

# KLASIFIKASI SENTIMEN DAN PEMETAAN EMOSI BERBASIS LEKSIKON TERHADAP KETIDAK PERCAYAAN DPR RI DI MEDIA SOSIAL X MENGGUNAKAN METODE SVM

Risky Riswanto<sup>1\*</sup>, Irfan Pratama<sup>2</sup>

<sup>1</sup>Universitas Mercu Buana Yogyakarta, Ngroropoh, Condongcatur, Kecamatan. Depok Kabupaten. Sleman, Daerah Istimewa Yogyakarta 55281

## Keywords:

Analisis Sentimen, Emotional Lexicon, Lexicon Based, Support Vector Machine, TF IDF.

## Correspondent Email:

wanto.r19@gmail.com

**Abstrak.** Penelitian ini menganalisis karakteristik emosi masyarakat serta menguji performa model klasifikasi *Support Vector Machine* (SVM) berbasis leksikon (*Emotional Lexicon*) dalam memetakan sepuluh dimensi emosi terkait fenomena mosi tidak percaya terhadap DPR RI pasca-demonstrasi Agustus 2025 di media sosial X. Menggunakan 19.454 data opini publik, dataset terbagi ke dalam Sentimen Negatif (8.882 data), Sentimen Netral (5.968 data), dan Sentimen Positif (4.604 data). Pada pengujian polaritas sentimen (Layer 1), algoritma SVM berbasis SMOTE mencapai *Test Accuracy* sebesar 85,45%. Sementara pada klasifikasi emosi multikelas (Layer 2) dengan pembagian data split 80:20, model SVM menghasilkan *Test Accuracy* sebesar 75,53%, dengan performa tertinggi pada emosi Netral (*F1-score* 0,88) dan Marah (*F1-score* 0,82). Hasil analisis menunjukkan bahwa sentimen negatif didominasi secara masif oleh emosi Marah (5.488 data) dan Kecewa (873 data) sebagai manifestasi kemarahan kolektif. Kategori Netral mencerminkan ketidakpuasan pasif berdiskusi formal, sementara visualisasi *word cloud* menegaskan kata "Rakyat", "DPRRI", dan "Bubar" sebagai poros utama tuntutan publik. Meski demikian, metode *Emotional Lexicon* (*EmoLex*) masih memiliki keterbatasan dalam mendeteksi sarkasme politik secara akurat. Penelitian ini menyarankan pemanfaatan arsitektur *Deep Learning* atau *Transformer* serta kernel non-linear untuk pengembangan klasifikasi emosi selanjutnya.



Copyright © 2026 The Author(s), Published by Jurnal Informatika dan Teknik Elektro Terapan. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

**Abstract.** This research analyzes the emotional characteristics of society and tests the performance of the lexicon-based Support Vector Machine (SVM) classification model (*Emotional Lexicon*) in mapping ten emotional dimensions related to the phenomenon of no confidence in the DPR RI after the August 2025 demonstration on social media. Using 19,454 public opinion data, the dataset is divided into Negative Sentiment (8,882 data), Neutral Sentiment (5,968 data), and Positive Sentiment (4,604 data). In sentiment polarity testing (Layer 1), the SMOTE-based SVM algorithm achieved *Test Accuracy* of 85.45%. Meanwhile, in the multiclass emotion classification (Layer 2) with a split data division of 80:20, the SVM model produced *Test Accuracy* of 75.53%, with the highest performance in Neutral (*F1-score* 0.88) and Angry (*F1-score* 0.82) emotions. The results of the analysis show that negative sentiment is dominated massively by emotions of Anger (5,488 data) and Disappointment (873 data) as manifestations of collective anger. The Neutral category reflects passive dissatisfaction with formal diction, while word cloud visualizations emphasize the words "People", "DPRRI", and "Dissolve" as the main axes of public demand. However, the Emotional

*Lexicon (EmoLex) method still has limitations in accurately detecting political sarcasm. This research suggests utilizing the Deep Learning or Transformer architecture as well as the non-linear kernel for the development of subsequent emotional classifications.*

## 1. PENDAHULUAN

Dewan Perwakilan Rakyat (DPR) merupakan lembaga negara lingkup politik hukum dan perundang-undangan yang saat ini menghadapi gelombang ketidakpercayaan publik yang masif [1]. Fenomena aksi demonstrasi Agustus 2025 yang dipicu oleh isu kenaikan tunjangan anggota DPR serta beredarnya informasi kunjungan kerja keluar negeri di tengah situasi krusial, memicu lonjakan opini kritis di media sosial X melalui tagar seperti #bubarkandpr dan #resetindonesia [2], [3], [4]. Volume data yang sangat besar (*Big Data*) dari diskursus mosi tidak percaya ini merekam krisis kepercayaan publik yang mendalam [5]. Namun, analisis sentimen tradisional yang hanya mengklasifikasikan teks ke dalam kategori positif, negatif, atau netral gagal menangkap akar emosional spesifik dibalik kritik tersebut, sehingga menyulitkan pemetaan esensi keluhan publik secara akurat.

Ketidakmampuan membedakan apakah sentimen negatif tersebut berakar dari emosi marah, kecewa, atau jijik menjadi batasan utama dalam analisis teks politik tradisional. Terlebih lagi, opini di media sosial X sering kali menggunakan ragam diksi formal atau gaya bahasa sarkasme yang menyamarkan emosi negatif ke dalam bentuk teks terstruktur, sehingga berpotensi salah teridentifikasi oleh sistem standar.

Untuk mengatasi keterbatasan tersebut, penelitian ini mengusulkan penerapan pendekatan *Natural Language Processing* (NLP) menggunakan *Emotional Lexicon* (NRC Emotion Lexicon/Emolex) untuk mengekstraksi dan memetakan sepuluh dimensi emosi politik secara otomatis [6], [7]. Kebaruan dari penelitian ini terletak pada integrasi analisis sentimen dan emosi leksikon dalam satu kerangka kerja diskursus politik [8], [9]. Pendekatan ini tidak hanya mengukur kuantitas ketidakpercayaan publik terhadap DPR di media sosial X, tetapi juga mengungkap spektrum emosi spesifik di balik polaritas sentimen tersebut secara lebih komprehensif.

## 2. TINJAUAN PUSTAKA

Penelitian mengenai analisis sentimen dalam lingkup politik hukum dan evaluasi kerja negara memperlihatkan urgensi yang sangat tinggi di era digital [1] mengkaji representasi politik hukum Dewan Perwakilan Rakyat (DPR) dengan memetakan opini publik di X menggunakan algoritma *Naïve Bayes*, yang menghasilkan tingkat akurasi sebesar 80% dari total 1.546 data *crawling*. Fenomena momentum krisis politik tunggal, seperti aksi demonstrasi massa Agustus 2025 yang memicu gelombang mosi tidak percaya publik terhadap tubuh DPR RI, umumnya melibatkan luapan psikologis yang jauh lebih kompleks. Gerakan demonstrasi yang membawa 17+8 tuntutan rakyat tersebut mencerminkan puncak kemarahan masyarakat atas fungsi representasi yang dinilai gagal [5]. Lini masa mengenai eskalasi gerakan ini terekam sejak aksi unjuk rasa geopolitik yang puncaknya terjadi pada Demo Agustus 2025 [8], Krisis kepercayaan publik yang mendalam ini terekam secara masif sebagai data berskala besar (*Big Data*) di media sosial X yang memerlukan reformasi mendalam [5].

Untuk memetakan data berskala besar tersebut, penggunaan metode pembelajaran mesin terbimbing (*supervised learning*) yang tangguh sangat diperlukan. Sebagai rujukan utama dalam penerapan algoritma berbasis geometri pada ulasan digital [9] melakukan analisis sentimen terhadap ulasan aplikasi DANA menggunakan algoritma *Support Vector Machine* (SVM). Melalui pengumpulan data sebanyak 10.000 ulasan berbasis *web scraping*, penelitian tersebut menerapkan pra-pemrosesan teks yang meliputi *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*, serta menggunakan pendekatan *lexicon-based* untuk pelabelan otomatis ke dalam kategori positif, negatif, dan netral yang menghasilkan tingkat akurasi tertinggi sebesar 94,2% [9]. Penyelarasan metodologi ini juga diperkuat oleh temuan [10] yang menganalisis respons emosional pada interaksi digital politik

menggunakan kombinasi fitur TF-IDF, yang membuktikan bahwa pemetaan emosi mampu memberikan akurasi tinggi meskipun terdapat bias atensi penonton pada sesi komunikasi politik tertentu.

Untuk mengatasi keterbatasan dimensi polaritas tradisional (positif, negatif, netral) seperti pada penelitian [5], [9], pemanfaatan kamus leksikon emosi mulai dikembangkan dalam pemrosesan bahasa alami (*Natural Language Processing*). [6], [7] berhasil menerapkan pendekatan berbasis *NRC Emotion Lexicon* (EmoLex) untuk mengeksplorasi kondisi emosi multikelas dari komentar publik di media sosial secara otomatis.

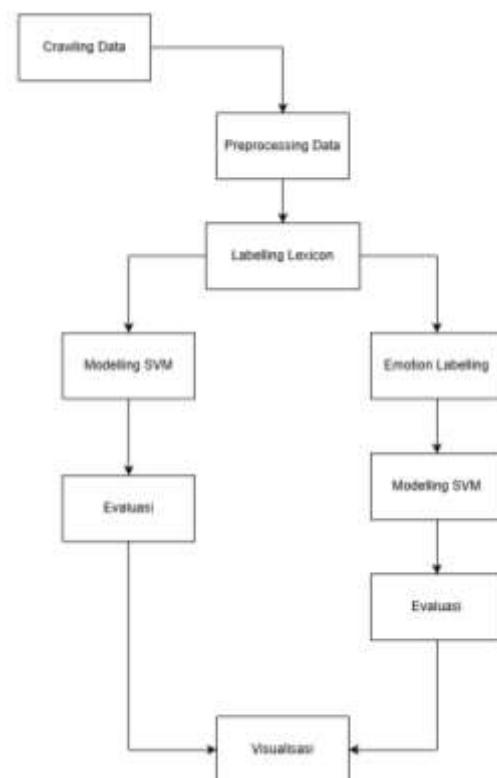
Kamus *EmoLex* ini dikembangkan pertama kali oleh [11] dengan total berisi 14.182 kata (*unigram*) serta menggunakan skema *association score* bernilai biner [12]. Penggunaan kamus leksikon ini memegang peran krusial dalam mengidentifikasi dimensi emosional serta kekuatan variasi emosi tersemat di dalam kata-kata opini masyarakat [13]. Walaupun demikian, tantangan utama pada teks opini berskala besar adalah penggunaan ragam gaya bahasa sarkasme politik terstruktur yang kerap mengecoh instrumen leksikon standar [6], [14]. Oleh karena itu, ketangguhan algoritma klasifikasi berbasis geometri seperti SVM sangat di butuhkan karena karakteristik pembentukan bidang pemisah optimal (*hyperplane*) miliknya terbukti efisien meminimalkan risiko *overfitting* pada klasifikasi teks berdimensi tinggi [4].

Integrasi pendekatan *lexicon-based* EmoLex dan algoritma SVM menciptakan sinergi metodologis yang tangguh dalam mengatasi ambiguitas dan derau bahasa, seperti sarkasme politik di media sosial. Melalui pemetaan biner EmoLex, teks opini publik tidak terstruktur ditransformasikan menjadi matriks fitur emosi multikelas yang spesifik dan mendalam. Dimensi fitur yang kompleks dan berdimensi tinggi (*sparse data*) tersebut kemudian diselesaikan secara optimal oleh SVM melalui pencarian bidang pemisah terbaik (*hyperplane*) untuk meminimalkan risiko *overfitting*. Penggabungan kedua pendekatan ini menjadi fondasi utama dalam penelitian ini untuk mengklasifikasikan 10 dimensi emosi publik secara akurat terkait isu ketidakpercayaan terhadap DPR RI.

Permasalahan ketidakseimbangan data (*data imbalance*) pada data media sosial X dapat diatasi menggunakan teknik *resampling Minority Oversampling Technique* (SMOTE). Penerapan SMOTE terbukti mampu mendongkrak metrik *Precision*, *Recall*, dan *F1-Score* secara signifikan, khususnya pada kelas minoritas [15]. Sinergi antara *EmoLex*, SMOTE, dan SVM menghasilkan kerangka metodologi yang presisi untuk pelabelan sentimen dan pemetaan emosi publik.

### 3. METODE PENELITIAN

Metode penelitian ini berfokus pada pendeteksi emosi publik selama periode deo Dewan Perwakilan Rakyat (DPR) 25 Agustus 2025 [9]. Penelitian ini menggunakan pendekatan kuantitatif menerapkan teknik Natural Language Processing (NLP) melalui metode hybrid. Diagram alur metodologi dapat di lihat pada Gambar 1.



Gambar 1. Metode Penelitian

### 3.1. Pengumpulan Data

Pengumpulan data yang dilakukan menggunakan teknik *crawling data* menggunakan *tweet harvest* dengan bahasa pemrograman *Python*. Data yang diambil difokuskan pada percakapan publik di media sosial X yang merefleksikan mosi tidak percaya serta krisis ketidakpercayaan terhadap lembaga legislatif, dengan menargetkan tagar-tagar (#) spesifik seperti #bubarkandpr, #resetindonesia, #ruuperampasanaset, dan #reformasidpr pada rentang waktu 25 Agustus hingga 30 Agustus 2025. Selain itu, untuk memperkaya cukupan analisis, pengumpulan data ini dilengkapi dengan data sekunder dari platform *Kaggle*, sehingga menghasilkan total 19454 data.

### 3.2. Preprocessing Data

*Preprocessing Data* bertujuan untuk menyiapkan data sebelum pemrosesan lebih jauh dilakukan secara umum, data *preprocessing* ini dilakukan untuk menghilangkan karakter-karakter serta kata atau simbol-simbol yang tidak diperlukan dalam proses analisis. Tahapan *preprocessing data* mengadopsi langkah-langkah berikut [16].

#### 3.2.1. Cleaning

*Cleaning* bertujuan pembersihan data melibatkan penggunaan sub fungsi *library* untuk membersihkan tanda baca, angka, simbol, ASCII, dan tautan. Sub fungsi ini bekerja dengan mengganti *string* atau atribut kosong untuk *string* target berdasarkan pola yang ditentukan [17].

#### 3.2.2. Case Folding

*Case Folding* merupakan proses penulisan huruf dengan mengubah setiap huruf menjadi *lowercase* (huruf kecil). Ini hanya dapat menulis ulang huruf A hingga Z dan semua karakter lainnya akan hilang [18].

#### 3.2.3. Stopword Removal (id)

*Stopword Removal* kata-kata umum yang sering digunakan dan dianggap tidak memiliki makna. Seperti kata “yang”, “dan”, “ke”, sehingga perlu dilakukan penghapusan. Tujuannya untuk memperoleh kata-kata dalam *tweet* yang relevan dengan analisis dokumen. Pada tahap ini peneliti menghapus *stopword* berbahasa Indonesia terlebih dahulu

menggunakan database kamus *stopword* yang disediakan oleh [5].

#### 3.2.4. Tokenizing

*Tokenizing* atau tokenisasi adalah proses membagi kalimat menjadi bagian-bagian yang disebut token. Token dapat dibentuk dari kata, frasa, atau elemen bermakna lainnya. Pada tahap ini, proses tokenisasi dilakukan dengan bantuan Natural Language Toolkit (NLTK), sebuah pustaka *Python* yang berfokus pada *preprocessing data* [19].

#### 3.2.5. Normalisasi

Normalisasi berfungsi untuk memperoleh bentuk dasar dari kata-kata tersebut. Misalnya, kata “berlari” akan dinormalisasikan menjadi “lari” agar dapat dianggap sebagai bentuk yang sama [20].

#### 3.2.6. Stemming

Tahapan *stemming* bertujuan mengembalikan setiap kata ke bentuk akarnya, seperti “yg” menjadi “yang” yang tidak bisa di baca untuk mengurangi variabilitas kata [21].

### 3.3. Labelling Lexicon

Setelah data dibersihkan pada tahapan *preprocessing data*, dilakukan pelabelan menggunakan metode *lexicon*. Penentuan dilakukan pada data teks berupa kalimat yang memiliki kata pada kamus *lexicon* yang terdiri dari kata negatif dan positif. Kata yang teridentifikasi dalam kamus *lexicon* akan dihitung skor nya sesuai dengan jumlah kata pada setiap teks atau kalimat [21].

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

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

Dimana  $S_{positive}$  ( $S_{pos}$ ) adalah bobot dari kalimat yang didapatkan melalui penjumlahan  $n$  skor polaritas kata opini positif  $S_{negative}$  ( $S_{neg}$ ) adalah bobot dari kalimat yang didapatkan melalui penjumlahan  $n$  skor polaritas kata opini negatif. Dari persamaan nilai sentimen (*sent*) dalam satu kalimat maka diperoleh persamaan 3 untuk menentukan

orientasi sentimen dengan perbandingan jumlah nilai positif, negatif, dan netral.

$$sentence_{sent} \begin{cases} positif \text{ if } S_{pos} > S_{neg} \\ neutral \text{ if } S_{pos} = S_{neg} \\ negative \text{ if } S_{pos} < S_{neg} \end{cases} \quad (3)$$

Jika dalam suatu teks memiliki jumlah kata positif lebih banyak dari kata negatif, maka data teks tersebut akan di labeli sentimen positif. Jika dalam suatu teks memiliki jumlah kata positif lebih sedikit dari kata negatif, maka data teks tersebut akan di labeli sentimen negatif. Jika dalam suatu teks memiliki jumlah kata positif sama dengan kata negatif, maka data teks tersebut akan di labeli sentimen netral.

### 3.4. NRC Emotion Lexicon

*NRC Emotion Lexicon* digunakan dalam berbagai aplikasi pemrosesan bahasa alami, analisis sentimen, serta riset dimana pemahaman terhadap dimensi emosional dalam teks memiliki peran penting. *NRC Emotion Lexicon* ini mendukung algoritma dan model untuk mengidentifikasi emosi yang tersemat dalam kata-kata, membantu dalam menggambarkan variasi dan tingkat kekuatan emosi yang mungkin terungkap dalam suatu teks [13].

### 3.5. Pembobotan emosi dengan NRC Emotion Lexicon (id)

*NRC Emotion Lexicon* atau yang bisa disebut juga *EmoLex* dikembangkan oleh [11]. Leksikon ini merupakan daftar kata-kata bahasa Indonesia terkait dengan emosi dan sentimen yang terdiri dari 14.182 kata (*unigram*). Dua sentimen berupa positif dan negatif serta delapan emosi dasar menurut Plutchik, yakni *anger* (marah), *anticipation* (antisipasi), *disgust* (muak atau jijik), *fear* (takut), *joy* (senang atau bahagia), *sadness* (sedih), *surprise* (terkejut), dan *trust* (percaya). Di dalam leksikon ini terdapat *association score*. Masing-masing kata diberi skor 0 atau 1. Angka 0 menunjukkan kata yang tidak berhubungan, sedangkan angka 1 ditujukan untuk kata yang berhubungan [5]. Penggunaan *EmoLex* dalam penelitian ini dipilih karena mampu memetakan spektrum emosi secara komprehensif hingga ke tingkat granularitas emosi spesifik, bukan sekadar

membagi teks ke dalam polaritas sentimen biner.

### 3.6. Perhitungan TF-IDF

Metode yang dikenal sebagai *Term Frequency-Inverse Document Frequency* (TF-IDF) digunakan untuk menentukan jumlah bobot yang terkait dengan setiap kata yang telah di ekstrak. Dalam model pembobotan TF-IDF, frekuensi kata (TF) dan frekuensi dokumen balik (IDF) di integrasikan untuk menghitung jumlah kata yang sering digunakan dalam proses pencarian data. Frekuensi kata (TF) menghitung jumlah kata yang muncul dalam satu dokumen (*tweet*), yang dianggap sebagai kata umum yang tidak penting untuk dinilai. Metode TF-IDF menggabungkan dua gagasan: Pertama, nilai TF untuk setiap kata dihitung dengan bobot awal 1, kemudian dengan menggunakan persamaan 2 menghitung nilai IDF [18].

$$IDF(Word) = \log\left(\frac{tf}{df}\right) \quad (4)$$

Metrik IDF (*word*) digunakan untuk setiap kata yang dicari, dimana *td* jumlah total dokumen yang ada dan *df* adalah jumlah kata yang muncul di seluruh dokumen [18].

### 3.7. Penyeimbangan Data Menggunakan SMOTE

Untuk mengatasi bias klasifikasi akibat ketidakseimbangan data, teknik *resampling* berupa *Synthetic Minority Oversampling Technique* (SMOTE) diterapkan pada tahap pra-pemrosesan data. Bekerja pada tingkat ruang fitur (*feature space*) dengan cara membangkitkan sampel sintetis baru untuk kelas minoritas. Pembentukan sampel sintetis oleh SMOTE dirumuskan pada persamaan berikut:

$$x_{syn} = x_i + (x_{knn} - x_i) \times \delta \quad (5)$$

Keterangan:

$x_{syn}$ = Data sintesis yang akan dibuat

$x_i$ = Data yang akan direplikasi

$x_{knn}$ = Data tetangga terdekat

$\delta$ = Nilai random antara 0 dan 1

Melalui penambahan sampel sintetis ini, SMOTE menyeimbangkan proporsi antar-kelas sehingga algoritma SVM dapat mengenali pola spektrum sentimen dan emosi publik secara adil tanpa terdistorsi oleh dominasi kelas tertentu.

### 3.8. Klasifikasi *Support Vector Machine* (SVM)

Pada tahap ini, *tweet* dikumpulkan dalam dua kategori negatif dan positif. Dalam analisis ini, *Support Vector Machine* (SVM) adalah algoritma yang di gunakan. Setelah data diklasifikasikan, kualitas model yang dibuat dengan mempertahankan akurasi yang diperlukan akan dievaluasi [18]. Proses pelatihan SVM dalam penelitian ini dilakukan dengan menggunakan data latih yang telah di labeli emosi, kemudian diujikan terhadap data uji untuk mengukur akurasi dan performa klasifikasi. Setiap *tweet* diubah menjadi representasi vektor berbasis frekuensi kata dan dikombinasikan dengan fitur emosional dari *lexicon*. Kelebihan SVM adalah kemampuannya dalam menangani data yang tidak seimbang, serta bekerja baik dalam data teks berlabel ganda seperti emosi [4].

### 3.9. Evaluasi Model

Evaluasi performa model dilakukan untuk mengukur sejauh mana algoritma *Support Vector Machine* (SVM) mampu mengklasifikasikan data sentimen dan emosi publik yang telah di labeli menggunakan pendekatan *Lexicon-based*. Pengujian ini menggunakan skema pembagian data (*data splitting*) dengan proposi 80% untuk data pelatihan (*training*) dan 20% untuk data pengujian (*testing*). Penilaian dilakukan menggunakan empat metrik utama, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*, yang umum digunakan dalam klasifikasi multi- kelas. *Accuracy* mengukur proporsi prediksi yang benar terhadap seluruh data uji, dan dirumuskan sebagai berikut.

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

Pertama, Akurasi digunakan untuk mengukur ketepatan total model dalam mengategorikan seluruh data sentimen dan emosi secara menyeluruh dengan TP (*true positive*), TN (*true negative*), FP (*false positive*), dan FN (*false negative*). *Precision* digunakan untuk mengukur seberapa akurat model mengidentifikasi kelas positif sebagai kelas yang benar positif, statis ini tidak dapat menunjukkan dengan jelas jumlah kelas positif aktual yang memiliki hasil prediksi yang akurat.

Persamaan untuk metrik ini adalah sebagai berikut:

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

Selanjutnya, *Recall* digunakan untuk mengukur sensitivitas model dalam menjaring seluruh aspirasi negatif netizen agar tidak ada data kritik atau kemarahan terhadap DPR yang terlewatkan dalam analisis dengan rumus:

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

Terakhir, *F1-Score* diterapkan sebagai nilai rata-rata harmonik yang menyeimbangkan antara *precision* dan *recall*, yang sangat krusial untuk menjaga stabilitas hasil klasifikasi mengingat distribusi emosi pada isu krisis kepercayaan ini cenderung tidak seimbang (*imbalanced data*), dengan rumus:

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

## 4. HASIL DAN PEMBAHASAN

### 4.1. Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan melalui platform media sosial X menggunakan *tools Tweet Harvest* pada *Google Collab*. Proses pengumpulan data dilakukan dengan menerapkan tagar (#) sebagai kata kunci dari “#bubarkandpr”, “#resetindonesia”, “#reformasiDPR”, dan “#RUUperampasanaset” yang berkaitan dengan Demo DPR Agustus 2025. Pengambilan data dilaksanakan pada 25 Agustus – 5 Oktober 2025, mencakup unggahan sebelum dan sesudah pertandingan. Melalui proses ini diperoleh total 2.120 *tweet*, data itu disimpan dalam format CSV. Selain data primer dari X, penelitian ini juga mengintegrasikan dataset dari Kaggle sebagai data pembantu (*secondary data*). Penambahan data dari Kaggle bertujuan untuk memperkaya model *Machine Learning*, sehingga hasil klasifikasi sentimen dan emosi memiliki tingkat validitas yang lebih tinggi. Dengan ini, data yang diperoleh total 19.454 komentar.

### 4.2. Preprocessing

Tahap *preprocessing* teks merupakan langkah krusial dalam penelitian ini untuk memastikan kualitas data sebelum masuk tahap klasifikasi. Mengingat karakteristik data yang diambil dari media sosial X memiliki tingkat derau (*noise*) yang tinggi, seperti penggunaan

singkatan, istilah *slang*, serta simbol-simbol non-alfabet, maka proses pembersihan menjadi penentu akurasi model SVM yang dibangun. Proses ini dimulai dengan *cleansing* untuk menghapus elemen yang tidak memiliki nilai semantik seperti *URL*, mentions, dan hashtags, dilanjutkan dengan *case folding* untuk menyeragamkan seluruh karakter menjadi huruf kecil. Selanjutnya, dilakukan tahap Normalisasi yang secara khusus menangani kata-kata tidak baku terkait isu politik seperti 'fufufafa' atau 'mulyono' menggunakan kamus kostum, serta tahap *tokenizing* untuk memecah kalimat menjadi satuan kata tunggal. Proses diakhiri dengan *stopword removal* dan Stemming menggunakan pustaka Sastrawi untuk mereduksi kata menjadi bentuk dasarnya, sehingga algoritma dapat fokus pada esensi emosi dan sentimen yang terkandung dalam setiap unggahan publik terkait mosi tidak percaya terhadap DPR.

Tabel 1. Hasil *Preprocessing*

| Tahapan Proses   | Hasil                                                                                                                                                                                                                                                                                    |
|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data Mentah      | Dan meninggal 3 orang dr dalam gedung DPRD yg dibakar anda demo menyampaikan aspirasi berubah menjadi seorang pembunuh korban jg tdk bersalah spt ojol yg meninggal terus skrg siapa yg disalahkan karena pendemo yg awalnya mulai pembakaran.<br>#BubarkanDPR<br>#resetindonesia #chaos |
| <i>Cleansing</i> | Dan meninggal orang dr dalam gedung DPRD yg dibakar anda demo menyampaikan aspirasi berubah menjadi seorang pembunuh korban jg tdk bersalah spt ojol yg meninggal terus skrg siapa yg disalahkan karena pendemo yg awalnya mulai pembakaran                                              |

| Tahapan Proses          | Hasil                                                                                                                                                                                                                                                                                                                                                                          |
|-------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Case Folding</i>     | dan meninggal orang dr dalam gedung dprd yg dibakar anda demo menyampaikan aspirasi berubah menjadi seorang pembunuh korban jg tdk bersalah spt ojol yg meninggal terus skrg siapa yg disalahkan karena pendemo yg awalnya mulai pembakaran                                                                                                                                    |
| <i>Normalization</i>    | dan meninggal orang dari dalam gedung dprd yang dibakar anda demo menyampaikan aspirasi berubah menjadi seorang pembunuh korban juga tidak bersalah seperti ojol yang meninggal terus sekarang siapa yang disalahkan karena pendemo yang awalnya mulai pembakaran                                                                                                              |
| <i>Tokenizing</i>       | ['dan', 'meninggal', 'orang', 'dari', 'dalam', 'gedung', 'dprd', 'yang', 'dibakar', 'anda', 'demo', 'menyampaikan', 'aspirasi', 'berubah', 'menjadi', 'seorang', 'pembunuh', 'korban', 'juga', 'tidak', 'bersalah', 'seperti', 'ojol', 'yang', 'meninggal', 'terus', 'sekarang', 'siapa', 'yang', 'disalahkan', 'karena', 'pendemo', 'yang', 'awalnya', 'mulai', 'pembakaran'] |
| <i>Stopword Removal</i> | ['meninggal', 'orang', 'gedung', 'dprd', 'dibakar', 'demo', 'menyampaikan', 'aspirasi', 'berubah', 'pembunuh', 'korban', 'bersalah', 'ojol', 'meninggal', 'disalahkan',                                                                                                                                                                                                        |

| Tahapan Proses  | Hasil                                                                                                                                                                  |
|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                 | 'pendemo', 'awalnya', 'pembakaran']                                                                                                                                    |
| <i>Stemming</i> | ['tinggal', 'orang', 'gedung', 'dprd', 'bakar', 'demo', 'sampai', 'aspirasi', 'ubah', 'bunuh', 'korban', 'salah', 'ojol', 'tinggal', 'salah', 'demo', 'awal', 'bakar'] |

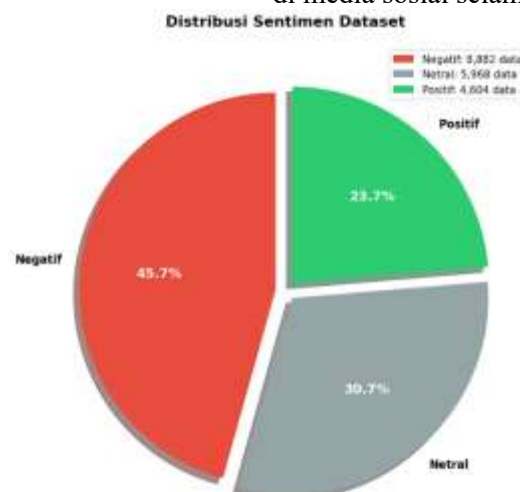
#### 4.3. Hasil Pelabelan Sentimen Menggunakan *Lexicon-Based*

Setelah label target terbentuk melalui pendekatan *lexicon-based*, data tersebut digunakan sebagai data latih dan data uji untuk proses klasifikasi menggunakan algoritma *Support Vector Machine* (SVM). Proses ini diawali dengan mentransformasi data teks menjadi bentuk vektor numerik melalui metode TF-IDF agar dapat diproses oleh model. Sebagai gambaran awal mengenai data yang digunakan dalam fase pelatihan dan pengujian, Tabel 2 sampel pelabelan lexicon based menyajikan potongan data hasil *scoring* leksikon yang menunjukkan keterkaitan antara akumulasi bobot kata dengan label polaritas yang dihasilkan.

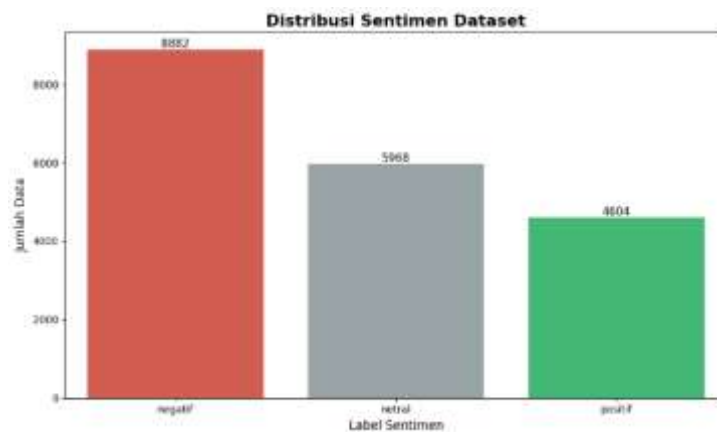
Tabel 2. Hasil Pelabelan *Lexicon Based*

| Text                                                | Sentiment | Polarity |
|-----------------------------------------------------|-----------|----------|
| tum, partai politik, pemerintah, kader, dprri, r... | 2         | Positif  |
| dengerin, tuhan, sri mulyani, efisiensi, tapi,...   | -10       | Negatif  |
| bikin, banjir, indonesia gelap, reset indonesia     | 4         | Positif  |
| kontrol, polisi, babi, reset indonesia, polisip...  | -9        | Negatif  |
| gila, uji, orang, indonesia, mudah, anggota, t...   | -19       | Negatif  |

Hasil klasifikasi akhir yang dilakukan oleh model SVM terhadap seluruh dataset menunjukkan kecenderungan opini yang sangat jelas. Sebaran hasil klasifikasi tersebut divisualisasikan dalam Gambar 2 dan Gambar 3, di mana sentimen negatif mendominasi opini publik dengan persentase sebesar 45,7% atau sebanyak 8.882 data. Dominasi ini mencerminkan kuatnya arus mosi tidak percaya di media sosial selama periode penelitian.



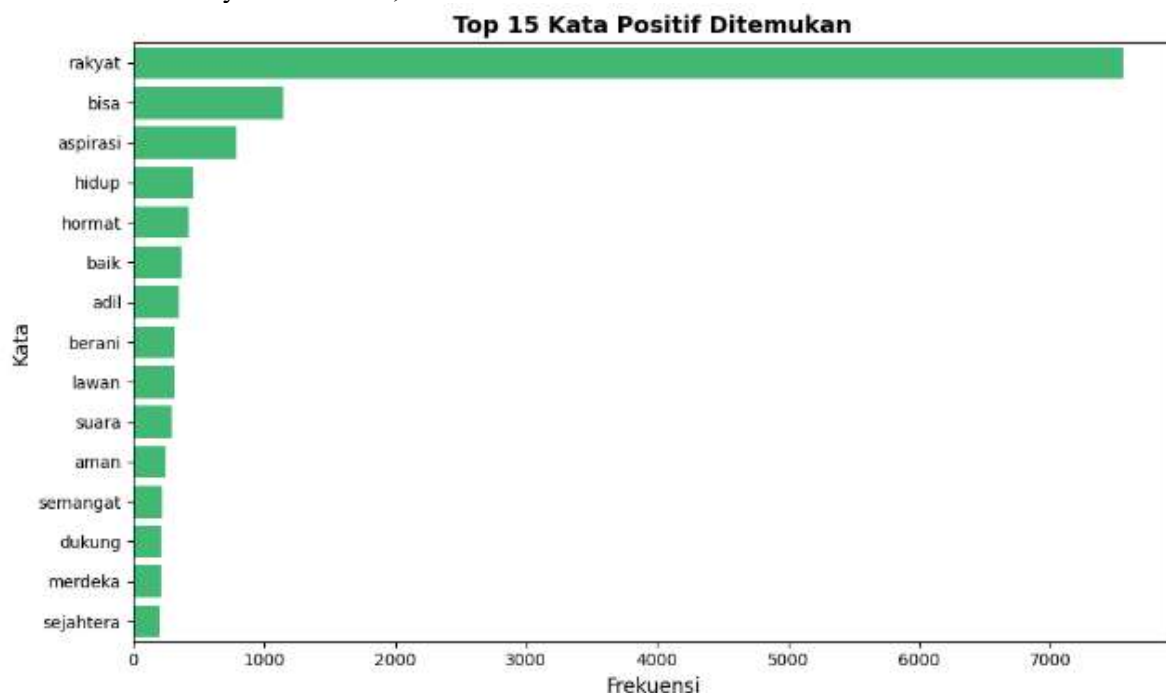
Gambar 2. Pie Chart Distribusi Emosi



Gambar 3. Bar Chart Distribusi Sentimen Dataset

Untuk memahami substansi dari setiap kategori sentimen yang telah berhasil diklasifikasikan oleh SVM, dilakukan ekstraksi kata kunci yang paling representatif pada masing-masing kelas. Gambar 4 memperlihatkan bahwa narasi negatif didominasi oleh kata-kata provokatif seperti 'bubar', 'demonstrasi', dan 'koruptor', yang menjadi indikator utama ketidakpuasan publik. Sebaliknya, pada Gambar 5, muncul istilah-istilah seperti 'rakyat', 'aspirasi', dan 'bisa' yang merepresentasikan harapan atau dukungan terbatas masyarakat. Temuan ini membuktikan bahwa model SVM mampu menangkap esensi pesan dari setiap dokumen sesuai dengan orientasi sentimen nya. Di sisi lain, kelas netral

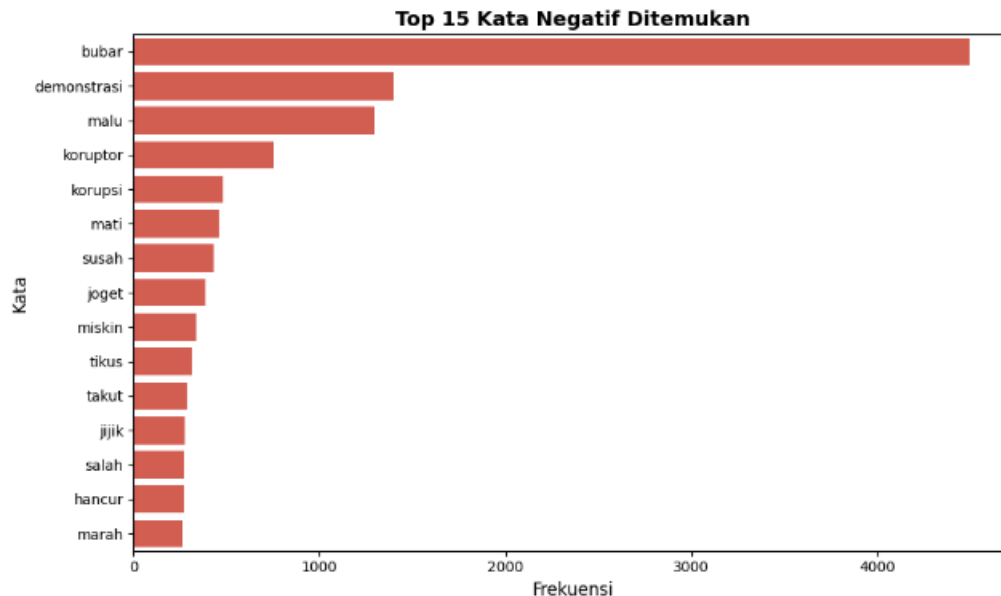
yang diwakili oleh kata kunci objektif seperti 'rapat', 'paripurna', dan 'jadwal' mengonfirmasi kemampuan model dalam mengisolasi teks yang murni bersifat informatif tanpa bias emosi. Kemunculan kluster kata ini secara masif mencerminkan dinamika eskalasi mosi tidak percaya publik yang terekam secara riil pada media sosial X. Melalui pola penyebaran kata tersebut, terlihat jelas bagaimana krisis kepercayaan sosiopolitik terhadap fungsi representasi DPR RI pasca-Demo Agustus 2025 berhasil ditransformasikan secara akurat oleh algoritma berbasis geometri hyperplane SVM ke dalam dimensi komputasional.



Gambar 4. Top 15 Kata Positif

Berdasarkan Gambar 5, ekstraksi kata kunci menunjukkan dominasi narasi kritis yang sangat kuat. Kata 'bubar' muncul sebagai istilah dengan frekuensi tertinggi, disusul oleh 'demonstrasi' dan 'malu'. Kemunculan kata-kata seperti 'koruptor', 'korupsi', 'mati', hingga istilah sindiran seperti 'joget' dan 'tikus' secara empiris

mencerminkan ekspresi ketidakpuasan serta krisis kepercayaan sosiopolitik masyarakat terhadap DPR. Pola ini membuktikan kemampuan model SVM dalam menangkap esensi pesan sesuai orientasi sentimen negatif publik.



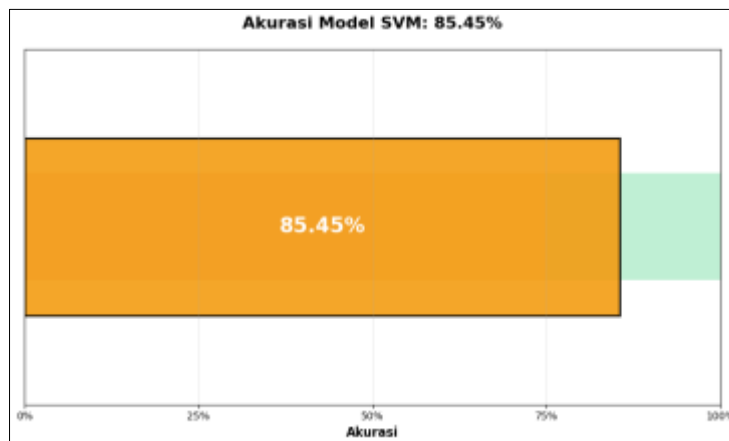
Gambar 5. Top 15 Kata Negatif

Melalui visualisasi distribusi dan temuan kata kunci tersebut, terlihat bahwa model SVM telah berhasil mengklasifikasikan data dengan pola yang konsisten terhadap label *lexicon-based*. Namun, efektivitas sebuah model klasifikasi tidak hanya diukur dari kemampuannya menghasilkan distribusi label, tetapi juga dari tingkat ketepatan prediksinya dibandingkan dengan label sebenarnya. Oleh karena itu, tahap berikutnya adalah melakukan evaluasi performa model untuk mengetahui seberapa besar tingkat keberhasilan algoritma SVM dalam menggeneralisasi data. Pengujian ini dilakukan dengan mengukur nilai akurasi

global serta melihat detail persebaran prediksi melalui metrik *Confusion Matrix* dan *Classification Report* yang akan dijabarkan pada subbab evaluasi berikut.

#### 4.4. Modelling SVM terhadap Pelabelan *Lexicon-Based*

Pada tahap evaluasi ini, performa model diuji untuk menentukan kualitas klasifikasi yang telah dihasilkan. Model SVM mencapai tingkat akurasi sebesar 85,45%. Angka ini mengindikasikan bahwa model memiliki keandalan yang sangat baik dalam memetakan opini publik terkait isu DPR.



Gambar 6. Hasil Akurasi Model SVM

Untuk menganalisis sebaran prediksi benar dan salah secara lebih mendalam, digunakan visualisasi Gambar 7 Berdasarkan matriks tersebut, terlihat bahwa model sangat kuat dalam mengenali sentimen negatif dengan 1.776 data yang terklasifikasi secara akurat. Matriks ini juga memperlihatkan tingkat kesalahan klasifikasi yang relatif rendah di antara kelas-kelas yang ada, yang membuktikan bahwa pemisahan fitur teks menggunakan *Kernel Linear* bekerja secara efektif pada dataset mosi tidak percaya ini.

Analisis mendalam terhadap visualisasi tersebut juga mengindikasikan bahwa tingginya akurasi pada klasifikasi sentimen negatif dipengaruhi secara signifikan oleh karakteristik sebaran kata (*vocabulary density*) dalam dataset. Dominasi sentimen negatif sebesar 1.776 data mencerminkan polaritas emosi masyarakat yang cenderung homogen saat menyuarakan mosi tidak percaya di media sosial. Pola tekstual pada kelas ini memiliki leksikon yang kuat, tegas, dan repetitif, sehingga memudahkan fungsi pembatas (*hyperplane*) dari *Kernel Linear* untuk menarik garis pemisah yang tegas dari kelas lainnya. Di

Tabel 3. *Classification Report* Performa SVM *Lexicon-Based*

|           | precision | recall | F1-score | support |
|-----------|-----------|--------|----------|---------|
| negatif   | 0.89      | 0.84   | 0.86     | 1776    |
| netral    | 0.91      | 0.91   | 0.91     | 1194    |
| positif   | 0.73      | 0.81   | 0.77     | 923     |
| accuracy  |           |        | 0.85     | 3891    |
| macro avg | 0.84      | 0.85   | 0.85     | 3891    |

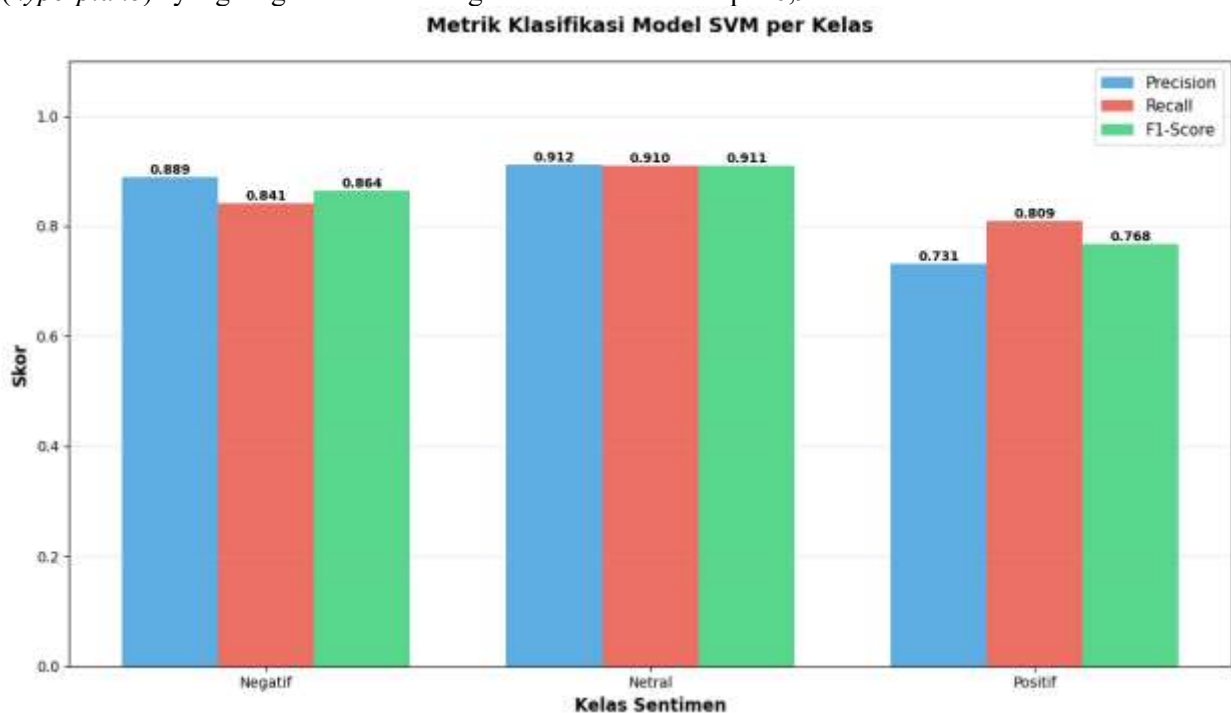
sisi lain, rendahnya tingkat kesalahan klasifikasi antar-kelas menunjukkan bahwa model tidak mengalami bias yang ekstrem, meskipun terdapat ketimpangan jumlah data (*data imbalance*) yang umum dijumpai pada analisis teks media sosial. Keberhasilan ini tidak lepas dari penerapan teknik resampling *Synthetic Minority Oversampling Technique* (SMOTE) pada tahap pra-pemrosesan data, yang secara efektif menyeimbangkan distribusi antar-kelas sehingga algoritma SVM dapat mengenali pola kelas minoritas secara optimal. Hal ini juga mengonfirmasi bahwa penanganan ketidakseimbangan kelas yang dipadukan dengan pembersihan teks (*text preprocessing*) serta pembobotan fitur telah mampu mereduksi derau (*noise*) sekaligus mempertahankan esensi semantik dari dokumen. Dengan demikian, hasil pencocokan linear ini tidak hanya memvalidasi keandalan teoretis dari algoritma yang digunakan, tetapi juga memberikan pembuktian empiris bahwa pendekatan komputasional integratif ini sangat adaptif dalam memetakan fenomena sosiologis berupa ketidakpuasan publik secara akurat dan objektif melalui ekstraksi korpus digital yang masif.

|              | precision | recall | F1-score | support |
|--------------|-----------|--------|----------|---------|
| weighted avg | 0.86      | 0.95   | 0.86     | 3891    |

Sebagaimana disajikan pada Tabel 3, pengujian *layer 1* menunjukkan bahwa model SVM memiliki performa prediksi yang sangat baik secara keseluruhan dengan akurasi mencapai 85% dari total 3.891 sampel data uji. Sentimen netral mencatatkan performa tertinggi dengan *F1-score* sebesar 91% (*precision* 91%

dan *recall* 91%; 1.194 sampel). Sentimen negatif menyusul dengan *F1-score* sebesar 86% (*precision* 89% dan *recall* 84%; 1.776 sampel), sedangkan sentimen positif memperoleh *F1-score* sebesar 77% (*precision* 73% dan *recall* 81%; 923 sampel). Tingginya konsistensi performa pada ketiga kelas ini didukung oleh ketersediaan sampel data uji (*support*) yang kuat pada setiap kategori polaritas. Ketersediaan data yang melimpah ini sangat menguntungkan bagi algoritma SVM dalam membentuk bidang pemisah optimal (*hyperplane*) yang tegas untuk mengisolasi

masing-masing kelas. Akibatnya, model mampu mengenali karakteristik bahasa, pola opini yang kritis, serta pernyataan publik yang bersifat informatif atau objektif (netral) di media sosial X secara akurat sebelum melangkah pada tingkatan analisis emosi yang lebih spesifik. Perbandingan metrik antar kelas tersebut divisualisasikan kembali pada Gambar 7 untuk mempermudah identifikasi kelas dengan performa terbaik. Dari grafik tersebut, dapat disimpulkan bahwa kelas Netral memiliki stabilitas tertinggi dengan nilai metrik mencapai 0,911.



Gambar 7. Metrik Klasifikasi Model SVM

Secara keseluruhan, hasil evaluasi yang ditunjukkan oleh Gambar 7 mempertegas keberhasilan integrasi antara pelabelan *lexicon-based* dengan klasifikasi *Support Vector Machine*. Tingginya nilai metrik pada seluruh kelas, khususnya pada sentimen negatif yang menjadi fokus utama mosi tidak percaya, membuktikan bahwa fitur-fitur linguistik yang diekstraksi melalui TF-IDF telah mampu mewakili polaritas opini publik secara akurat. Meskipun terdapat selisih performa pada kelas positif, pencapaian akurasi global yang melampaui angka 86% memberikan validasi bahwa model ini memiliki tingkat reliabilitas yang sangat tinggi. Dengan demikian, sistem

yang dikembangkan tidak hanya mampu melakukan klasifikasi secara otomatis dalam skala besar, tetapi juga dapat menjadi instrumen yang sah untuk memetakan dinamika ketidakpuasan publik terhadap lembaga legislatif di media sosial.

#### 4.5. Emotional Lexicon Labelling

Setelah dilakukan klasifikasi sentimen secara umum menggunakan algoritma SVM, tahap analisis selanjutnya adalah melakukan *Emotional Lexicon Labelling* guna membedakan dimensi emosional yang terkandung dalam opini publik. Berbeda dengan analisis sentimen yang hanya membagi data ke dalam polaritas positif, negatif, dan netral, pendekatan

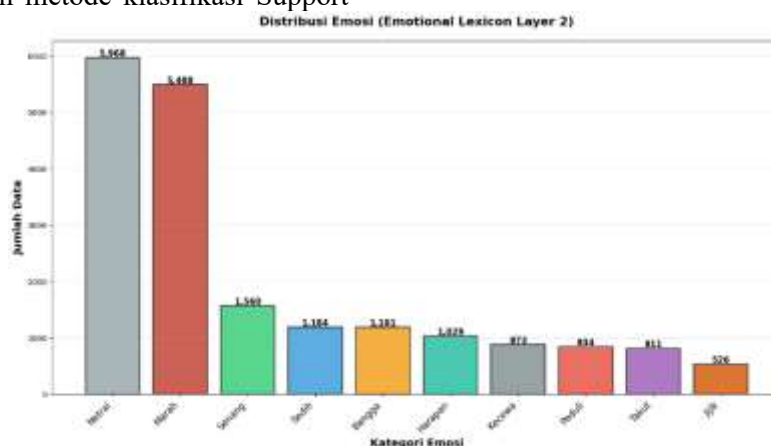
*emotional lexicon* (EmoLex) pada tahap ini bertujuan untuk mengidentifikasi kategori emosi spesifik yang mendasari munculnya mosi tidak percaya tersebut. Dengan mencocokkan kata-kata kunci dalam dataset terhadap kamus emosi yang telah ditentukan, setiap dokumen tweet dianalisis untuk mendeteksi keberadaan emosi dominan seperti amarah (*anger*), ketakutan (*fear*), hingga kekecewaan. Analisis emosi ini sangat krusial dalam penelitian ini karena mosi tidak percaya terhadap lembaga legislatif bukan sekadar ekspresi negatif yang bersifat umum, melainkan sebuah manifestasi dari akumulasi emosi publik yang lebih mendalam. Visualisasi mengenai persebaran emosi ini disajikan secara mendetail pada Gambar 8 berikut.

Melalui visualisasi pada Gambar 8 tersebut, karakteristik emosi publik dapat dipetakan secara lebih komprehensif guna memahami dinamika psikologis massa di balik narasi mosi tidak percaya. Kehadiran emosi spesifik seperti amarah dan ketakutan yang mendominasi korpus data mengonfirmasi bahwa ketidakpuasan terhadap lembaga legislatif dipicu oleh isu-isu struktural yang menyentuh rasa keadilan masyarakat. Proses pelabelan berbasis leksikon ini berhasil mengungkap bahwa di balik polaritas sentimen negatif yang telah diklasifikasikan oleh SVM pada tahap sebelumnya, terdapat intensitas emosi yang bervariasi dan tidak bersifat tunggal. Dengan demikian, integrasi antara pendekatan pembelajaran mesin dan analisis leksikon emosi ini memberikan dimensi analisis yang lebih kaya, objektif, serta mendalam terhadap fenomena sosiologis digital.

Sebagai simpulan dari tahapan analisis ini, penggabungan metode klasifikasi Support

Vector Machine dan eksplorasi *Emotional Lexicon Labelling* berhasil memberikan pembuktian empiris yang utuh mengenai kompleksitas opini publik di media sosial. Konstruksi sentimen negatif yang awalnya terdeteksi secara masif oleh model SVM kini tidak lagi dipandang sebagai sekadar angka statistik makro, melainkan telah terbedah secara mikro menjadi rincian emosional yang merepresentasikan gejolak sosial yang nyata. Keberadaan emosi amarah, ketakutan, hingga kekecewaan yang terekstraksi dari kluster data mosi tidak percaya ini menjadi indikator penting bahwa diskursus digital mengenai lembaga legislatif didorong oleh tuntutan akuntabilitas yang sangat kuat. Secara metodologis, keberhasilan pendekatan hibrida ini menegaskan bahwa analisis teks berbasis komputasi mampu melampaui batas klasifikasi polaritas sederhana dengan menyajikan kedalaman konteks psikologis penuturnya.

Validitas hasil yang ditunjukkan melalui visualisasi sebaran emosi ini juga memberikan kontribusi penting bagi studi sosiologi digital, khususnya dalam memetakan bagaimana ketidakpuasan struktural ditransformasikan menjadi narasi kolektif di ruang publik virtual. Dengan demikian, seluruh rangkaian visualisasi dan interpretasi data yang telah dipaparkan tidak hanya menjawab rumusan masalah mengenai arah sentimen masyarakat, melainkan juga meletakkan landasan teoretis yang kuat untuk memahami faktor-faktor pemicu utama di balik runtuhnya ekspektasi publik terhadap performa lembaga perwakilan rakyat dalam fenomena mosi tidak percaya ini.



Gambar 8. Distribusi Emosi Lexicon Layer 2

Berdasarkan Gambar 8, emosi Marah (*Anger*) menjadi emosi spesifik paling dominan (5.488 data) di luar kategori Netral (5.968 data), yang menegaskan bahwa ketidakpercayaan publik terhadap DPR RI berakar dari kemarahan kolektif di media sosial X. Spektrum emosi spesifik lainnya mencakup Senang (1.560), Sedih (1.184), Bangga (1.181), Harapan (1.029), Kecewa (873), Peduli (834), Takut (811), dan Jijik (526). Ketika dikorelasikan dengan *layer* 1, sentimen negatif dibentuk oleh akumulasi emosi *Marah*, *Sedih*,

*Kecewa*, *Takut*, dan *Jijik* sebagai narasi penolakan publik, sedangkan sentimen positif diwakili emosi *Senang*, *Bangga*, *Harapan*, dan *Peduli* yang merefleksikan optimisme serta solidaritas gerakan sipil. Pemetaan ini membuktikan keberagaman dimensi emosi yang mendasari polaritas sentimen publik secara komprehensif. Untuk memahami proses pelabelan emosi ini bekerja pada level kata, Tabel 4 menunjukkan contoh pembobotan kata-kata kunci terhadap delapan emosi dasar.

Tabel 4. Sampel Skor *Emotional Lexicon*

| Kata              | Anger    | Anticipation | Disgust  | Fear     | Joy      | Sadness  | Surprise | Trust    |
|-------------------|----------|--------------|----------|----------|----------|----------|----------|----------|
| dpr               | 0        | 0            | 0        | 0        | 0        | 0        | 0        | 0        |
| bangsat           | 1        | 0            | 1        | 0        | 0        | 0        | 0        | 0        |
| becus             | 1        | 0            | 1        | 0        | 0        | 0        | 0        | 0        |
| kerja             | 0        | 0            | 0        | 0        | 0        | 0        | 0        | 1        |
| bubar             | 1        | 0            | 0        | 1        | 0        | 1        | 0        | 0        |
| mahasiswa         | 0        | 0            | 0        | 0        | 0        | 0        | 0        | 1        |
| prestasi          | 0        | 1            | 0        | 0        | 1        | 0        | 1        | 1        |
| curi              | 1        | 0            | 1        | 1        | 0        | 0        | 0        | 0        |
| <b>Total Skor</b> | <b>4</b> | <b>1</b>     | <b>3</b> | <b>2</b> | <b>1</b> | <b>1</b> | <b>1</b> | <b>3</b> |

Tabel 4 memperlihatkan bahwa kata-kata yang kerap muncul dalam narasi mosi tidak percaya memiliki keterikatan kuat dengan lebih dari satu dimensi emosi. Sebagai contoh, kata 'bangsat' dan 'becus' memberikan kontribusi skor pada kategori *Anger* dan *Disgust*, sementara kata 'bubar' secara sistematis memicu skor pada kategori *Anger*, *Fear*, dan *Sadness*. Akumulasi skor pada kategori *Anger* (Total Skor: 4) dan *Disgust* (Total Skor: 3) dalam sampel tersebut membuktikan bahwa pemilihan kata oleh masyarakat memang diarahkan untuk mengekspresikan kemarahan yang mendalam. Mekanisme pembobotan ini memastikan bahwa klasifikasi emosi yang dihasilkan pada Gambar 9 memiliki basis linguistik yang kuat dan relevan dengan konteks kritik terhadap DPR.

Untuk memberikan gambaran yang lebih konkret mengenai representasi emosi

dalam data nyata, Tabel 5 menyajikan persebaran jumlah data untuk setiap kategori emosi lengkap dengan sampel tweet yang mewakilinya. Melalui visualisasi ini, terlihat jelas bahwa selain kategori Netral (5.968 data), emosi Marah (4.757 data) menjadi ekspresi yang paling dominan dalam mosi tidak percaya.

Contoh kalimat pada kategori tersebut sering kali menggunakan diksi kasar dan tagar perlawanan yang menunjukkan tingginya tensi kemarahan masyarakat. Sebaliknya, kemunculan emosi seperti Bangga (2.429 data) dan Kecewa (2.038 data) menunjukkan adanya dualitas respons, di mana publik merasa kecewa terhadap lembaga namun tetap bangga dalam menyuarakan persatuan dan solidaritas aksi di media sosial.

Tabel 5. Distribusi Data Emosi

| Contoh Kalimat | Emosi                                              | Jumlah Data |
|----------------|----------------------------------------------------|-------------|
| Netral         | bener dibikin banjir indonesiagelap resetindonesia | 5968        |

| Contoh Kalimat | Emosi                                                                                                                                                                                                                                                                        | Jumlah Data |
|----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|
| Marah          | kontol polisi babi resetindonesia polisipembunuh polisimusuhbersama                                                                                                                                                                                                          | 4757        |
| Bangga         | harusnya ketum parpol segera perintahkan kadernya di dpr ri terutama namanama yang kini jadi musuh masyarakat agar segera mengundurkan diri sebagai bentuk tanggung jawab dan empati ke rakyat sebelum terlambat resetindonesia affankurniawan                               | 2429        |
| Kecewa         | gila ujian orang indonesia gua mohon banget untuk di waktu yang mudah an dekat ga ada lagi yang namanya memilih pemimpin karena rasa kasihanfomo polisipembunuh resetindonesia indoneiagelap acab                                                                            | 2038        |
| Senang         | ranah sipil sempit makin terdesak kerakusan penguasa dan didesak tamengtameng setianya lantas siapa pemegang kekuasaan tertinggi masih rakyat kan sekarang seperti rakyat belum berkuasa dan merdeka penuh perjuangkan rebut resetindonesia                                  | 1493        |
| Jijik          | gue sampe muntah liat semua berita yang beredar muak banget gue sama pemerintah ripdemokrasi resetindonesia                                                                                                                                                                  | 909         |
| Sedih          | dengerin tuh sri mulyani efisiensi tai kucing yg jd korban rakyat kecil rakyat semakin terjepit elit dikasih tunjangan yg makin melejit resetindonesia aparatkeparat aparatkontol                                                                                            | 791         |
| Harapan        | bisa ga si dpr besok pagi bgtt paripurna mendesak buat respon keos ini itu anggota yg keburu kabur gagu jawab pertanyaan media tone deaf sm ga deket dg konstituennya langsung copot aja tuntutan paw demi menjaga marwah partai dan bersihbersih lumbung resetindonesia     | 514         |
| Takut          | polisi yang baik cuma polisi tidur resetindonesia forever                                                                                                                                                                                                                    | 387         |
| Peduli         | agaknya sinuwun sudah mulai menapakkan kaki republik ini memberikan sang raja sebuah permainan dg ancaman memotong uang saku pembangunan daerahnya tentu beliau enggan menyanggah atau bahkan melawan karena tidak etis beliau memilih membersamai perjuangan resetindonesia | 168         |

#### 4.6. Modelling SVM terhadap *Emotion*

Tahap selanjutnya dalam penelitian ini adalah melakukan pemodelan *Support Vector Machine* (SVM) untuk mengklasifikasikan kategori emosi berdasarkan label yang telah dihasilkan oleh *Emotional Lexicon*. Berbeda dengan klasifikasi sentimen sebelumnya yang bersifat tri-polar, pemodelan pada tahap ini memiliki tingkat kompleksitas yang lebih tinggi karena model harus mampu membedakan karakteristik linguistik di antara sepuluh kategori emosi yang berbeda. SVM dipelajari

untuk mengenali pola fitur TF-IDF yang mencirikan emosi amarah, kebanggaan, hingga kekecewaan, dengan tujuan untuk melihat sejauh mana algoritma ini dapat menggeneralisasi pelabelan emosi otomatis secara konsisten.

Untuk melihat efektivitas model dalam melakukan generalisasi, Tabel 6 menyajikan perbandingan antara label aktual dari leksikon dengan hasil prediksi model SVM. Meskipun terdapat beberapa ketidaksesuaian prediksi pada kategori yang memiliki ambiguitas tinggi, model secara konsisten mampu memprediksi

dengan tepat kategori emosi yang memiliki fitur kata kunci yang kuat, seperti terlihat pada

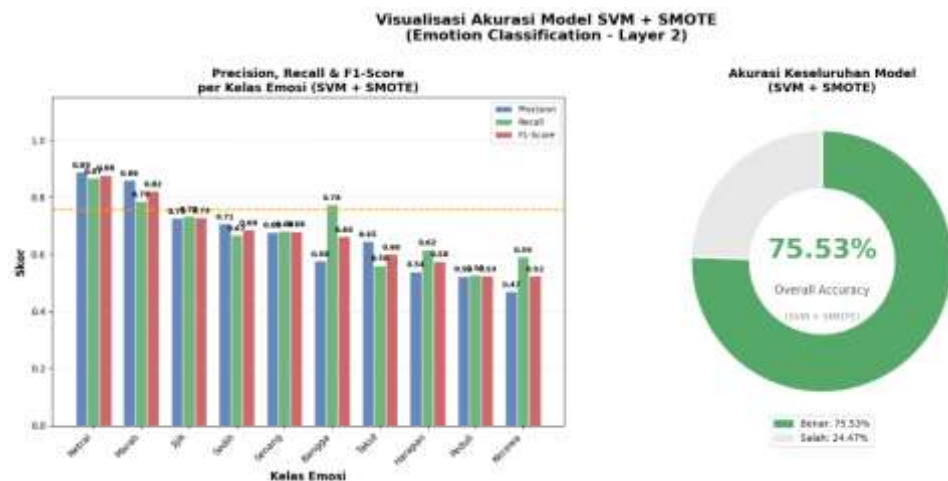
sampel data ke-4, ke-5, dan ke-7 yang diprediksi secara akurat sebagai emosi 'marah'.

Tabel 6. Sampel Hasil Prediksi Emosi Multikelas

| No | Text                                                        | Actual Emotion | Predicted Emotion SVM |
|----|-------------------------------------------------------------|----------------|-----------------------|
| 1  | demonstrasi, serta, anarkis, tidak, anarkis, suara, tida... | marah          | senang                |
| 2  | belum, rumah, pusing, hancur, kaya, rumah, sahroni          | kecewa         | sedih                 |
| 3  | hasil, mana, wkwkwkwkwkk, kocak                             | netral         | netral                |
| 4  | tuju, dprri, bubar, tidak, perintah, tinggi, presiden, tuju | marah          | marah                 |
| 5  | bicara, kosong, kerja, bisa, bodohin, lagiudah, bubar, d... | marah          | marah                 |

Tingkat akurasi yang dicapai oleh model SVM sebesar 75,53% membuktikan pada gambar 9 bahwa penerapan teknik resampling SMOTE (*Synthetic Minority Over-sampling Technique*) terbukti efektif dalam mengatasi ketimpangan jumlah data (class imbalance), sehingga algoritma SVM dengan kernel Linear

memiliki kemampuan generalisasi yang baik dalam mempelajari pola bobot TF-IDF untuk mengklasifikasikan kategori emosi publik. Performa tertinggi dicapai pada sentimen Netral dan Marah, yang menunjukkan model klasifikasi SVM mampu mengenali fitur-fitur dari kedua emosi tersebut secara konsisten.



Gambar 9. Visualisasi Akurasi Model SVM

Setelah tahap pemodelan selesai, dilakukan dengan menguji model pada Data Split dengan rasio 80:20, di mana 80% data digunakan sebagai data latih dan 20% sisanya sebagai data uji (*Data Test*) guna mengukur kemampuan generalisasi algoritma SVM dalam mengklasifikasikan dimensi emosi publik. Berdasarkan hasil pengujian terhadap data uji

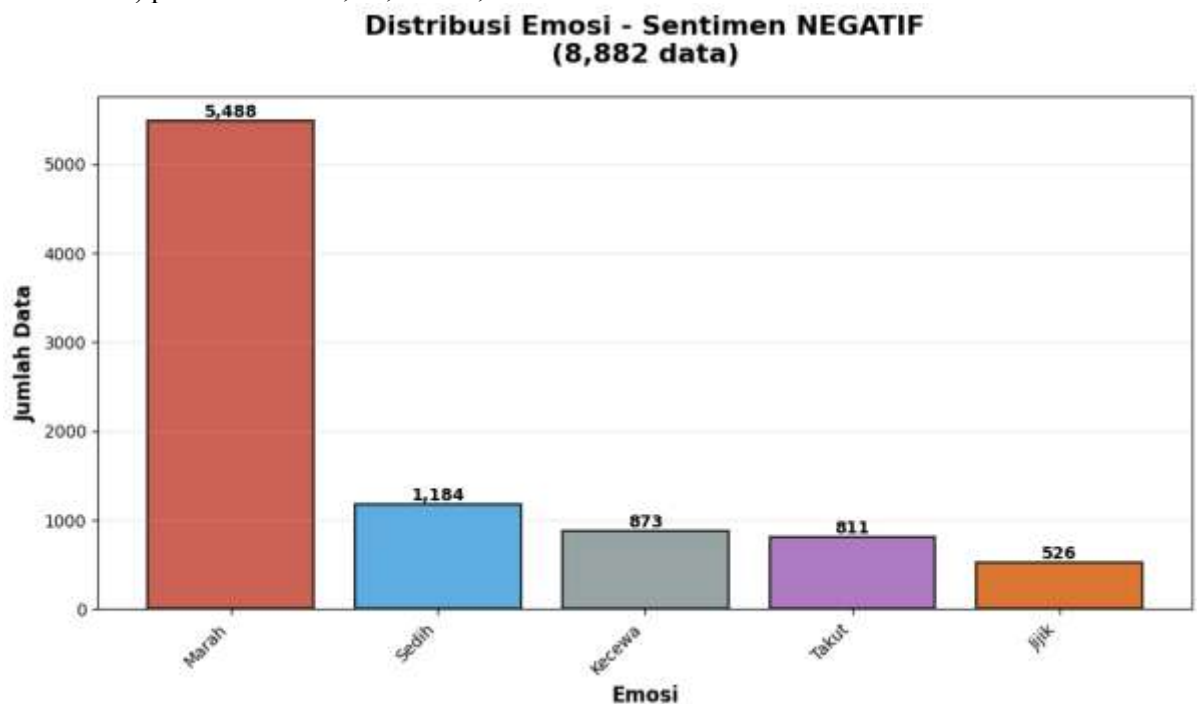
tersebut, model SVM berhasil mencapai nilai Test Accuracy sebesar 75,53%. Meskipun angka ini berada di bawah akurasi klasifikasi sentimen, hasil tersebut menunjukkan keandalan model yang tinggi dalam menangani klasifikasi multikelas yang terdiri dari sepuluh kategori emosi berbeda. Performa detail dari pengujian data split ini dirangkum dalam tabel berikut.

Tabel 7. Classification Report Evaluasi SVM terhadap Emotional Lexicon

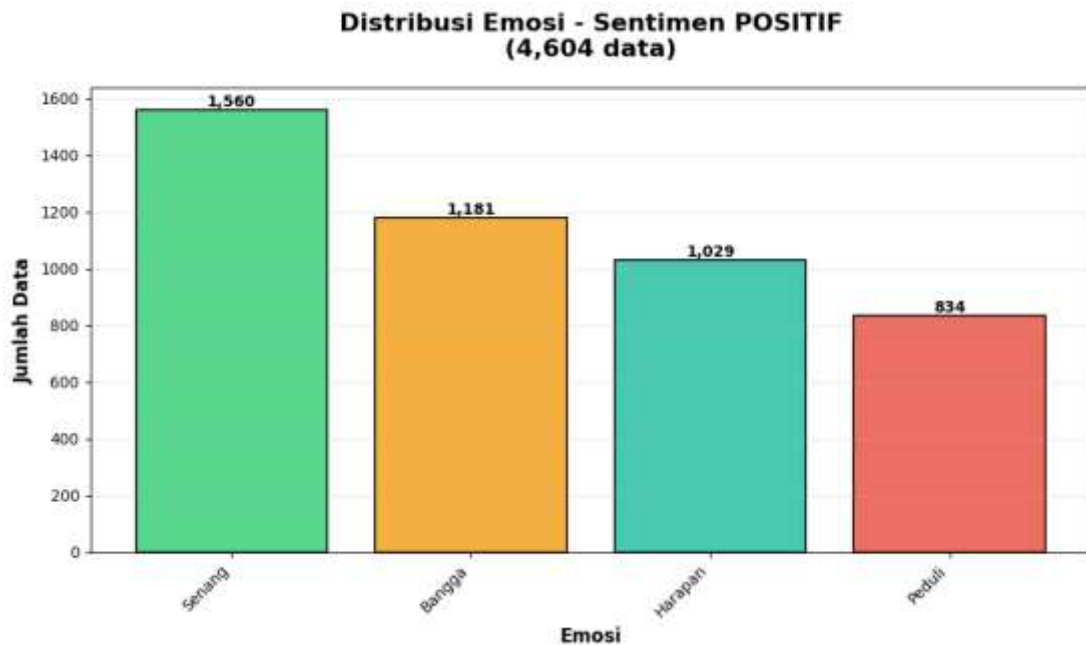
| Emotion         | Precision | Recall | F1-Score    | Support     |
|-----------------|-----------|--------|-------------|-------------|
| Netral          | 0.89      | 0.87   | 0.88        | 1194        |
| Marah           | 0.86      | 0.79   | 0.82        | 1098        |
| Senang          | 0.68      | 0.68   | 0.68        | 312         |
| Sedih           | 0.71      | 0.67   | 0.69        | 237         |
| Bangga          | 0.58      | 0.78   | 0.66        | 236         |
| Harapan         | 0.54      | 0.62   | 0.58        | 206         |
| Kecewa          | 0.47      | 0.59   | 0.52        | 174         |
| Peduli          | 0.52      | 0.53   | 0.53        | 167         |
| Takut           | 0.65      | 0.56   | 0.6         | 162         |
| Jijik           | 0.73      | 0.73   | 0.73        | 105         |
| <b>Accuracy</b> |           |        | <b>0.76</b> | <b>3891</b> |

Berdasarkan Tabel 7, model klasifikasi menunjukkan performa yang sangat optimal, terutama pada kategori Netral (*F1-Score* 0,88) dan Marah (*F1-Score* 0,82). Capaian ini mengindikasikan bahwa algoritma *Support Vector Machine* (SVM) sangat efektif dalam mengenali karakteristik serta preferensi fitur teks yang berkaitan dengan narasi objektif/kritik formal maupun ekspresi kemarahan publik. Ketajaman prediksi ini diperjelas melalui matriks kebingungan (*confusion matrix*) pada Gambar 10, 11, dan 12,

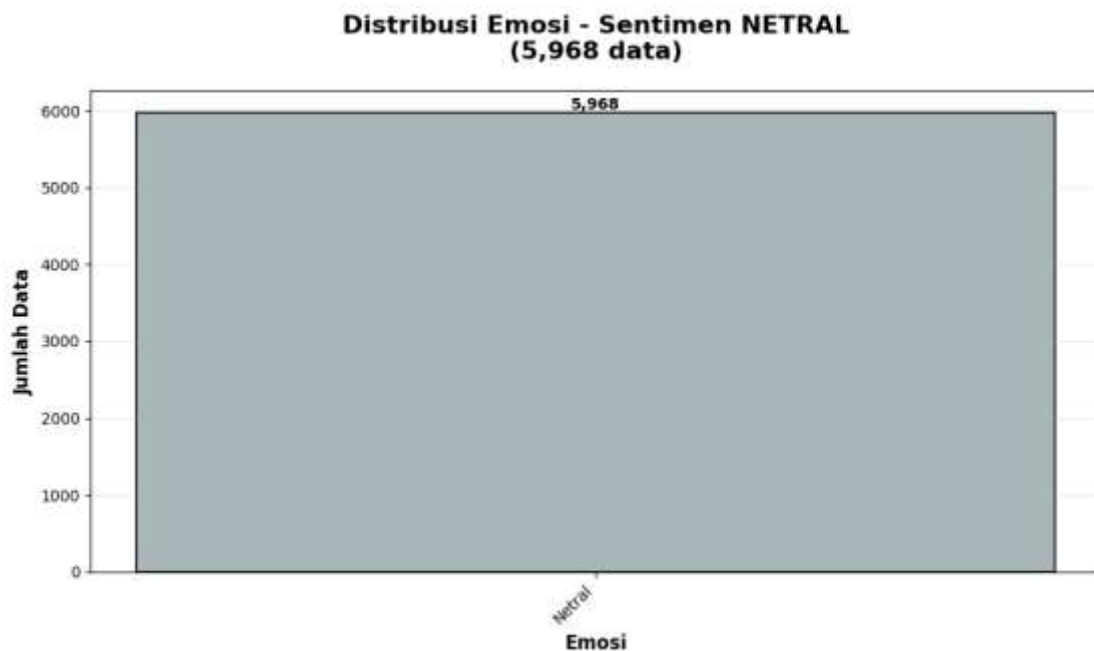
di mana konsentrasi angka dominan berada pada diagonal utama, menandakan mayoritas data uji terklasifikasi secara akurat sesuai label aktualnya. Distribusi perbandingan metrik evaluasi ini juga divisualisasikan secara komprehensif pada grafik bar chart di Gambar 13. Untuk memahami dinamika emosi secara lebih mendalam, analisis distribusi emosi kemudian diperluas ke dalam setiap kategori sentimen guna membedah muatan emosional di balik label positif, negatif, dan netral melalui visualisasi berikut.



Gambar 10. Distribusi Emosi Sentimen Negatif



Gambar 11. Distribusi Emosi Sentimen Positif



Gambar 12. Distribusi Emosi Sentimen Netral

Penelitian ini mengklasifikasikan data ke dalam tiga kelompok sentimen utama, yaitu Sentimen Negatif dengan total 8.882 data, Sentimen Netral sebanyak 5.968 data, dan Sentimen Positif sebesar 4.604 data. Pada kelompok Sentimen Negatif, emosi Marah menjadi motor utama dengan 5.488 data, disusul secara beruntutan oleh Sedih (1.184 data), Kecewa (873 data), Takut (811 data), dan Jijik (526 data), yang membuktikan bahwa

narasi negatif publik didominasi oleh kemarahan kolektif terhadap situasi yang berkembang. Sementara itu, pada kelompok Sentimen Positif, emosi Senang tampil mendominasi dengan 1.560 data, yang diikuti oleh emosi Bangga (1.181 data), Harapan (1.029 data), dan Peduli (834 data). Hasil ini merepresentasikan apresiasi publik terhadap solidaritas aksi serta harapan akan adanya perubahan. Adapun pada kelompok Sentimen

Netral, seluruh data (5.968 data) secara konsisten teridentifikasi tunggal sebagai emosi Netral.

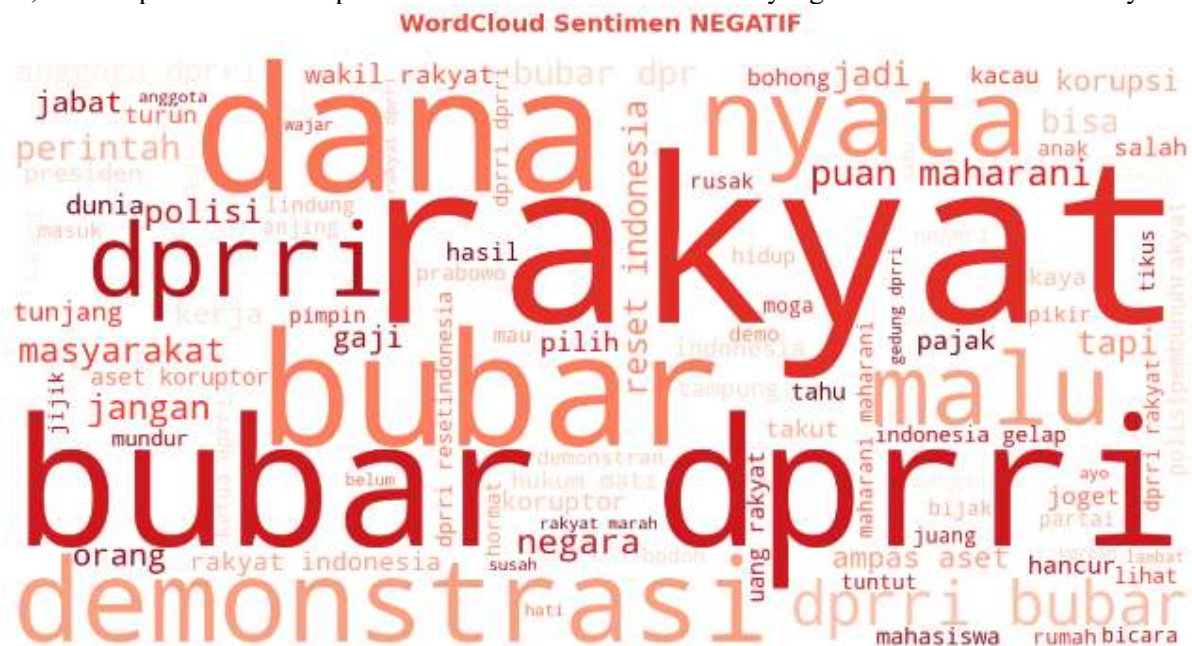
Menariknya, pada kelompok Sentimen Netral (5.968 data), seluruh data secara konsisten teridentifikasi sebagai emosi Netral. Namun, perlu dicatat bahwa dalam konteks mosi tidak percaya, emosi netral ini sebetulnya sering kali mengandung unsur sarkasme atau sindiran tajam terhadap situasi politik yang terjadi. Meskipun teks tidak mengandung kata-kata emosional yang meledak-ledak, penggunaan gaya bahasa ironi yang membenturkan fakta dengan diksi yang tampak "sopan" tetap membawa beban kritik negatif yang substansial. Model mengategorikannya sebagai netral dikarenakan keterbatasan sistem *Emotional Lexicon* dalam menangkap ambiguitas makna pada kalimat sarkasme.

Secara keseluruhan, evaluasi ini menegaskan bahwa algoritma SVM mampu bertindak sebagai instrumen pengklasifikasi emosi yang andal dengan tingkat akurasi 75,54%. Capaian ini tidak lepas dari efektivitas

penerapan teknik resampling SMOTE (Synthetic Minority Over-sampling Technique) yang berhasil mengeliminasi masalah ketimpangan distribusi data (class imbalance) pada kategori emosi minoritas. Hasil pada klasifikasi multikelas ini memberikan validasi teknis bahwa narasi mosi tidak percaya di media sosial X memiliki pola emosional yang terstruktur. Integrasi antara teknik resampling, hasil data split, dan sebaran emosi per sentimen memberikan gambaran komprehensif bahwa kemarahan kolektif terhadap lembaga legislatif merupakan pilar utama yang menggerakkan diskursus publik dalam penelitian ini.

#### 4.7. Word Cloud

Sebagai tahap akhir dari analisis data, dilakukan visualisasi menggunakan *Word Cloud* untuk memetakan kata-kata yang paling sering muncul (*term frequency*) dalam kumpulan data mosi tidak percaya terhadap DPR. Visualisasi ini berfungsi sebagai representasi grafis yang merangkum inti dari diskursus publik berdasarkan tiga kategori sentimen yang telah dianalisis sebelumnya.



Gambar 13. Word Cloud Sentimen Negatif



Gambar 14. Word Cloud Sentimen Positif



Gambar 15. Word Cloud Sentimen Netral

Berdasarkan perbandingan ketiga visualisasi pada Gambar 16, 17, dan 18, dapat ditarik analisis bahwa dominasi narasi perlawanan pada sentimen negatif terlihat jelas melalui kata "rakyat", "dprri", "bubar", "dana", dan "nyata" yang merepresentasikan kemarahan kolektif publik terhadap isu transparansi anggaran. Hal ini selaras dengan dominasi emosi Marah dan Kecewa yang telah di validasi

sebagai motor utama penggerak mosi tidak percaya dalam dataset ini. Di sisi lain, kelompok sentimen positif menunjukkan adanya dimensi solidaritas dan harapan melalui kemunculan kata "mahasiswa" dan "perintah", yang mengindikasikan bahwa emosi Bangga dalam konteks ini lebih merujuk pada apresiasi terhadap gerakan aksi mahasiswa demi

kepentingan rakyat daripada dukungan terhadap lembaga terkait.

Sementara itu, visualisasi pada *Word Cloud* netral menunjukkan kemunculan signifikan kata kunci seperti "maharani", "puan", "indonesia", dan "reset", yang mengonfirmasi analisis pada subbab 4.7 bahwa sentimen netral dalam mosi tidak percaya sebetulnya membawa beban kritik terselubung. Penggunaan istilah "reset indonesia" serta penyebutan nama tokoh politik secara spesifik menunjukkan adanya narasi informatif yang bersifat skeptis terhadap kondisi politik saat ini, namun disampaikan tanpa kata-kata emosional yang eksplisit. Hal ini membuktikan bahwa kelompok sentimen yang secara konsisten diidentifikasi oleh model SVM sebagai emosi Netral ini tetap mengandung pesan ketidakpuasan substansial yang disampaikan melalui diksi yang lebih formal, deskriptif, dan objektif.

## 5. KESIMPULAN

Berdasarkan hasil penelitian, opini publik terkait mosi tidak percaya DPR RI di media sosial X didominasi oleh sentimen negatif sebesar 8.882 data (45,7%). Pada klasifikasi polaritas sentimen, penerapan algoritma *Support Vector Machine* (SVM) berbasis leksikon dan SMOTE terbukti sangat andal dengan meraih *Test Accuracy* sebesar 85,45%, di mana performa tertinggi dicatatkan oleh sentimen Netral (*F1-score* 0,91) dan Negatif (*F1-score* 0,86). Selanjutnya pada tingkatan emosi spesifik, model SVM mencatatkan *Test Accuracy* sebesar 75,53% pada klasifikasi sepuluh dimensi emosi, dengan emosi Marah (5.488 data) dan Kecewa (873 data) menjadi pendorong utama ketidakpercayaan publik. Kategori Netral (5.968 data) umumnya mencerminkan ketidakpuasan pasif berdiksi formal, sementara visualisasi word cloud menegaskan kata "Rakyat", "DPRRI", dan "Bubar" sebagai poros utama tuntutan publik. Meski demikian, metode *Emotional Lexicon* (*EmoLex*) masih memiliki keterbatasan dalam mendeteksi sarkasme politik, sehingga gaya bahasa ironi sering kali teridentifikasi sebagai emosi Netral.

Untuk penelitian selanjutnya, disarankan memanfaatkan model *Deep Learning* atau *Transformer* (seperti IndoBERT) guna mengatasi ambiguitas makna dan sarkasme politik. Selain itu, eksperimen menggunakan kernel non-linear

pada SVM (seperti RBF atau *Polynomial*) serta penambahan dimensi analisis temporal perlu dilakukan untuk menguji peningkatan akurasi dan memetakan fluktuasi emosi publik secara dinamis.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Irfan Pratama, S.Kom., M.Eng., MCE. MCF. atas bimbingan, waktu, dan arahan akademisnya yang sangat berharga sepanjang penyusunan penelitian ini. Apresiasi mendalam juga disampaikan kepada Edwin Nasli F. atas dukungan finansial, moral, dan motivasi luar biasa yang menjadi fondasi utama kelancaran pendidikan penulis. Terima kasih pula kepada Program Studi Sistem Informasi Universitas Mercu Buana Yogyakarta serta seluruh pihak yang telah memberikan bantuan teknis maupun ruang diskusi ilmiah hingga manuskrip jurnal ini dapat diselesaikan dengan baik.

## DAFTAR PUSTAKA

- [1] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.
- [2] Amirullah, "Kronologi Demo Memprotes DPR Hingga Meluas Berubah Penjarahan," *Tempo.co*. Accessed: Oct. 09, 2025. [Online]. Available: <https://www.tempo.co/politik/kronologi-demo-memprotes-dpr-hingga-meluas-berubah-penjarahan-2065182>
- [3] CNBC Indonesia, "Kronologi Demo 25-29 Agustus: Demo DPR Melebar Jadi Amarah ke Polisi," *CNBC Indonesia*. Accessed: Oct. 09, 2025. [Online]. Available: <https://www.cnbcindonesia.com/news/20250830145726-4-662882/kronologi-demo-25-29-agustus-demo-dpr-melebar-jadi-amarah-ke-polisi>
- [4] A. Paramartha *et al.*, "Analisis Emosional Masyarakat Terhadap Tagar #Kaburajadulu Di Twitter Menggunakan Support Vector Machines," *Integr. Perspect. Soc. Sci. J.*, vol. 2, no. 4, p. 7374, 2025.
- [5] danSyafriandi Afifah Salsabilah Putri\*, Eujeniatul Jannah, Dodi Vionanda, "Implementation Of Text Mining For Emotion Detection Using The Lexicon Method (Case Study: Tweets About Covid-19)," *Telematika*,

- vol. 18, no. 1, p. 49, 2025, doi: 10.31315/telematika.v18i1.4341.
- [6] A. N. Rohman, E. Utami, and S. Raharjo, "Deteksi Kondisi Emosi pada Media Sosial Menggunakan Pendekatan Leksikon dan Natural Language Processing," *Eksplora Inform.*, vol. 9, no. 1, pp. 70–76, 2019, doi: 10.30864/eksplora.v9i1.277.
- [7] A. Hamzah and R. Y. Ariyana, "Klasifikasi Emosi Berbasis Emolex dari Komentar Evaluasi Akademik Mahasiswa," *Techno.Com*, vol. 23, no. 2, pp. 457–468, 2024, doi: 10.62411/tc.v23i2.10058.
- [8] R. A. N. Rizka and Zaehol Fatah, "Analisis Sentimen Komentar YouTube terhadap Tragedi Demo 25 Agustus Menggunakan Pendekatan Lexicon-Based," *J. Mhs. Tek. Inform.*, vol. 4, no. 2, pp. 217–224, 2025, doi: 10.35473/jamastika.v4i2.4525.
- [9] I. T. Cahya, A. Irma, I. Ali, O. Nurdiawan, and Martanto, "ANALISIS SENTIMEN ULASAN APLIKASI DANA DI PLAY STORE MENGGUNAKAN SUPPORT VECTOR," vol. 14, no. 2, 2026.
- [10] F. Amrullah and A. Solichin, "Analisis Emosi Pada Live Chat Youtube 'Mata Najwa: 3 Bacapres Bicara Gagasan' Menggunakan Pendekatan Lexicon dan Algoritma Naive Bayes," *J. Ticom Technol. Inf. Commun.*, vol. 12, no. 3, pp. 121–128, 2024, doi: 10.70309/ticom.v12i3.132.
- [11] D. S. M. Mohammad and D. P. Turney, "NRC Word-Emotion Association Lexicon (aka EmoLex)," Dr. Saif M. Mohammad. Accessed: Oct. 12, 2025. [Online]. Available: <https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>
- [12] Y. A. N. Jannah and R. B. Prasetyo, "Analisis Sentimen dan Emosi Publik pada Awal Pandemi COVID-19 Berdasarkan Data Twitter dengan Pendekatan Berbasis Leksikon," *Semin. Nas. Off. Stat.*, vol. 2022, no. 1, pp. 597–608, 2022, doi: 10.34123/semnasoffstat.v2022i1.1483.
- [13] P. W. Ciady *et al.*, "Bot Discord Dengan Metode Analisis Sentimen Berbasis Leksikon," *J. Mhs. Tek. Inform.*, vol. 8, no. 5, pp. 10493–10498, 2024.
- [14] B. Susarianto, T. Nizami, and T. Informatika, "Analisis Sentimen Publik Pengesahan UU TNI Di Media Sosial X Menggunakan SVM," no. 2, pp. 1257–1269, 2025.
- [15] Z. Rani and B. K. Khotimah, "SAPI DI TWITTER MENGGUNAKAN KOMBINASI METODE K-MEANS DAN SUPPORT VECTOR MACHINE," vol. 13, no. 1, 2025.
- [16] D. C. Rahmadani, S. Khomsah, and M. Y. Fathoni, "Analisis Emosi Wisatawan Menggunakan Metode Lexicon Text Analysis," *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 1, pp. 13–27, 2024, doi: 10.28932/jutisi.v10i1.6690.
- [17] R. Dwiyanaputra, S. I. Murpratiwi, and A. Aranta, "Analisis Sentimen Pengguna Platform Media Sosial X Pada Topik Pemilihan Presiden 2024 Menggunakan Perbandingan Model," vol. 9, no. 1, pp. 626–634, 2025.
- [18] F. Muttakin, N. Andrika, and S. Salsabila, "Sentiment Analysis of Shoe Product Reviews on Indonesian E-Commerce Platform Using Lexicon Based and Support Vector Machine," *J. Tek. Inform.*, vol. 6, no. 2, pp. 839–854, 2025, doi: 10.52436/1.jutif.2025.6.2.3800.
- [19] H. Hoiriyah, N. Qomariya, A. K. Darmawan, M. Walid, and Y. Efenie, "Sentiment Analysis on Lgbt Issues in Indonesia With Lexicon-Based and Support Vector Machine Algorithms," *J. Pilar Nusa Mandiri*, vol. 19, no. 1, pp. 27–36, 2023, doi: 10.33480/pilar.v19i1.4183.
- [20] I. H. Kusuma and N. Cahyono, "Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor," *J. Inform. J. Pengemb. IT*, vol. 8, no. 3, pp. 302–307, 2023, doi: 10.30591/jpit.v8i3.5734.
- [21] A. R. Ismail and Raden Bagus Fajriya Hakim, "Implementasi Lexicon Based Untuk Analisis Sentimen Dalam Menentukan Rekomendasi Pantai Di DI Yogyakarta Berdasarkan Data Twitter," *Emerg. Stat. Data Sci. J.*, vol. 1, no. 1, pp. 37–46, 2023, doi: 10.20885/esds.vol1.iss.1.art5.