

# Klasifikasi Sentimen Masyarakat di Twitter Terhadap Ganjar Pranowo dengan Metode Support Vector Machine

Syaiful Azhar, Yusra\*, Muhammad Fikry, Surya Agustian, Iis Afrianty

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: <sup>1</sup>11950115229@students.uin-suska.ac.id, <sup>2\*</sup>yusra@uin-suska.ac.id, <sup>3</sup>muhammad.fikry@uin-suska.ac.id,

<sup>4</sup>surya.agustian@uin-suska.ac.id, <sup>5</sup>iis.afrianty@uin-suska.ac.id

Email Penulis Korespondensi: yusra@uin-suska.ac.id

**Abstrak**—Klasifikasi sentimen masyarakat terhadap Ganjar Pranowo di Twitter dapat memberikan wawasan tentang popularitas, dukungan, atau kritik yang diterimanya. Penelitian ini bertujuan untuk mengklasifikasikan sentimen masyarakat terhadap Ganjar Pranowo di Twitter menggunakan metode Support Vector Machine (SVM). Data penelitian terdiri dari 4000 tweet yang dikumpulkan dari Twitter. Setelah melalui tahap preprocessing, tweet tersebut diklasifikasikan menggunakan SVM menjadi kelas positif atau negatif. Metode klasifikasi dioptimasi untuk menghasilkan model yang paling optimal dengan cara pengujian pengaruh tahapan seleksi fitur dan SVM parameter tuning. Data dibagi menjadi 80% training (TRAIN\_SET) dan 20% testing (TEST\_SET). Model optimal divalidasi melalui 10% dari TRAIN\_SET yang diambil secara acak untuk data validasi. Variasi eksperimen untuk penelusuran model optimal dilakukan sebanyak 16 experiment. Hasil validasi tertinggi (top rank 4 model) diujikan pada TEST\_SET, menghasilkan F1-score sebesar 84.13%, 84.13%, 84.13% dan 84.13% untuk ID experiment 1, 7, 14 dan 16 secara berturut-turut. Dalam penelitian ini, SVM cukup efektif untuk mengklasifikasikan sentimen tweet terkait Ganjar Pranowo di Twitter.

**Kata Kunci:** Twitter; Ganjar Pranowo; Klasifikasi; Support Vector Machine; Python

**Abstract**—The classification of public sentiment towards Ganjar Pranowo on Twitter can provide insights into his popularity, support, or criticism. This research aims to classify public sentiment towards Ganjar Pranowo on Twitter using the Support Vector Machine (SVM) method. The research data consists of 4000 tweets collected from Twitter. After undergoing preprocessing, these tweets are classified using SVM into positive or negative classes. The classification method is optimized to produce the most optimal model by testing the influence of feature selection stages and SVM parameter tuning. The data is divided into 80% training (TRAIN\_SET) and 20% testing (TEST\_SET). The optimal model is validated using 10% of the randomly selected TRAIN\_SET for validation data. Sixteen experiments are conducted to explore the optimal model, with the highest validation results (top rank 4 models) tested on the TEST\_SET, yielding F1-scores of 84.13%, 84.13%, 84.13%, and 84.13% for experiment IDs 1, 7, 14, and 16, respectively. In this research, SVM proves to be sufficiently effective in classifying sentiment-related tweets about Ganjar Pranowo on Twitter.

**Keywords:** Twitter; Ganjar Pranowo; Classification; Support Vector Machine;

## 1. PENDAHULUAN

Media sosial di gunakan oleh politisi demi memperkuat kepercayaan publik dan menciptakan persepsi positif di kalangan masyarakat. Media sosial menjadi salah satu bentuk komunikasi masa yang paling populer karena memberikan penyebaran informasi yang efisien dan memfasilitasi komunikasi interaktif dua arah, di mana memungkinkan pengguna berinteraksi secara online dengan orang lain secara bersama-sama [1][2][3]. Twitter adalah suatu platform media sosial yang memungkinkan pengguna untuk mencari informasi mengenai beragam topik seperti bisnis, hiburan, ekonomi, politik, dan topik lainnya. Dalam konteks penelitian jurnal, Twitter merupakan sumber potensial untuk memperoleh data opini pengguna terhadap suatu audiens. Melalui postingan *tweet* yang jumlahnya sangat besar, pengguna Twitter memiliki kesempatan untuk mengungkapkan pandangan dan pendapat mereka mengenai peristiwa yang terjadi di sekitar komunitas, termasuk di antaranya produk atau layanan [4][5].

Ganjar Pranowo, Gubernur Jawa Tengah yang menjabat dua periode sejak 23 Agustus 2023, disorot sebagai calon presiden potensial di tahun 2024. Dengan rekam jejaknya yang mencuri perhatian selama kepemimpinannya dan masa sebelumnya sebagai anggota DPR RI, Ganjar Pranowo semakin dikenal dan didekati oleh masyarakat. Polemik dan opini dari berbagai lapisan masyarakat memberikan gambaran menyeluruh, menciptakan momentum untuk menganalisis polarisasi dan opini publik [6]. Ganjar Pranowo menggunakan media sosial untuk komunikasi interaktif dengan masyarakat, membentuk citra dirinya dan mendapatkan keuntungan dalam komunikasi politik. Twitter, selain mempermudah komunikasi politik, juga meningkatkan pemahaman terhadap sudut pandang politik kelompok pengguna media sosial. Dalam penelitian, banyaknya pengguna media sosial memberikan peluang bagi peneliti untuk menganalisis data sentimen dari status atau konten yang diposting oleh mereka [1].

Dalam penelitian ini, dilakukan analisis sentimen terhadap status atau *tweet* pengguna Twitter untuk mengungkapkan opini publik terhadap Ganjar Pranowo. Penggunaan algoritma klasifikasi teks memiliki kelebihan dan kekurangan. Metode SVM dipilih karena mampu memberikan akurasi tinggi dalam mengklasifikasikan pola [7]. Analisis sentimen diperlukan untuk memahami pandangan dan respons masyarakat terhadap Ganjar Pranowo, yang dapat memberikan wawasan tentang popularitasnya, dukungan, atau kritik yang diterimanya. Dengan menganalisis sentimen, penelitian ini dapat mengidentifikasi pola-pola umum dalam pendapat publik terkait tokoh tersebut [8]. Dalam konteks analisis sentimen, SVM digunakan sebagai algoritma klasifikasi untuk mengklasifikasikan *tweet* berdasarkan kategori sentimen positif atau negatif. Tujuan utama analisis sentimen dalam penelitian ini adalah menguji akurasi metode klasifikasi SVM dalam mengkategorikan tweet menjadi sentimen positif atau negatif terkait Ganjar Pranowo. Selain itu, penelitian ini bertujuan untuk menganalisis distribusi sentimen positif dan negatif guna memahami perasaan umum

masyarakat terhadap tokoh tersebut [9]. Penelitian ini sangat penting karena dapat memberikan gambaran akurat tentang dukungan publik terhadap Ganjar Pranowo melalui analisis sentimen terhadap tweet. Selain itu, penelitian ini juga mendeteksi kritik atau kontroversi yang dapat mempengaruhi citra tokoh tersebut. Hasilnya memberikan informasi strategis untuk merancang komunikasi yang lebih efektif, menjaga dukungan positif, dan mengatasi isu-isu kontroversial. Dengan demikian, penelitian ini memiliki urgensi dalam memberikan wawasan mendalam terhadap dinamika opini publik terkait Ganjar Pranowo.

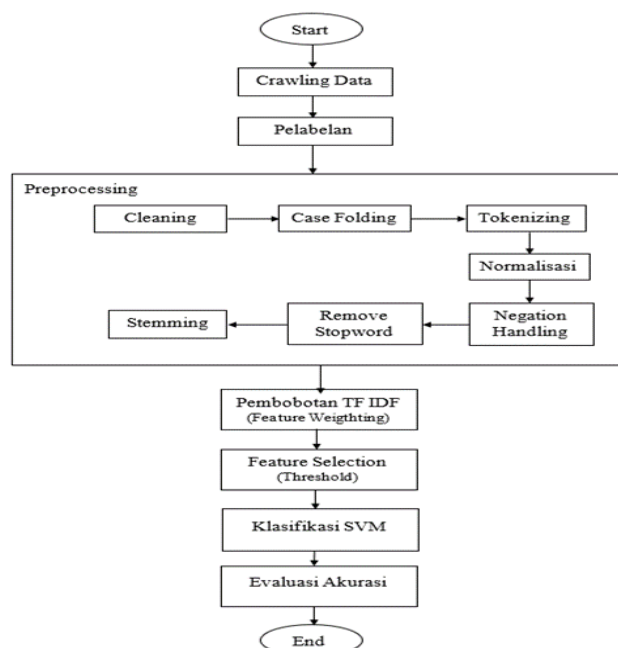
Terdapat beberapa penelitian yang dilakukan dengan membandingkan SVM dengan *Naïve Bayes*, metode klasifikasi SVM menunjukkan akurasi yang lebih tinggi sebesar 81,46% dibandingkan dengan metode klasifikasi *Naïve Bayes* (75,41%) dalam analisis sentimen ulasan bahasa Indonesia di Google Play Store [10]. Penelitian selanjutnya juga mendukung keefektifan algoritma SVM, mencapai akurasi terbaik sebesar 84,37% dengan menggunakan kernel RBF dalam analisis sentimen opini terkait maskapai penerbangan di Twitter [11]. Penelitian selanjutnya menyimpulkan bahwa metode klasifikasi *Naïve Bayes* memberikan akurasi lebih tinggi sebesar 85,59% dibandingkan dengan metode SVM (84,41%) dalam menganalisis sentimen masyarakat terhadap vaksinasi Covid-19 [12]. Penelitian selanjutnya menunjukkan akurasi sebesar 78,18% dalam analisis sentimen terkait sistem ganjil genap di Tol Bekasi menggunakan algoritma SVM [13]. Pada penelitian berikutnya, SVM unggul dibandingkan metode konvensional lainnya, terlihat pada Tabel 4 untuk *binary classification* (HS dan AB) dan *multi-label classification* (kelas level dengan 3 label). Akurasi tertinggi pada kelas HS, AB, dan level dari HS mencapai 84.52%, 87.02%, dan 76.43% dengan *MinMax scaling* pada vektor input *sentence embeddings FastText* dalam penerapan SVM dan *FastText* untuk mendeteksi *Hate Speech* dan *Abusive* di Twitter [14].

Penelitian ini memiliki kesenjangan (GAP) yang signifikan dalam pemrosesan data teks untuk menganalisis opini positif dan negatif tentang Ganjar Pranowo di Twitter. Fokus pada pengembangan kombinasi optimal dalam tahap preprocessing menyoroti kekurangan metode sebelumnya dan menekankan pentingnya memperbaiki kualitas data teks. Keterbatasan yang ada dalam menganalisis opini di media sosial, terutama Twitter, dengan menggunakan SVM, juga menjadi fokus, menunjukkan keinginan untuk meningkatkan ketepatan model dalam konteks tersebut. GAP ini membuka peluang untuk mengisi celah pengetahuan dan meningkatkan metodologi analisis sentimen, dengan harapan bahwa hasil penelitian ini dapat memberikan kontribusi yang lebih besar terhadap pemahaman opini publik terhadap Ganjar Pranowo di platform media sosial.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Tapap ini melibatkan pengumpulan fakta dan informasi dari platform media sosial Twitter, yaitu data *tweet* yang mencantumkan akun Twitter @GanjarPranowo, yang menjadi sumber data utama. Untuk menjalankan penelitian ini, sangat penting untuk memiliki alur proses penelitian yang terstruktur agar penelitian dapat dilakukan sesuai dengan rencana. Tahapan penelitian dapat dilihat pada Gambar 1 sebagai panduan melaksanakan penelitian.



Gambar 1. Tahapan Penelitian

Gambar 1 menunjukkan tahapan awal dalam pengolahan data *tweet* mentah setelah *crawling*. Proses berlanjut dengan pelabelan (positif/negatif), lalu *preprocessing* melibatkan *cleaning*, *case folding*, *tokenizing*, normalisasi,

*negation handling*, penghapusan *stopword*, dan *stemming*. Setelah *preprocessing*, data masuk ke tahap pembobotan TF.IDF, diikuti *Feature Selection* dengan *threshold*, lalu Klasifikasi SVM digunakan, dan terakhir, model dievaluasi untuk akurasi. Tahapan ini membentuk rangkaian pengolahan data *tweet* dari awal hingga evaluasi model.

## 2.2 Pengumpulan Data

Pada tahap pengumpulan data, dilakukan proses *crawling* menggunakan API Twitter dengan menggunakan modul *tweepy* dalam bahasa pemrograman Python. Data yang diperoleh adalah kumpulan *tweet* yang ditujukan kepada akun Twitter @GanjarPranowo, dengan menggunakan kata kunci yang mencantumkan *username* tersebut. Selanjutnya, dataset *tweet* yang diperoleh disimpan dalam format CSV (*Comma Separated Values*) sebagai bentuk penyimpanan yang lebih terstruktur dan siap untuk diproses lebih lanjut. Jumlah total data yang berhasil dikumpulkan untuk penelitian ini adalah sebanyak 4000 data *tweet*. Yang diambil pada tanggal 29 November 2022 sampai tanggal 3 Maret 2023 [4][15].

## 2.3 Pelabelan Data

Pada tahap ini, dilakukan pelabelan terhadap data *tweet* yang telah diperoleh. Tujuannya adalah menilai apakah *tweet* itu termasuk dalam kategori positif atau negatif. Misalnya, jika *tweet* tersebut mengandung sentimen positif terhadap Ganjar Pranowo, maka akan diberi label positif. Sebaliknya, jika *tweet* tersebut mengandung sentimen negatif, maka akan diberi label negatif. Setelah proses pelabelan, ditemukan bahwa terdapat 2000 *tweet* yang diberi label positif dan 2000 *tweet* lainnya yang diberi label negatif.

## 2.4 Preprocessing

Tahap *preprocessing* bertujuan untuk membersihkan dan mempersiapkan data *tweet* sebelum dilakukan analisis lebih lanjut [16]. Beberapa tahap yang biasa dilakukan dalam *preprocessing* adalah sebagai berikut:

- Cleaning*: Sederhanakan teks dengan menghapus tanda baca, link, dan karakter khusus yang tidak diperlukan.
- Case Folding*: Consistenkan pemrosesan dengan meratakan seluruh teks ke huruf kecil atau huruf besar.
- Tokenizing*: Pecah teks menjadi unit-unit kecil seperti kata-kata atau token untuk analisis lebih lanjut.
- Normalisasi: Ubah kata-kata menjadi bentuk dasar atau normal, contohnya pada kata yang *typo* atau singkatan yang tidak baku menjadi kata baku.
- Negation Handling*: Temukan dan ubah kata-kata yang dipengaruhi oleh negasi, seperti "tidak" atau "bukan".
- Remove Stopword*: Hapus kata-kata umum yang kurang informatif, seperti "dan", "di", atau "untuk".
- Stemming*: Ubah kata-kata ke bentuk dasar dengan menghilangkan akhiran atau awalan khusus.

## 2.5 Feature Weighting

Setelah tahap *preprocessing*, Kata-kata dalam *tweet* akan diberikan bobot menggunakan metode *Term Frequency-Inverse Document Frequency* (TF.IDF). TF.IDF mengukur seberapa signifikan suatu kata dalam *tweet* berdasarkan frekuensi kemunculannya dalam *tweet* tersebut dan kemunculannya di seluruh dataset.

## 2.6 Feature Selection (Threshold)

Tahap ini melibatkan seleksi fitur berupa kata-kata/token dari *Bag Of Word*(BOW) yang berpengaruh signifikan dalam klasifikasi. Fitur-fitur yang memenuhi  $TF.IDF \text{ threshold} \geq 0,001$  akan dipilih sebagai fitur penting yang digunakan dalam klasifikasi.

## 2.7 Klafisikasi

Pada tahap ini, data dibagi menjadi dua bagian, yaitu 80% sebagai data *training* (*TRAIN\_SET*) dan 20% sebagai data *testing* (*TEST\_SET*). Data *training* digunakan dalam melatih algoritma dan mengembangkan model, dan data *testing* digunakan sebagai referensi dalam menguji kinerja algoritma yang telah dilatih sebelumnya. Setelah itu, dilakukan pemodelan menggunakan metode SVM [4][17]. Pada tahap pemrosesan menggunakan kernel RBF (*Radial Basis Function*) dan *grid search* untuk mencari pasangan parameter *C* dan *gamma* terbaik. Kemudian menggunakan teknik *K-fold Cross Validation*. Yang digunakan untuk memisahkan data latih dan data *testing* agar tidak terjadi perulangan.

## 2.8 Evaluasi

Evaluasi model analisis umumnya melibatkan perbandingan antara nilai positif dan negatif untuk menghasilkan kesimpulan dari hasil klasifikasi. Untuk menghitung nilai tersebut, data yang digunakan merupakan data *testing* yang sudah diketahui sentimennya. Dengan membandingkan sentimen yang diprediksi oleh model dengan sentimen yang sebenarnya, dapat dievaluasi sejauh mana model berhasil dalam mengklasifikasikan dengan akurasi.

# 3. HASIL DAN PEMBAHASAN

## 3.1 Pengumpulan Data

Data *tweet* yang diperoleh melalui indeksasi di Twitter melewati tahap seleksi data. Tahap ini, dipilih atribut yang akan digunakan untuk membersihkan data dan klasifikasi lebih lanjut. Jumlah data adalah 4000 data kemudian data akan dibagi menjadi 80% data (3200 *tweet*) sebagai *TRAIN\_SET* dan 20% data (800 *tweet*) sebagai *TEST\_SET*. Berikut jumlah

data *training* dan *testing*. *TEST\_SET* sebagai data uji, tidak pernah terlihat selama proses *training* dan pembentukan fitur *Bag of Word* dan *TF.IDF*. Kemudian untuk proses pencarian model yang optimal, dari *TRAIN\_SET* yang berjumlah 3200 *tweet*, akan dibagi lagi dengan komposisi 90% data *training* dan 10% data validasi. Pada 90% data *training* tersebut akan digunakan untuk mencari parameter *C* dan *gamma* terbaik dan mendapatkan *K-fold Cross Validation* dengan melakukan perulangan menggunakan nilai *K* = 10 (*10-fold cross validation*). Data validasi digunakan untuk melihat performa sejauh mana keakuratan mesin dalam klasifikasi. Tabel 1 Berikut contoh hasil *Crawling Data*.

**Tabel 1.** Contoh Hasil *Crawling Data*

No	Komentar	Sentimen
1	b'@GanjaranApp @ganjarpranowo sehat selalu pak Ganjar Pranowo banyak yg mendukung bapak'	Positif
2	b'@BijiPot @ch_chotimah2 @ganjarpranowo Tidak pantas bicara negara sendiri di negri orang yg tidak sesuai kenyataan, khianat namanya. <a href="https://t.co/vPbizV21QL">https://t.co/vPbizV21QL</a>	Negatif

Pada Tabel 1 merupakan contoh hasil *crawling* data yang sudah dilabel (positif/negatif) dan data tersebut akan digunakan untuk tahap *preprocessing*.

### 3.2 Preprocessing Data

*Preprocessing* data melibatkan pemrosesan dari hasil *crawling* menjadi data bersih yang siap untuk klasifikasi. Berikut langkah-langkah *preprocessing* [18]. Kalimat (1) adalah kalimat awal sebelum proses, dan kalimat (2) adalah kalimat hasil pada proses yang bersangkutan

#### a. Cleaning

Mengubah data yang tidak perlu, tidak akurat, atau tidak sesuai tujuan analisis, seperti URL, tanda baca, simbol, dan emotikon. Berikut menunjukkan hasil proses *Cleaning*

1. b'@BijiPot @ch\_chotimah2 @ganjarpranowo Tidak pantas bicara negara sendiri di negri orang yg tidak sesuai kenyataan, khianat namanya. <https://t.co/vPbizV21QL>
2. Tidak pantas bicara negara sendiri di negri orang yg tidak sesuai kenyataan, khianat namanya

#### b. Case Folding

Bertujuan untuk menghilangkan perbedaan huruf besar dan huruf kecil dalam teks tertentu. berikut hasil proses *case folding* [17][10].

1. Tidak pantas bicara negara sendiri di negri orang yg tidak sesuai kenyataan, khianat namanya.
2. tidak pantas bicara negara sendiri di negri orang yg tidak sesuai kenyataan, khianat namanya.

#### c. Tokenizing

Proses memecah teks atau urutan karakter menjadi unit yang lebih kecil, yang disebut token. Token dapat berupa kata, frasa, kalimat, atau bahkan unit subkata individual, bergantung pada tingkat perincian yang diinginkan untuk analisis atau pemrosesan. berikut hasil tahapan *Tokenizing* [19].

1. tidak pantas bicara negara sendiri di negri orang yg tidak sesuai kenyataan, khianat namanya.
2. "tidak", "pantas", "bicara", "negara", "sendiri", "di", "negri", "orang", "yg", "tidak", "sesuai", "kenyataan", "khianat", "namanya".

#### d. Normalisasi

Proses ini akan memperbaiki ejaan yang salah pada kata pada *tweet*. berikut hasil tahapan Normalisasi.

1. "tidak", "pantas", "bicara", "negara", "sendiri", "di", "negri", "orang", "yg", "tidak", "sesuai", "kenyataan", "khianat", "namanya".
2. "tidak", "pantas", "bicara", "negara", "sendiri", "di", "negri", "orang", "yang", "tidak", "sesuai", "kenyataan", "khianat", "namanya".

#### e. Negation Handling

Pemrosesan teks yang fokus pada penanganan kalimat yang mengandung negasi. Dalam analisis sentimen, negasi dapat membalikkan makna kata-kata atau kalimat secara keseluruhan. Berikut hasil tahapan *Negation Handling*.

1. "tidak", "pantas", "bicara", "negara", "sendiri", "di", "negri", "orang", "yang", "tidak", "sesuai", "kenyataan", "khianat", "namanya"
2. layak bicara negara sendiri di negri orang yang tidak sesuai kenyataan khianat namanya

#### f. Stopword Removal

Tahap penghilangan kata-kata tidak penting, seperti kata sambung dan kata keterangan, dari teks yang tidak memberikan kontribusi penting dalam pemodelan data. Berikut Hasil dari tahap *Stopword Removal* [19].

1. layak bicara negara sendiri di negri orang yang tidak sesuai kenyataan khianat Namanya
2. layak bicara negara negri orang sesuai kenyataan khianat

#### g. Stemming

Proses dalam pemrosesan teks yang memiliki tujuan mengubah kata menjadi bentuk dasarnya atau kata dasar [15]. Proses ini melibatkan penghapusan akhiran atau awalan kata sehingga hanya tersisa akar kata yang lebih umum. Berikut Hasil dari proses *stemming*.

1. layak bicara negara negri orang sesuai kenyataan khianat
2. layak bicara negara negri orang sesuai kenyataan khianat

### 3.3 Feature Weighting

Untuk pembobotan fitur menggunakan TF.IDF, menggunakan data *training* untuk perhitungan TF.IDF menggunakan *binary Sklearn TfidfVectorizer*

### 3.4 Feature Selection

Proses *feature selection*, memilih kata-kata (fitur/token) yang signifikan digunakan untuk klasifikasi berdasarkan *threshold*. Missal pada *experiment 4*, dimana fitur yang dipilih adalah *stemming*, *stopword*, normalisasi, dan *negation handling* aktif, diperoleh jumlah feature (token BOW) sebanyak 5980 kata. Setelah diterapkan TF.IDF *threshold* 0,001, diperoleh token sebanyak 136 kata.

## 3.5 Klasifikasi

### 3.5.1 Experiment Setup

*Experiment* dilakukan untuk menemukan model SVM yang paling optimal, yaitu yang dapat menghasilkan performa terbaiknya. Pada tahap ini akan dilakukan beberapa *experiment* untuk menyelidiki pengaruh penerapan beberapa langkah *text preprocessing*, yaitu penerapan *stemming*, normalisasi, *negation handling* dan penghapusan *stopword*. Komposisi dari tahap *preprocessing* tersebut dijelaskan pada Tabel 2 *Experiment Setup* berikut. dengan tanda centang menyatakan tahapan tersebut diterapkan, dan tanda silang menyatakan tahapan tersebut tidak diterapkan dalam tahap *text preprocessing*.

**Tabel 2.** *Experiment Setup*

<i>ID Experiment</i>	<i>Stemming</i>	<i>Stopword</i>	<i>Normalisasi</i>	<i>Negation Handling</i>
1	✓	✗	✗	✗
2	✓	✓	✗	✗
3	✓	✓	✓	✗
4	✓	✓	✓	✓
5	✗	✓	✗	✗
6	✓	✗	✓	✗
7	✗	✗	✗	✓
8	✗	✓	✓	✗
9	✗	✓	✓	✓
10	✓	✗	✓	✓
11	✓	✓	✗	✓
12	✗	✗	✓	✓
13	✓	✗	✓	✗
14	✗	✗	✗	✗
15	✗	✗	✓	✗
16	✓	✗	✗	✓

Pada Tabel 2 *Experiment Setup* terlihat dilakukan proses menerapkan dan tidak menerapkan beberapa proses *preprocessing* dengan tanda centang menyatakan tahapan tersebut diterapkan, dan tanda silang menyatakan tahapan tersebut tidak diterapkan dalam tahap *text preprocessing*.

### 3.5.2 Parameter Tuning

Pada setiap *experiment* yang dilakukan sesuai Tabel 2, dijalankan juga pencarian parameter optimal dari SVM. Banyak penelitian *men-tuning* parameter *C* dan *gamma*, dan jenis kernel SVM yang dipakai untuk klasifikasi. Dalam penelitian ini, kernel yang akan dipakai adalah RBF, karena berdasarkan penelitian klasifikasi sentimen [20], [21] dan deteksi *hate speech* [14], kernel RBF paling baik dipakai untuk klasifikasi. Sehingga dalam penelitian ini hanya parameter  $C=\{1,10,100,1000\}$  dan  $gamma=\{0.1, 0.01, 0.001, 0.0001\}$  saja yang akan di *tuning* menggunakan *grid search*[21] Hasil penelusuran model optimal melalui proses seleksi fitur dan parameter *tuning* yang telah dilakukan dengan *10-fold Cross Validation* dapat dilihat secara lengkap pada Tabel 3. Parameter *C* dan *Gamma* adalah parameter terbaik dari masing-masing skema *experiment*, dan akurasi rata-rata yang diperoleh dari validasi terhadap 10 percobaan.

**Tabel 3.** Parameter *Tuning* dan Akurasi *10-fold Cross Validation*

<i>ID Experiment</i>	<i>C</i>	<i>Gamma</i>	Akurasi <i>K-fold Cross Validation</i>
1	1000	0.0001	86.52%

2	1000	0.0001	84.72%
3	1000	0.0001	84.65%
4	1000	0.0001	84.65%
5	1000	0.0001	84.40%
6	10	0.001	87.11%
7	1000	0.0001	86.66%
8	1000	0.0001	84.16%
9	1000	0.0001	84.16%
10	10	0.001	87.11%
11	1000	0.0001	84.72%
12	10	0.001	87.11%
13	10	0.001	87.11%
14	1000	0.0001	86.66%
15	10	0.001	87.11%
16	1000	0.0001	86.52%

Pada Tabel 3. Parameter *C* dan *Gamma* dapat dilihat bahwa hasil penelusuran model optimal melalui proses seleksi fitur dan parameter *tuning* yang telah dilakukan dengan *10-fold Cross Validation* menghasilkan parameter terbaik dari masing-masing skema *experiment*, dan akurasi rata-rata yang diperoleh dari validasi terhadap 10 percobaan.

### 3.5.3 Penelusuran Model Optimal

Setelah pasangan parameter terbaik didapatkan, maka parameter tersebut digunakan kembali untuk melakukan pelatihan pada data *training* (90% dari data *TRAIN\_SET*), dan hasilnya digunakan untuk memprediksi data validasi (10% dari data *TRAIN\_SET*). Hasil evaluasi training dan validasi tersebut disajikan dalam Tabel 4.

**Tabel 4** . Evaluasi terhadap Data *Training* dan Data Validasi

<i>ID Experiment</i>	<i>Data Training</i>		<i>Data Validasi</i>	
	Akurasi	<i>F1-score</i>	Akurasi	<i>F1-score</i>
1	90.97%	90.95%	<b>89.68%</b>	<b>89.66%</b>
2	87.70%	87.61%	87.50%	87.43%
3	88.22%	88.17%	87.18%	87.11%
4	88.22%	88.17%	87.18%	87.11%
5	87.81%	87.73%	86.56%	86.48%
6	91.56%	91.54%	88.43%	88.42%
7	90.93%	90.91%	<b>88.75%</b>	<b>88.73%</b>
8	88.22%	88.18%	86.56%	86.49%
9	88.22%	88.18%	86.56%	86.49%
10	91.56%	91.54%	88.43%	88.42%
11	87.70%	87.61%	87.50%	88.43%
12	91.52%	91.51%	88.12%	88.11%
13	91.56%	91.54%	88.43%	88.42%
14	90.93%	90.91%	<b>88.75%</b>	<b>88.73%</b>
15	91.52%	91.51%	88.12%	88.11%
16	90.97%	90.95%	<b>89.68%</b>	<b>89.66%</b>

Pada Tabel 4 dapat dilihat bahwa terdapat 3 akurasi tertinggi dan kombinasi yang memiliki nilai tertinggi yaitu pada *ID experiment* ke-1 dengan nilai akurasi data *training* sebesar 90.97%, *F1 Score* 90.95%, Akurasi Data Validasi sebesar 89.68%, *F1 Score* 89.66%, *ID experiment* ke-7 dengan nilai akurasi data *training* sebesar 90.93%, *F1 Score* 90.91%, Akurasi Data Validasi sebesar 88.75%, *F1 Score* 88.73%, pada *ID experiment* ke-14 dengan nilai akurasi data *training* sebesar 90.93%, *F1 Score* 90.91%, Akurasi Data Validasi sebesar 88.75%, *F1 Score* 88.73%, dan pada *ID experiment* ke-16 dengan nilai akurasi data *training* sebesar 90.97%, *F1 Score* 90.95%, Akurasi Data Validasi sebesar 89.68%, *F1 Score* 89.66%. Maka model yang paling optimal adalah model *ID experiment* 1, 7 diikuti oleh *ID experiment* 14 dan 16.

### 3.5.4 WordCloud

*WordCloud* adalah representasi visual dari frekuensi kata yang terdapat dalam suatu koleksi dokumen. *WordCloud* Digunakan untuk memberikan representasi visual tentang kata-kata paling sering muncul atau penting pada teks tersebut [22]. Hasil *WordCloud* pada data sentimen Ganjar setelah melalui tahap *preprocessing* dan *feature selection* pada *id experiment* 14 (model optimal terpilih) ditunjukkan pada Gambar 2.



- [1] S. Rahmah, "Personal Branding Ganjar Pranowo untuk Membangun Komunikasi Politik di Media Sosial Instagram," *Jurnal Interaksi : Jurnal Ilmu Komunikasi*, vol. 5, no. 1, pp. 94–101, 2021, doi: 10.30596/interaksi.v5i1.5584.
- [2] D. Fatma Sjoraida, R. Dewi, A. Noorlistyanto Adi, and A. Kirana Dipa, "Penggunaan Media Sosial Dalam Membangun Reputasi Anggota Legislatif di Jawa Barat," *PROfesi Humas*, vol. 6, no. 1, pp. 89–110, 2021.
- [3] I. M. P. Gede, P. Pasek, O. Mahawardana, and P. R. Nurbawa, "Analisis Sentimen Berdasarkan Opini Dari Media Sosial Twitter Terhadap 'Figure Pemimpin' Menggunakan Python," *Jurnal Manajemen Dan Teknologi Informasi*, vol. 13, no. 1, pp. 22–28, 2023.
- [4] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *Jurnal Media Informatika Budidarma*, vol. 4, no. 3, pp. 650–658, Jul. 2020, doi: 10.30865/mib.v4i3.2181.
- [5] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "Penerapan Algoritma SVM Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," *Eductic - Scientific Journal of Informatics Education*, vol. 7, no. 1, pp. 1–11, 2020, doi: 10.21107/edutic.v7i1.8779.
- [6] D. Sandi, E. Utami, and Kusnawi, "Analisis Sentimen Publik Terhadap Elektabilitas Ganjar Pranowo Di Tahun Politik 2024 Di Twitter Dengan Algoritma KNN Dan Naïve Bayes," *Jurnal Media Informatika Budidarma*, vol. 7, no. 3, pp. 1097–1108, 2023, doi: 10.30865/mib.v7i3.6298.
- [7] N. Aprilia Putri, H. April Lia, H. Shofa Shofia, T. Tukino, and P. Bayu, "Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM Pada Media Sosial Twitter," *JOINTECS (Journal of Information Technology and Computer Science)*, vol. 8, no. 1, pp. 11–18, 2023.
- [8] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru di Twitter Menggunakan Algoritma Klasifikasi," *Jurnal Teknoinfo*, vol. 14, no. 2, pp. 115–124, Jul. 2020, doi: 10.33365/jti.v14i2.679.
- [9] O. Zoellanda A.Tane, K. Muslim Lhaksmana, and F. Nhita, "Analisis Sentimen Pada Twitter Tentang Calon Presiden 2019 Menggunakan Metode SVM (Support Vector Machine)," *eProceedings of Engineering*, vol. 6, no. 2, pp. 9716–9725, 2019.
- [10] L. B. Ilmawan and M. A. Mude, "Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes Untuk Analisis Sentimen Pada Ulasan Tekstual di Google Play Store," *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, pp. 154–161, 2020, doi: 10.33096/ilkom.v12i2.597.154-161.
- [11] A. M. Pravina, I. Cholissodin, and P. P. Adikara, "Analisis Sentimen Tentang Opini Maskapai Penerbangan Pada Dokumen Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 3, pp. 2789–2797, 2019.
- [12] B. Laurensz and S. Eko, "Analisis Sentimen Masyarakat Terhadap Tindakan Vaksinasi Dalam Upaya Mengatasi Pandemi Covid-19," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 10, no. 2, pp. 118–123, 2021, doi: 10.22146/jnteti.v10i2.1421.
- [13] H. S. Utama, D. Rosiyadi, B. S. Prakoso, and D. Ariadarma, "Analisis Sentimen Sistem Ganjil Genap di Tol Bekasi Menggunakan Algoritma Support Vector Machine," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 3, no. 2, pp. 243–250, 2019, doi: 10.29207/resti.v3i2.1050.
- [14] A. Zikri, A. Zikri, and S. Agustian, "Penerapan Support Vector Machine dan FastText untuk Mendeteksi Hate Speech dan Abusive pada Twitter," *Jurnal Media Informatika Budidarma*, vol. 7, no. 1, pp. 436–443, 2023, doi: 10.30865/mib.v7i1.5408.
- [15] 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, 2021, doi: 10.25126/jtiik.0813944.
- [16] D. Anjas Ramadhan, "Analisis Sentimen Program Acara di SCTV Pada Twitter Menggunakan Metode Naive Bayes dan Support Vector Machine," *eProceedings of Engineering*, vol. 6, no. 2, pp. 9736–9743, 2019.
- [17] B. Pamungkas, "Analisis Sentimen Twitter Menggunakan Metode Support Vector Machine (SVM) pada Kasus Benih Lobster 2020," *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, vol. 3, no. 2, pp. 10–20, 2021.
- [18] M. I. Petiwi, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine," *Jurnal Media Informatika Budidarma*, vol. 6, no. 1, pp. 542–550, Jan. 2022, doi: 10.30865/mib.v6i1.3530.
- [19] R. Wati and S. Ernawati, "Analisis Sentimen Persepsi Publik Mengenai PPKM Pada Twitter Berbasis SVM Menggunakan Python," *Jurnal Teknik Informatika Unika St. Thomas (JTIUST)*, vol. 06, no. 02, pp. 241–247, 2021.
- [20] M. Sahbuddin and S. Agustian, "Support Vector Machine Method with Word2vec For Covid-19 Vaccine Sentiment Classification On Twitter," *Journal Of Informatics And Telecommunication Engineering*, vol. 6, no. 1, pp. 288–297, Jul. 2022, doi: 10.31289/jite.v6i1.7534.
- [21] M. M. Kusair and S. Agustian, "SVM Method With FastText Representation Featurefor Classification Of Twitter Sentiments Regarding The Covid-19 Vaccination Program," *Jurnal Teknologi Informasi Dan Komunikasi Digital Zone*, vol. 13, no. 1, Mei 2022, pp. 140–150, 2022.
- [22] I. B. G. Sarasvananda, D. Selivan, M. L. Radhitya, and I. N. T. A. Putra, "Analisis Sentimen Pada Pembelajaran Daring Di Indonesia Melalui Twitter Menggunakan Naïve Bayes Classifier," *SINTECH Journal*, vol. 5, no. 2, pp. 227–233, 2022.