi secara efektif.

Meskipun penerapan TF-IDF dan SVM telah banyak dikaji pada berbagai topik seperti isu ekonomi, hiburan, maupun kebijakan publik lainnya, sejauh penelusuran penulis belum ditemukan penelitian yang secara khusus menganalisis sentimen masyarakat terhadap wacana pembentukan Provinsi Luwu Raya, terutama berdasarkan data komentar dari platform TikTok. Padahal, pemahaman terhadap kecenderungan sentimen publik atas isu ini penting sebagai bahan pertimbangan bagi para pemangku kepentingan, baik pemerintah daerah, tim pengusul pemekaran, maupun akademisi, dalam menyusun strategi komunikasi publik serta mengevaluasi tingkat dukungan masyarakat terhadap rencana pembentukan daerah otonom baru tersebut.

Berdasarkan uraian di atas, penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap pembentukan Provinsi Luwu Raya berdasarkan komentar pada video TikTok menggunakan metode TF-IDF sebagai teknik ekstraksi fitur dan algoritma *Support Vector Machine* sebagai model klasifikasi. Data komentar yang telah dikumpulkan melalui proses *crawling* terlebih dahulu diolah melalui tahapan *text preprocessing*, meliputi *cleaning*, *case folding*, *tokenizing*, normalisasi kata tidak baku, *stopword removal*, dan *stemming*, sebelum dilakukan pelabelan sentimen dan proses klasifikasi. Kinerja model yang dibangun selanjutnya dievaluasi menggunakan metrik akurasi, presisi, *recall*, F1-score, serta validasi silang (*cross validation*) untuk memastikan konsistensi hasil klasifikasi. Melalui penelitian ini, diharapkan dapat diperoleh gambaran objektif mengenai kecenderungan opini masyarakat terhadap wacana pembentukan Provinsi Luwu Raya, sekaligus memberikan kontribusi metodologis dalam penerapan TF-IDF dan SVM pada kasus analisis sentimen isu pemerintahan daerah berbasis data media sosial TikTok.

II. METODE

Penelitian ini dilaksanakan melalui delapan tahapan utama sebagaimana ditunjukkan pada Gambar 1, yaitu identifikasi masalah, pengumpulan dataset, *text preprocessing*, pelabelan sentimen, ekstraksi fitur TF-IDF, *train-test split*, training SVM, dan evaluasi model. Kedelapan tahapan tersebut dilakukan secara berurutan agar data yang digunakan pada proses klasifikasi telah bersih, konsisten, dan representatif terhadap opini masyarakat yang sesungguhnya.



Gambar 1. Diagram Alir Tahapan Penelitian

A. Identifikasi Masalah

Tahap ini merupakan tahap awal penelitian yang bertujuan merumuskan permasalahan yang akan diteliti, yaitu bagaimana mengklasifikasikan sentimen masyarakat terhadap wacana pembentukan Provinsi Luwu Raya secara otomatis dan akurat berdasarkan komentar pada platform TikTok. Besarnya volume komentar serta penggunaan bahasa yang tidak baku menyebabkan analisis sentimen secara manual menjadi tidak efisien, sehingga dibutuhkan pendekatan berbasis machine learning, yaitu kombinasi metode TF-IDF sebagai teknik ekstraksi fitur dan algoritma *Support Vector Machine* (SVM) sebagai model klasifikasi.

B. Pengumpulan Dataset dan Text Processing

Data penelitian dikumpulkan dari kolom komentar sejumlah video TikTok yang memuat pemberitaan dan pembahasan mengenai wacana pembentukan Provinsi Luwu Raya. Pengumpulan data dilakukan melalui teknik web scraping, sehingga diperoleh data mentah dengan atribut berupa teks komentar (*text*), jumlah tanda suka (*diggCount*), jumlah balasan (*replyCommentTotal*), waktu unggah komentar (*createTimeISO*), nama pengguna (*uniqueId*), tautan video asal komentar (*videoWebUrl*), dan identitas unik pengguna (*uid*). Setelah dilakukan penyaringan terhadap data yang tidak relevan, seperti komentar kosong, komentar berupa stiker atau GIF, serta komentar yang hanya berisi tanda baca, diperoleh dataset akhir sebanyak 3.227 komentar yang digunakan pada tahap penelitian selanjutnya.

1) *Cleaning Text*, yaitu proses menghapus tautan (URL), mention akun (@username), tanda pagar (#), tag HTML, angka, emoji, dan tanda baca dari teks komentar menggunakan pencocokan pola berbasis ekspresi reguler (regex).

2) *Hapus Noise*, yaitu membuang komentar yang tidak mengandung informasi bermakna, seperti komentar kosong, komentar berupa kata “sticker”, “stiker”, “gif”, “video”, atau yang hanya terdiri atas tanda baca.

3) *Case Folding*, yaitu penyeragaman seluruh huruf pada teks menjadi huruf kecil (lowercase) untuk menjaga konsistensi penulisan.

4) *Tokenization*, yaitu proses memecah teks komentar menjadi satuan kata (token) menggunakan fungsi `word_tokenize` dari pustaka Natural Language Toolkit (NLTK) [6].

5) *Normalization*, yaitu proses mengubah kata gaul, singkatan, atau kata tidak baku yang lazim digunakan pada komentar media sosial (misalnya “gak”, “yg”, “dr”, “udh”) menjadi bentuk kata baku (“tidak”, “yang”, “dari”, “sudah”) berdasarkan kamus normalisasi yang disusun secara manual.

6) *Stopword Removal*, yaitu penghapusan kata-kata umum yang tidak memiliki makna signifikan terhadap sentimen, menggunakan `StopWordRemoverFactory` dari pustaka Sastrawi [7]. Kata negasi seperti “tidak”, “tak”, “bukan”, “belum”, dan “jangan” sengaja dipertahankan dan tidak dihapus dari daftar stopwords, karena kata-kata tersebut berperan penting dalam menentukan polaritas sentimen suatu kalimat.

7) *Stemming*, yaitu proses mengubah kata berimbuhan menjadi kata dasar menggunakan Stemmer dari pustaka Sastrawi [7], yang menerapkan algoritma confix-stripping berbasis algoritma Nazief dan Adriani.

Setelah seluruh tahapan preprocessing selesai, token hasil stemming digabungkan kembali menjadi satu kesatuan kalimat (stemmed text) sebagai representasi akhir setiap komentar yang siap digunakan pada tahap ekstraksi fitur.

C. Pelabelan Sentimen

Oleh karena data komentar tidak memiliki label sentimen secara langsung, penelitian ini menerapkan pelabelan otomatis berbasis kamus (lexicon-based labeling). Disusun dua daftar kata acuan, yaitu daftar kata positif yang memuat istilah-istilah yang mencerminkan dukungan atau optimisme terhadap pemekaran, seperti “setuju”, “dukung”, “mantap”, “sepakat”, dan “pemekaran”, serta daftar kata negatif yang memuat istilah penolakan, kekecewaan, atau kritik, seperti “tolak”, “gagal”, “korupsi”, dan “kecewa”. Setiap token pada komentar yang telah melalui proses stemming dicocokkan dengan kedua daftar kata tersebut, kemudian jumlah kemunculan kata positif dan kata negatif pada setiap komentar dibandingkan. Setiap komentar selanjutnya diberi label sentimen sebagai berikut:

1) 0. Negatif, apabila jumlah kata negatif lebih banyak daripada kata positif.

2) 1. Netral, apabila jumlah kata positif dan negatif sama, termasuk ketika tidak ditemukan kata acuan pada komentar tersebut.

3) 2. Positif, apabila jumlah kata positif lebih banyak daripada kata negatif.

Melalui proses ini, diperoleh komposisi data sebanyak 2.196 komentar berlabel netral (1), 868 komentar berlabel negatif (0), dan 163 komentar berlabel positif (2).

D. TF-IDF

Data teks hasil preprocessing selanjutnya diubah menjadi representasi numerik menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) melalui *TfidfVectorizer* pada pustaka *scikit-learn* [8]. Metode ini memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam suatu komentar (term frequency) serta seberapa jarang kata tersebut muncul pada keseluruhan komentar (inverse document frequency), sehingga kata yang sering muncul pada satu komentar namun jarang muncul di komentar lain akan memperoleh bobot yang lebih tinggi [1]. Perhitungan TF-IDF dirumuskan pada Persamaan (1) hingga (3).

$$TF(t,d) = f(t,d) / \sum f(t,d) \quad (1)$$

$$IDF(t) = \log(N / df(t)) \quad (2)$$

$$TF-IDF(t,d) = TF(t,d) \times IDF(t) \quad (3)$$

dengan t menyatakan suatu term (kata atau frasa), d menyatakan dokumen (komentar) tempat term tersebut muncul, $f(t,d)$ menyatakan jumlah kemunculan t pada d , N menyatakan jumlah total dokumen dalam korpus, dan $df(t)$ menyatakan jumlah dokumen yang mengandung term t . Pada penelitian ini, ekstraksi fitur dilakukan dengan parameter *max features* sebesar 5.000, rentang *n-gram* (1,2) sehingga fitur berupa kombinasi unigram dan bigram, serta *min df* sebesar 3 untuk mengabaikan term yang muncul pada kurang dari tiga dokumen agar mengurangi fitur yang jarang muncul dan tidak representatif. Proses ini menghasilkan matriks fitur berukuran 3.227 baris (jumlah komentar) dan 1.710 kolom (jumlah fitur kata/frasa unik) yang selanjutnya digunakan sebagai input pada tahap klasifikasi.

E. Train-Test Split

Dataset yang telah direpresentasikan dalam bentuk matriks TF-IDF kemudian dibagi menjadi data latih dan data uji menggunakan fungsi *train_test_split* pada pustaka *scikit-learn* [6], dengan proporsi 80% data latih dan 20% data uji. Pembagian data dilakukan secara stratifikasi (*stratify*) berdasarkan label sentimen untuk menjaga proporsi masing-masing kelas antara data latih dan data uji tetap seimbang, mengingat distribusi kelas pada dataset tidak merata. Nilai *random_state* sebesar 42 digunakan untuk menjamin bahwa hasil pembagian data bersifat konsisten dan dapat direproduksi pada pengujian berikutnya. Melalui skema ini, diperoleh 2.581 data latih dan 646 data uji.

F. Training SVM

Proses klasifikasi sentimen dilakukan menggunakan algoritma *Support Vector Machine* (SVM) dengan implementasi *LinearSVC* pada pustaka *scikit-learn* [6]. SVM bekerja dengan mencari *hyperplane* pemisah terbaik yang mampu memaksimalkan margin antar kelas data pada ruang fitur berdimensi tinggi [1]. Untuk kasus klasifikasi

tiga kelas sentimen (negatif, netral, dan positif) pada penelitian ini, *LinearSVC* menerapkan strategi *one-vs-rest*, yaitu membangun satu model pemisah untuk setiap kelas terhadap seluruh kelas lainnya. Persamaan umum *hyperplane* pemisah dan fungsi objektif SVM dirumuskan pada Persamaan (4) dan (5).

$$w^T x + b = 0 \quad (4)$$

$$\min (1/2) \|w\|^2, \quad y_i(w^T x_i + b) \geq 1 \quad (5)$$

dengan w menyatakan vektor bobot yang menentukan orientasi *hyperplane*, x menyatakan vektor fitur hasil ekstraksi TF-IDF, b menyatakan bias atau *intercept hyperplane*, serta x_i dan y_i masing-masing menyatakan vektor fitur dan label kelas dari data ke- i . Model dilatih menggunakan data latih (*training set*) dengan parameter *random_state* sebesar 42 agar hasil pelatihan model dapat direproduksi pada pengujian berikutnya.

G. Evaluasi

Evaluasi dilakukan untuk mengevaluasi kemampuan model SVM untuk mengkategorikan sentimen positif dan negatif terhadap data komentar pembentukan provinsi Luwu Raya di platform tiktok [9]. Proses evaluasi akan membagi data rasio 80% sebagai data *train* dan 20% sebagai data *test*. Model akan di uji oleh empat metrik utama, akurasi, presisi, *recall*, dan *f1-score*. Empat metrik digunakan untuk menilai seberapa model bisa mengidentifikasi kelas sentimen dengan benar [10] (Persamaan 6,7,8,9).

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (6)$$

$$Precision = TP / (TP + FP) \quad (7)$$

$$Recall = TP / (TP + FN) \quad (8)$$

$$F1-Score = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (9)$$

dengan *True Positive* (TP) menyatakan jumlah data yang diprediksi benar sesuai kelas sebenarnya, *True Negative* (TN) menyatakan jumlah data bukan kelas tersebut yang diprediksi benar, *False Positive* (FP) menyatakan jumlah data yang salah diprediksi sebagai kelas tersebut, dan *False Negative* (FN) menyatakan jumlah data kelas tersebut yang salah diprediksi sebagai kelas lain.

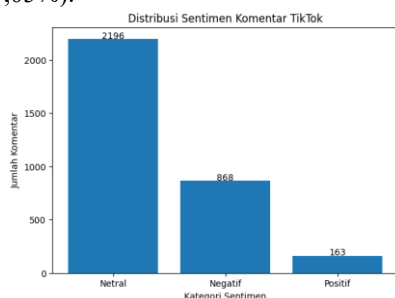
Selain evaluasi pada satu skema pembagian data, penelitian ini juga menerapkan validasi silang (*k-fold cross validation*) dengan $k=5$ menggunakan fungsi *cross_val_score* pada pustaka *scikit-learn* [6], guna menguji konsistensi dan kestabilan kinerja model pada subset data yang berbeda-beda serta memastikan bahwa hasil evaluasi tidak bergantung pada satu skema pembagian data tertentu.

III. HASIL DAN PEMBAHASAN

A. Distribusi Sentimen Komentar TikTok

Proses pelabelan otomatis berbasis kamus (*lexicon-based labeling*) terhadap 3.227 komentar TikTok terkait wacana pembentukan Provinsi Luwu Raya menghasilkan distribusi tiga kelas sentimen sebagaimana ditunjukkan pada Gambar 2. Sentimen netral mendominasi dengan 2.196 komentar (68,05%), diikuti sentimen negatif sebanyak 868

komentar (26,89%), dan sentimen positif sebanyak 163 komentar (5,05%).

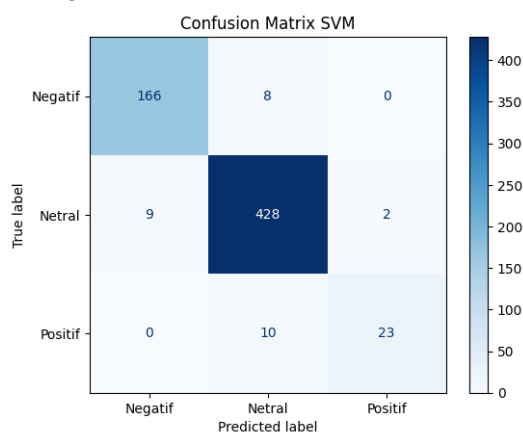


Gambar 2. Distribusi Sentimen Komentar TikTok terhadap Pembentukan Provinsi Luwu Raya

Dominasi sentimen netral mengindikasikan bahwa sebagian besar warganet menyampaikan komentar yang bersifat informatif, deskriptif, atau sekadar menanggapi pemberitaan tanpa menyatakan sikap eksplisit terhadap pemekaran. Sementara itu, proporsi komentar bersentimen negatif yang jauh lebih besar dibandingkan komentar positif menunjukkan bahwa kritik, kekecewaan, dan keluhan masyarakat (terutama terkait dampak aksi demonstrasi dan pemblokiran jalan) lebih banyak disuarakan secara terbuka di kolom komentar dibandingkan dukungan eksplisit terhadap wacana pembentukan Provinsi Luwu Raya. Kondisi ini sekaligus menunjukkan bahwa dataset penelitian bersifat tidak seimbang (imbalanced), khususnya pada kelas positif, yang perlu diperhatikan dalam menginterpretasikan kinerja model klasifikasi.

B. Hasil Klasifikasi Sentimen Menggunakan SVM

Model SVM yang telah dilatih pada 2.581 data latih selanjutnya diuji pada 646 data uji. Hasil prediksi model divisualisasikan dalam bentuk confusion matrix pada Gambar 3



Gambar 3. Confusion Matrix Hasil Klasifikasi SVM

Berdasarkan Gambar 3, dari 174 komentar berlabel negatif pada data uji, model berhasil mengklasifikasikan 166 komentar dengan benar dan hanya salah mengklasifikasikan 8 komentar sebagai netral. Pada kelas netral, dari 439 komentar, model berhasil memprediksi 428 komentar dengan benar, dengan 9 komentar salah diprediksi sebagai negatif dan 2 komentar sebagai positif. Adapun pada kelas positif, dari 33 komentar, model hanya berhasil mengklasifikasikan 23 komentar dengan benar, sedangkan

10 komentar lainnya salah diprediksi sebagai netral. Secara rinci, kinerja model pada setiap kelas ditunjukkan pada Tabel 1.

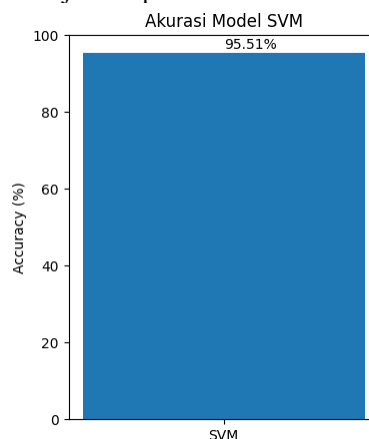
TABEL 1. HASIL EVALUASI KLASIFIKASI SVM PER KELAS

Kelas	Precision	Recall	F1-Score	Support
Negatif (0)	0.95	0.95	0.95	174
Netral (1)	0.96	0.97	0.97	439
Positif (2)	0.92	0.70	0.79	33
Accuracy			0.96	646
Macro avg	0.94	0.88	0.90	646
Weighted avg	0.95	0.96	0.95	646

Tabel 1 menunjukkan bahwa model memperoleh precision dan recall yang tinggi pada kelas negatif (0,95 dan 0,95) dan kelas netral (0,96 dan 0,97), namun recall pada kelas positif relatif lebih rendah, yaitu sebesar 0,70. Rendahnya recall pada kelas positif ini sejalan dengan proporsi data yang jauh lebih sedikit pada kelas tersebut (hanya 163 dari 3.227 komentar, atau sekitar 33 data pada data uji), sehingga model cenderung lebih sulit mengenali pola pada kelas minoritas dibandingkan dua kelas lainnya yang memiliki jumlah data jauh lebih banyak.

C. Akurasi Model SVM

Secara keseluruhan, model SVM dengan ekstraksi fitur TF-IDF memperoleh akurasi sebesar 95,51% pada data uji, sebagaimana ditunjukkan pada Gambar 4.



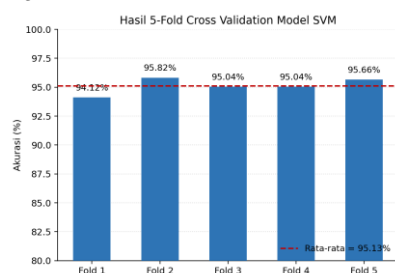
Gambar 4. Hasil Akurasi Model SVM

Nilai akurasi ini menunjukkan bahwa kombinasi metode TF-IDF dan SVM mampu [1], [5] mengklasifikasikan sentimen komentar TikTok terkait wacana pembentukan Provinsi Luwu Raya dengan tingkat ketepatan yang tinggi. Hasil ini juga memperkuat temuan penelitian terdahulu yang menunjukkan efektivitas kombinasi TF-IDF dan SVM pada berbagai konteks analisis sentimen media sosial berbahasa Indonesia, meskipun akurasi yang diperoleh pada penelitian ini relatif lebih tinggi, salah satunya dipengaruhi oleh dominannya kelas netral yang lebih mudah dikenali polanya oleh model.

D. Hasil Cross Validation

Untuk menguji konsistensi dan kestabilan kinerja model pada subset data yang berbeda-beda, dilakukan validasi silang (5-fold cross validation) menggunakan seluruh data

fitur TF-IDF. Hasil pengujian pada setiap *fold* ditunjukkan pada Gambar 5.



Gambar 5. Hasil 5-Fold Cross Validation Model SVM

Hasil *cross validation* menunjukkan akurasi yang stabil pada setiap fold, berkisar antara 94,12% hingga 95,82%, dengan rata-rata akurasi sebesar 95,13%. Nilai rata-rata *cross validation* ini sejalan dan tidak berbeda jauh dengan akurasi pada data uji tunggal (95,51%), yang mengindikasikan bahwa model SVM yang dibangun tidak mengalami *overfitting* dan mampu melakukan generalisasi dengan baik terhadap data komentar TikTok yang belum pernah dilihat sebelumnya.

E. Analisis Word Cloud Sentimen Positif dan Negatif

Untuk memperoleh gambaran kualitatif mengenai topik yang paling banyak dibicarakan pada masing-masing kelas sentimen, dilakukan visualisasi *word cloud* terhadap komentar berlabel positif dan negatif, sebagaimana ditunjukkan pada Gambar 6 dan Gambar 7.



Gambar 6. Word Cloud Komentar Bersentimen Positif

Pada Gambar 6, kata-kata yang paling menonjol pada komentar bersentimen positif antara lain “majulah”, “provinsi”, “mekar”, “mantap”, “dukung”, “prioritas”, dan “luwu raya”. Kemunculan kata-kata tersebut mencerminkan harapan dan optimisme sebagian warganet terhadap pembentukan Provinsi Luwu Raya, khususnya terkait harapan percepatan pembangunan, pemerataan akses, dan perbaikan tata kelola pemerintahan di wilayah tersebut.



Gambar 7. Word Cloud Komentar Bersentimen Negatif

Sebaliknya, pada Gambar 7, kata-kata yang dominan pada komentar bersentimen negatif antara lain “demo”, “tidak”, “jalan”, “tutup”, “logistik”, “ambulance”, dan “susah”. Pola ini menunjukkan bahwa sentimen negatif sebagian besar tidak diarahkan pada gagasan pembentukan provinsi itu sendiri, melainkan lebih banyak berupa kritik dan keluhan terhadap dampak aksi demonstrasi dan pemblokiran Jalan Trans Sulawesi, seperti terganggunya distribusi logistik, kelangkaan bahan bakar, serta terhambatnya akses kendaraan darurat termasuk ambulans. Menariknya, kata “ambulance” muncul pada kedua word cloud dengan konteks yang berbeda, pada komentar positif dikaitkan dengan harapan perbaikan fasilitas kesehatan pascapemekaran, sedangkan pada komentar negatif dikaitkan dengan keluhan terhambatnya layanan darurat akibat blokade jalan.

F. Pembahasan

Secara umum, hasil penelitian ini menunjukkan bahwa kombinasi metode TF-IDF dan SVM efektif digunakan untuk mengklasifikasikan sentimen masyarakat terhadap isu pemerintahan daerah, khususnya wacana pembentukan Provinsi Luwu Raya, berdasarkan data komentar TikTok yang tidak terstruktur. Akurasi sebesar 95,51% serta hasil *cross validation* yang konsisten pada kisaran 94–96% memperkuat temuan penelitian terdahulu mengenai keandalan SVM dengan pembobotan fitur TF-IDF pada data teks berbahasa Indonesia [1], [5], [6], [10] sekaligus menjadi kontribusi baru karena diterapkan pada konteks isu pemekaran daerah yang sebelumnya belum banyak dikaji.

Meskipun demikian, hasil evaluasi per kelas pada Tabel 1 mengungkap keterbatasan model dalam mengenali kelas positif, yang tercermin dari recall sebesar 0,70. Keterbatasan ini bersumber dari ketidakseimbangan jumlah data pada tahap pelabelan, di mana komentar bersentimen positif jumlahnya jauh lebih sedikit dibandingkan kelas netral dan negatif. Pada praktiknya, hal ini juga mencerminkan realitas di lapangan, yaitu bahwa opini publik yang tersaring melalui kolom komentar video pemberitaan cenderung lebih didominasi oleh komentar netral (informatif) dan kritis (negatif) dibandingkan dukungan eksplisit.

Temuan pada analisis word cloud turut memperkaya interpretasi hasil klasifikasi, dengan menunjukkan bahwa sentimen negatif pada dataset ini sebagian besar dipicu oleh dampak teknis aksi demonstrasi (pemblokiran jalan, gangguan logistik dan layanan darurat), bukan semata-mata penolakan terhadap gagasan pemekaran itu sendiri. Sementara itu, sentimen positif lebih banyak berkaitan dengan harapan pembangunan dan pemerataan akses. Implikasinya, para pemangku kepentingan, baik tim pengusul pemekaran maupun pemerintah daerah, perlu mempertimbangkan aspek teknis penyampaian aspirasi masyarakat, di samping substansi wacana pemekaran, dalam menyusun strategi komunikasi publik ke depan. Penelitian selanjutnya juga disarankan untuk menerapkan teknik penyeimbangan data (seperti oversampling) guna meningkatkan kemampuan model dalam mengenali kelas sentimen minoritas..

IV. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dan pembahasan yang telah diuraikan, dapat disimpulkan beberapa hal sebagai berikut. Pertama, kombinasi metode TF-IDF sebagai teknik ekstraksi fitur dan *Support Vector Machine* (SVM) sebagai algoritma klasifikasi terbukti efektif digunakan untuk mengklasifikasikan sentimen masyarakat terhadap wacana pembentukan Provinsi Luwu Raya berdasarkan 3.227 komentar TikTok, dengan capaian akurasi sebesar 95,51% pada data uji. Kedua, hasil validasi silang (5-fold cross validation) menunjukkan akurasi yang stabil pada kisaran 94,12%–95,82% dengan rata-rata 95,13%, yang sejalan dengan akurasi pada data uji tunggal sehingga mengindikasikan model tidak mengalami overfitting dan mampu melakukan generalisasi dengan baik. Ketiga, distribusi sentimen menunjukkan bahwa opini masyarakat pada kolom komentar TikTok didominasi oleh sentimen netral (68,05%), diikuti sentimen negatif (26,89%), dan sentimen positif (5,05%), yang mengindikasikan bahwa sebagian besar warganet cenderung menyampaikan komentar yang bersifat informatif dan kritis dibandingkan dukungan eksplisit. Keempat, hasil evaluasi per kelas menunjukkan model memiliki kinerja yang tinggi dalam mengenali kelas negatif dan netral, namun kinerja pada kelas positif masih terbatas (recall 0,70) akibat jumlah data pada kelas tersebut yang jauh lebih sedikit (data tidak seimbang/*imbalanced*). Kelima, analisis word cloud memperkuat interpretasi hasil klasifikasi dengan menunjukkan bahwa sentimen negatif sebagian besar dipicu oleh dampak teknis aksi demonstrasi, seperti terganggunya logistik, kelangkaan bahan bakar, dan terhambatnya akses layanan darurat, bukan semata-mata penolakan terhadap gagasan pemekaran Provinsi Luwu Raya itu sendiri. Secara keseluruhan, penelitian ini berhasil menjawab tujuan penelitian sekaligus mengisi kesenjangan penelitian (research gap) yang diuraikan pada bagian Pendahuluan, yaitu belum tersedianya kajian analisis sentimen berbasis TF-IDF dan SVM yang secara khusus membahas wacana pembentukan Provinsi Luwu Raya pada data media sosial TikTok.

Berdasarkan keterbatasan yang ditemukan, penelitian ini mengajukan beberapa saran bagi penelitian selanjutnya. Pertama, perlu diterapkan teknik penyeimbangan data (*data balancing*), seperti *oversampling* atau SMOTE, untuk meningkatkan kemampuan model dalam mengenali kelas sentimen minoritas, khususnya kelas positif. Kedua, proses pelabelan sentimen pada penelitian ini masih bersifat otomatis berbasis kamus (*lexicon-based*), sehingga penelitian selanjutnya disarankan untuk melibatkan validasi oleh pakar (*expert annotation*) atau kombinasi dengan pendekatan berbasis pembelajaran mendalam (*deep learning*), seperti IndoBERT, guna meningkatkan akurasi pelabelan. Ketiga, cakupan data dapat diperluas dengan menggabungkan sumber data dari

platform media sosial lain, seperti X (Twitter), Instagram, atau YouTube, untuk memperoleh gambaran opini publik yang lebih komprehensif. Keempat, penelitian selanjutnya dapat membandingkan kinerja SVM dengan algoritma klasifikasi lain, seperti Random Forest, Naïve Bayes, atau model berbasis deep learning, guna memperoleh model dengan performa yang lebih optimal. Kelima, secara praktis, hasil penelitian ini dapat menjadi bahan pertimbangan bagi tim pengusul pemekaran maupun pemerintah daerah dalam menyusun strategi komunikasi publik, khususnya dalam meminimalkan dampak teknis dari aksi penyampaian aspirasi masyarakat agar tidak menimbulkan persepsi negatif yang sesungguhnya tidak berkaitan langsung dengan substansi wacana pembentukan Provinsi Luwu Raya.

DAFTAR PUSTAKA

- [1] Yaskur Bearly Fernandes, Elin Haerani, Fadhilah Syafria, Muhammad Fikry, and Lola Oktavia, "Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pada Komentar Bitcoin Di Aplikasi X," *Bulletin of Computer Science Research*, vol. 6, no. 1, pp. 381–391, Dec. 2025, doi: 10.47065/bulletincsr.v6i1.928.
- [2] ANTARA News Makassar, "Ini alasan warga desak pembentukan Provinsi Luwu Raya," <https://makassar.antaranews.com/berita/620602/ini-alasan-warga-desak-pembentukan-provinsi-luwu-raya>.
- [3] H. Jalante, "The Formation of Greater Luwu Province and the Future of South Sulawesi: Empirical Analysis, Fiscal Capacity, and Contemporary Socio-Spatial Perspectives," *Jurnal Pallangga Praja (JPP)*, vol. 8, no. 1, pp. 79–91, May 2026, doi: 10.61076/jpp.v8i1.6318.
- [4] detikSusel, "Pemprov Sulsel Tegaskan Pemekaran Luwu Raya Kewenangan Pemerintah Pusat," <https://www.detik.com/sulsel/berita/d-8331667/pemprov-sulsel-tegaskan-pemekaran-luwu-raya-kewenangan-pemerintah-pusat>.
- [5] J. E. Br Sinulingga and H. C. K. Sitorus, "Analisis Sentimen Opini Masyarakat terhadap Film Horor Indonesia Menggunakan Metode SVM dan TF-IDF," *Jurnal Manajemen Informatika (JAMIKA)*, vol. 14, no. 1, pp. 42–53, Feb. 2024, doi: 10.34010/jamika.v14i1.11946.
- [6] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python: analyzing text with the natural language toolkit*. "O'Reilly Media, Inc.," 2009.
- [7] M. Adriani, J. Asian, B. Nazief, S. M. M. Tahaghoghi, and H. E. Williams, "Stemming Indonesian: A confix-stripping approach," *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 6, no. 4, pp. 1–33, 2007.
- [8] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.
- [9] Salsabila Dwi Fitri, Dewi Lestari, Rizqa Raaiqa Bintana, Reni Aryani, Mohamad Ilhami, and Yolla Noverina, "Implementasi Model Support Vector Machine Dalam Analisa Sentimen Masyarakat Mengenai Kebijakan Penerapan Aplikasi MyPertamina," *Bridge : Jurnal publikasi Sistem Informasi dan Telekomunikasi*, vol. 2, no. 2, pp. 176–193, Aug. 2024, doi: 10.62951/bridge.v2i3.180.
- [10] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 1, p. 147, Feb. 2021, doi: 10.25126/jtiik.0813944.