

Klasifikasi Sentimen *Support Vector Machine* Berbasis Optimasi Menyambut Pemilu 2024

Support Vector Machine Sentiment Classification Based on Optimization Welcoming 2024 Election

William Silalahi^{1*}, Adi Hartanto²

^{1,2}Program Studi Teknik Informatika, Fakultas Ilmu Komputer,
Universitas Mercu Buana
Jl. Raya Kranggan No.6, Jatisampurna 17433, Indonesia

*Corresponding author: 41519210046@student.mercubuana.ac.id

ABSTRAK

DOI:
[10.30595/jrst.v7i2.18133](https://doi.org/10.30595/jrst.v7i2.18133)

Histori Artikel:

Diajukan:
22/06/2023

Diterima:
16/08/2023

Diterbitkan:
15/09/2023

Ajang pesta demokrasi di tahun 2024 yang disertai dengan narasi politik akan kembali diadakan. Beragam komentar negatif dan isu hoaks mulai bermunculan di media sosial untuk menjatuhkan pihak oposisi. Hal ini menimbulkan ketidakpercayaan masyarakat hingga muncul golongan putih. Oleh karena itu, klasifikasi sentimen berdasarkan opini di *Twitter* dilakukan agar mengetahui bagaimana pandangan masyarakat mengenai fenomena ini. Algoritma yang dipakai adalah *Support Vector Machine* dan *Particle Swarm Optimization* dengan metode tambahan seperti *Term Frequency-Inverse Document Frequency* dalam pembuatan vektor dan *Synthetic Minority Oversampling Technique* untuk menyeimbangkan data pada setiap kelas. Karena algoritma machine learning tersebut bersifat supervised learning, maka pelabelan otomatis dilakukan menggunakan *Valence Aware Dictionary sEntiment Reasoner* sebagai data latih. *WordCloud* sebagai sarana pendukung dalam mempersiapkan pemilu tahun depan juga diterapkan. Hasil pelabelan otomatis mendapatkan 1000 data sentimen positif dan 400 data sentimen negatif. Perolehan akurasi *Support Vector Machine* tanpa optimasi dengan metode pengujian *split* 70% data latih dan 30% data uji adalah sebesar 87.33%. Sedangkan persentase akurasi *Support Vector Machine (Particle Swarm Optimization)* adalah 87.50%. *WordCloud* sentimen positif dan negatif menunjukkan bahwa masyarakat mendukung adanya pemilu 2024 meskipun ada yang kurang menyukai publik figur atau partai politik tertentu.

Kata Kunci: Pemilu, PSO, SVM, VADER, WordCloud

ABSTRACT

The democratic party event in 2024 which is accompanied by a political narrative will be held again. Various negative comments and hoax issues began to appear on social media to bring down the opposition. This created distrust in society and led to the emergence of non-voters. Therefore, the classification of sentiments based on opinions on *Twitter* is carried out in order to find out how the public views this phenomenon. The algorithms used are *Support Vector Machine* and *Particle Swarm Optimization* with additional methods such as *Term Frequency-Inverse Document Frequency* in vector creation and *Synthetic Minority Oversampling Technique* to balance data in each class. Because the machine learning algorithm is supervised learning, automatic labeling is done using the *Valence Aware Dictionary Sentiment Reasoner* as training data. *WordCloud* as a supporting tool in preparing for next year's elections is also implemented. Automatic labeling results get 1000 positive sentiment data and 400 negative sentiment data. The accuracy of the *Support Vector Machine* without optimization with the *split* test method of 70% training data and 30% test data is 87.33%. While the percentage accuracy of *Support Vector Machine (Particle Swarm Optimization)* is 87.50%.

WordCloud positive and negative sentiments show that the public supports the 2024 election even though there are those who do not like certain public figures or political parties.

Keywords: Election, PSO, SVM, VADER, WordCloud

1. PENDAHULUAN

Pemilihan umum (pemilu) akan diadakan pada tahun 2024 mendatang. Konsep Demokrasi dijalankan sesuai kebijakan UUD NRI 1945 Pasal 1 ayat (2), menyatakan bahwa “Kedaulatan berada di tangan rakyat dan dilaksanakan menurut Undang-Undang Dasar” (Diniyanto, 2022). Setiap 5 tahun sekali, para kandidat terpilih dilibatkan untuk mengisi kekosongan jabatan di bidang Eksekutif dan Legislatif (Juanita, 2020). Para kandidat partai politik memperkenalkan diri beserta kebijakan-kebijakan yang akan diimplementasikan pada saat terpilih kepada media agar dapat memperoleh suara dari masyarakat luas.

Beragam komentar negatif maupun penggiringan opini mulai mewarnai dunia perpolitikan di media sosial (Rohim & Wardana, 2019). Isu hoaks beserta konten yang berbau suku, agama, ras, dan antar golongan sering beredar di laman media sosial dengan tujuan memperburuk kandidat atau partai politik oposisi (Juanita, 2020). Di sisi lain, ada sebagian kandidat politisi atau partai yang menawarkan praktik finansial politik, perdagangan suara, dan lain-lain (Margiansyah, 2019). Akibatnya minat dan kepercayaan masyarakat terhadap partai politik menurun hingga memunculkan golongan putih (golput). Perangin (2018) menyebutkan bahwa jumlah golput dalam pemilu terus meningkat. Mulai dari tahun 1999 yaitu sebesar 10.21%, 23.345% di tahun 2004, 39.10% di tahun 2009 (Rohim & Wardana, 2019), 30.42% pada tahun 2014 (Yusri et al., 2019), dan penurunan persentase pada pemilu 2019 menjadi 18.03% (Widi, 2022). Salah satu indikator penilaian pemilu adalah partisipasi masyarakat itu sendiri (Juanita, 2020). Jadi ketika masyarakat yang berperan aktif sedikit, maka demokrasi telah gagal dijalankan dengan baik.

Terdapat beberapa algoritma yang digunakan dalam klasifikasi sentimen, diantaranya yaitu algoritma *Long Short Term Memory*, *Logistic Regression*, *K-Nearest Neighbor*, *Naive Bayes*, dan *Support Vector Machine* (Chiny et al., 2021). Penelitian klasifikasi sentimen dengan algoritma *Naive Bayes* dan *SMOTE* pada dataset “*LG G7 mobile unboxing*” berjumlah 6248 komentar mendapatkan akurasi sebesar 79.78% diikuti dengan *F1-Score* 83.72% (V. D, 2019). Lalu penelitian (Sautomo et al., 2020) pengelompokkan sentimen dari kata kunci

“Mudik” dengan pelabelan leksikon *Valence Aware Dictionary and sEntiment Reasoner*. Algoritma klasifikasi *Support Vector Machine* mendapatkan akurasi 87%. Sedangkan 82% untuk algoritma *Naive Bayes*. Kemudian dengan optimasi *Particle Swarm Optimization* terhadap *Support Vector Machine* dan *Naive Bayes* dari kata kunci “@WHO covid” setelah perlakuan *K-Fold Cross Validation* mendapatkan selisih akurasi sebesar 7.1% dengan persentase akurasi keduanya adalah 76.9%:69.8% pada penelitian (Hafidz & Liliana, 2021).

Contoh lain dari penerapan algoritma *Support Vector Machine* yaitu pada penelitian sentimen mengenai pemilihan tiga tokoh kandidat presiden di tahun 2024 dengan persentase akurasi *K-Fold Cross Validation* melampaui perolehan tingkat akurasi *Naive Bayes*, dimana mencapai 98.61%, 94.65%, dan 81.75% (Nardilasari et al., 2023). Selain itu, penelitian klasifikasi sentimen bertemakan Pemilu Indonesia di tahun 2019 menggunakan algoritma *Naive Bayes* pada dataset *Twitter* dengan rasio data latih dan data uji 70%:30% memperoleh akurasi data latih sebesar 81% dan 76% terhadap data uji (Juanita, 2020). Sedangkan klasifikasi sentimen mengenai pemilu di tahun 2024 dengan menggunakan algoritma *Long Short Term Memory* 80% *data training* dan 20% *data testing* mendapatkan rata-rata akurasi 78% (Firdlous et al., 2023). Adapun penelitian klasifikasi sentimen yang menerapkan algoritma *Support Vector Machine* dari dataset opini pemilu di Amerika tahun 2012, dimana tingkat akurasinya mencapai 93.20% (Chauhan et al., 2021).

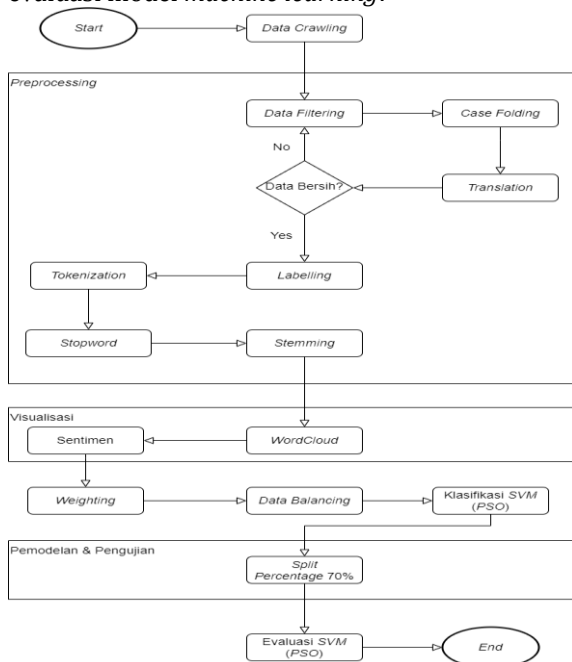
Berdasarkan pemaparan sebelumnya, maka penelitian ini mengangkat topik klasifikasi sentimen untuk mengetahui pandangan masyarakat mengenai pemilu 2024 yang akan datang. Dataset diambil dari *Twitter* karena total pengguna yang mencapai 145 juta di tahun 2019 dan kian meningkat (Juanita, 2020). Jumlah kelas sentimen yang digunakan ada 2, yaitu sentimen positif dan negatif. Kedua kelas ini akan diberikan secara otomatis berdasarkan kamus yang dimiliki oleh *Valence Aware Dictionary sEntiment Reasoner (VADER)*. Kelas-kelas ini akan dibagi menjadi data latih dan data uji dengan perbandingan 70:30 pada algoritma *Support Vector Machine (SVM)* dan optimasi *Particle Swarm Optimization (PSO)*. *SVM* dan *PSO* dipilih karena tingginya perolehan tingkat

akurasi dibandingkan algoritma lainnya dari beberapa penelitian terdahulu.

Dengan adanya penelitian ini, diharapkan algoritma SVM (PSO) dan WordCloud dalam klasifikasi sentimen berdasarkan opini masyarakat di media sosial terkait pemilu 2024 dapat digunakan sebagai acuan bagi pemerintah dalam mempersiapkan pemilu di tahun tersebut. Kata-kata dalam tweets yang menjadi tren pengaruh suatu sentimen dapat dilihat dan dianalisis untuk memberikan penilaian terbaik.

2. METODE PENELITIAN

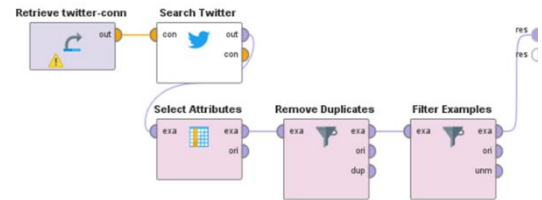
Metode penelitian ini adalah penelitian kuantitatif berdasarkan hasil perhitungan statistik dari pemodelan klasifikasi yang dilakukan dengan mengumpulkan 1400 data pada media sosial Twitter, dimana 1000 data merupakan sentimen positif dan 400 data yang lain merupakan sentimen negatif. Ketidakseimbangan data dapat mempengaruhi proses pelatihan pada model data mining karena beberapa instances pada kelas minoritas dapat dianggap sebagai gangguan dalam pembacaan pola data secara umum (Obiedat et al., 2022). Untuk menyeimbangkan kedua kelas, maka metode SMOTE terhadap kelas sentimen negatif diterapkan. Kelas netral hasil pelabelan otomatis akan dihilangkan dalam penelitian ini karena Algoritma Support Vector Machine bersifat linear. Pengolahan dataset menggunakan aplikasi analisis data mining bernama RapidMiner. Berikut ini adalah alur penelitian klasifikasi sentimen yang dilakukan, dimulai dari pengambilan data, preprocessing data, visualisasi, pemodelan dan pengujian serta evaluasi model machine learning:



Gambar 1. Flowchart Penelitian

2.1 Data Crawling

Data Crawling adalah proses pengambilan data berbasis Uniform Resource Locator (URL) di situs web berdasarkan hasil pencarian dan pengumpulan data yang relevan dengan kata kunci. Token API pada akun Twitter digunakan dalam mengakses data opini masyarakat dari aplikasi tersebut. Tanggal pengambilan pada tanggal 4 sampai 8 bulan Februari dan Maret. Kata kunci yang dimasukkan pada pencarian yaitu "pemilu 2024".

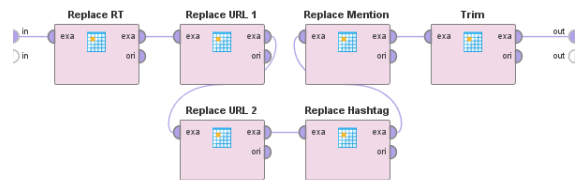


Gambar 2. Proses Crawling

2.2 Data Preprocessing

Tahapan ini merupakan tahapan reduksi data untuk meringankan beban komputasi pada saat pemodelan diterapkan. Berikut ini adalah bagian-bagian preprocessing yang termasuk di dalamnya:

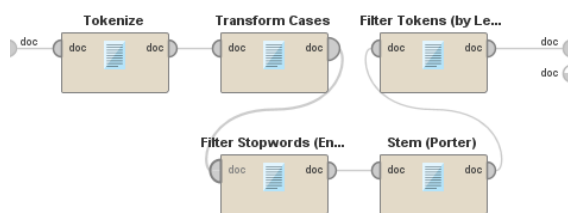
- Data filtering, meliputi penghapusan Uniform Resource Locator (URL), kata RT (retweet), simbol, tanda baca, angka, whitespace, dan duplikasi data.



Gambar 3. Proses Data Filtering

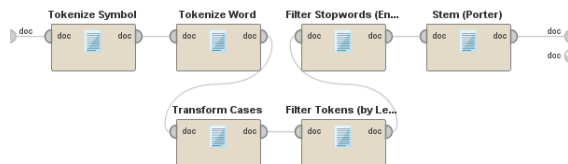
- Case folding, merupakan tahapan untuk mengubah setiap karakter berhuruf kapital menjadi huruf kecil tanpa terkecuali. Proses ini menghindari terjadinya variasi instances yang memiliki urutan huruf dan kesamaan makna.
- Translation, penerjemahan teks ke dalam Bahasa Inggris diimplementasikan agar pada saat pemberian kelas menggunakan VADER dapat bekerja secara efektif menggunakan kamus bahasa Inggris.
- Labeling, pemberian label diberikan berdasarkan kamus yang dimiliki oleh metode VADER. Hasilnya akan berupa angka desimal yang nantinya dikonversikan ke dalam kelas "Positif" jika skor lebih dari nol dan "Negatif" jika skor kurang dari nol. Sedangkan, skor yang memiliki nilai sama dengan nol akan dikonversikan ke nilai "Netral" dan dihapus.

- e. *Tokenization*, teknik yang digunakan untuk memecah suatu kalimat menjadi kata-kata penyusunnya. Masing-masing kata ini akan menjadi *instance* dan dihitung berdasarkan tingkat kemunculannya untuk kebutuhan proses *TF-IDF* maupun *WordCloud*.
- f. *Stopword*, penghapusan kata-kata umum yang sering digunakan, dimana dianggap tidak bermakna dalam proses analisis *Text Mining*. Jika *stopword* ini tetap ada, maka akan terjadi pergeseran jumlah *string* beserta frekuensinya pada kedua kelas.
- g. *Stemming*, berperan sebagai metode untuk mereduksi kata ke bentuk dasarnya (Kaburuan & Setiawan, 2023). Setiap kata yang memiliki imbuhan akan dihilangkan.



Gambar 4. Preprocessing WordCloud

Proses yang dilalui yaitu *Tokenize*, *Transform Cases*, *Stopwords (English Dictionary)*, *Stem (Porter)*, dan *Filter Tokens by Length*.



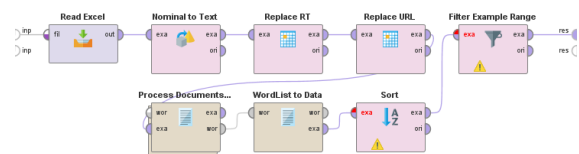
Gambar 5. Preprocessing SVM & SVM (PSO)

Kedua model pada gambar 5 memerlukan tahapan preprocessing seperti *Tokenize Symbol & Word*, *Transform Cases*, *Filter Tokens by Length*, *Filter Stopwords (English)*, dan *Stem (Porter)*.

2.3 Visualisasi

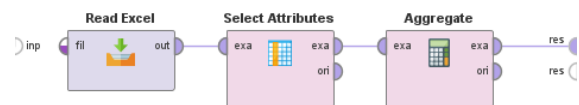
Agar mempermudah analisis hasil pengolahan data, maka penggambarannya akan dibagi menjadi dua tahapan seperti yang dijelaskan berikut ini:

- a. *WordCloud*, 20 *token* yang dominan dituangkan ke dalam suatu gambar untuk merepresentasikan kata-kata yang sering muncul pada sentimen positif dan negatif.



Gambar 6. Proses WordCloud

- b. Sentimen, visualisasi hasil pelabelan *VADER* berupa diagram untuk melihat besaran persentase sentimen positif yang menandakan dukungan masyarakat terhadap adanya pemilu 2024 maupun sentimen negatif sebagai masyarakat yang kurang menyetujui program tersebut.



Gambar 7. Proses Sentimen

2.4 Weighting

Weighting atau pembobotan merupakan teknik untuk mengubah token menjadi nilai berupa numerik. Data numerik ini berfungsi untuk pembentukan vektor matriks pada saat proses klasifikasi *machine learning*. Dalam penelitian ini, metode yang digunakan sebagai pembobotan adalah *TF-IDF*.

2.5 Data Balancing

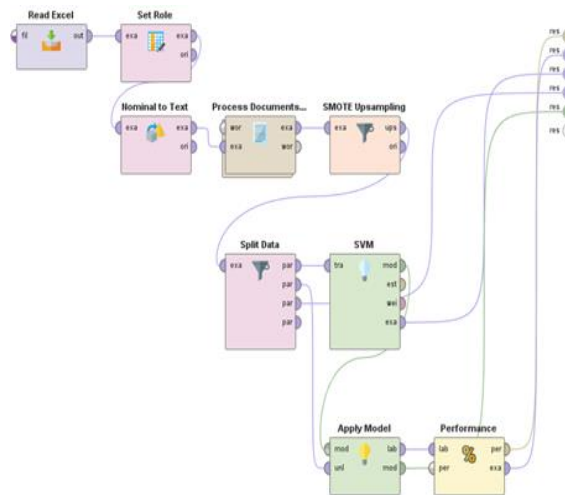
Data balancing atau penyeimbangan data dilakukan dengan tujuan membantu peningkatan akurasi yang akan diperoleh oleh hasil klasifikasi *machine learning*. Salah satu metode yang dapat dipakai adalah *SMOTE*. Dengan adanya *SMOTE*, kelas yang memiliki *instances* sedikit dibandingkan kelas lainnya akan dinaikkan menggunakan data sintesis sehingga kelas dengan jumlah *instances* terbesar tercapai. Dataset yang sebelumnya berjumlah 1400, dinaikkan menjadi 2000 data yang terdiri dari 1000 data sentimen positif dan 1000 data sentimen negatif.

2.6 Klasifikasi Machine Learning

Tahapan ini melibatkan *SVM* yang bekerja sendiri maupun dengan optimasi *PSO* untuk melihat perbedaan akurasi diantara keduanya Berikut adalah kedua proses tersebut:

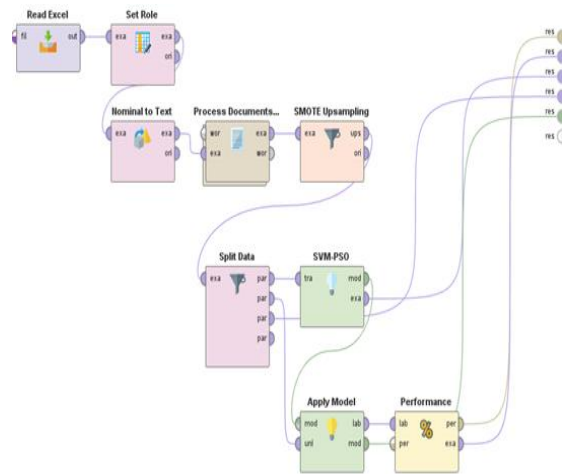
- a. *SVM*, model *Support Vector Machine (SVM)* merupakan algoritma *machine learning* berbasis *supervised learning*, dimana memerlukan data latih yang sudah diberikan label atau kelas. *SVM* menggunakan *kernel* pada ruang pencarian berdimensi tinggi. Dua (2) kelompok kelas dipisahkan berdasarkan pola tertentu sehingga *hyperplane* yang dipengaruhi *margin* antar kelas tersebut

dapat tercipta dengan maksimal (Risnanto et al., 2020).



Gambar 8. Proses Klasifikasi SVM

b. SVM (PSO), SVM kombinasi optimasi, dimana Particle Swarm Optimization (PSO) merupakan algoritma fitur seleksi sederhana untuk menemukan nilai atau pembobotan optimasi maupun pemilihan atribut diantara banyak atribut lainnya. Fitur seleksi berarti mencari kaitan antara satu atribut dengan yang lain dalam proses pembobotan (Arsi et al., 2021).



Gambar 9. Proses Klasifikasi SVM (PSO)

2.7 Evaluasi Machine Learning

Evaluasi algoritma SVM maupun dengan kombinasi PSO didasarkan pada pemodelan dengan total 2000 data berisi 1000 sentimen positif dan 1000 sentimen negatif. 600 data merupakan data sintesis dari metode SMOTE. Pengujian yang dilakukan yaitu 30% (600 data) dengan data latih sebesar 70% (1400 data) yang diambil secara acak.

3. HASIL DAN PEMBAHASAN

3.1 Preprocessing

a. Data filtering, keempat teks pada tabel 1 adalah sekumpulan opini masyarakat yang diunggah ke Twitter.

Tabel 1. Data Filtering

No.	Teks	Data Filtering
1	Suksesan Pemilu 2024, Jangan ada Politik Identitas dan SARA https://t.co/iHil9QpKK7	Suksesan Pemilu Jangan ada Politik Identitas dan SARA
2	RT @gittryani: Mari bersama kita cegah hoaks menjelang Pemilu 2024. #Pemilu2024 #PemiluSerentak2024 https://t.co/ukxZ00q...	Mari bersama kita cegah hoaks menjelang Pemilu
3	PN Jakarta Perintahkan 2024 Ditunda, Mahfud MD: Sensasi Berlebihan #Sindonews #BukanBeritaBiasa https://t.co/tnSMkRb8ot	PN Jakarta Pusat Perintahkan Pemilu Ditunda Mahfud MD Sensasi Berlebihan
4	RT @gustjahSaja: Putusan @democrazymedia Putusan PN Jakpus adalah bagian dari Drama Penundaan Pemilu 2024, banyak peran disana, ada peran Antagonis...	Putusan PN Jakpus adalah bagian dari Drama Penundaan Pemilu banyak peran disana ada peran Antagonis

Kalimat-kalimat yang memiliki URL, mention, tagar, simbol beserta angka telah dihapus pada kolom Data Filtering.

b. Case folding, seluruh teks yang memiliki awalan huruf, singkatan, dan ejaan-ejaan lain yang memakai huruf kapital akan dikonversikan ke dalam teks berisi huruf kecil.

Tabel 2. Case Folding

No.	Data Filtering	Case Folding
1	Suksesan Pemilu Jangan ada Politik Identitas dan SARA	suksesan pemilu jangan ada politik identitas dan sara

No.	Data Filtering	Case Folding
2	Mari bersama kita cegah tersebarnya berita hoaks menjelang Pemilu	mari bersama kita cegah tersebarnya berita hoaks menjelang pemilu
3	PN Jakarta Pusat Perintahkan Pemilu Ditunda Mahfud MD Sensasi Berlebihan	pn jakarta pusat perintahkan pemilu ditunda mahfud md sensasi berlebihan
4	Putusan PN Jakpus adalah bagian dari Drama Penundaan Pemilu banyak peran disana ada peran Antagonis	putusan pn jakpus adalah bagian dari drama penundaan pemilu banyak peran disana ada peran antagonis

Pada kolom *Case Folding*, setiap huruf kapital sudah diganti dengan huruf kecil tanpa terkecuali.

- c. *Translation*, dataset diubah menjadi teks berbahasa inggris. Formula *translate* pada *Google Spreadsheet* digunakan karena cara penerapannya sederhana.

Tabel 3. Translation

No.	Case Folding	Translation
1	suksesan pemilu jangan ada politik identitas dan sara	<i>successful elections there is no identity and sara politics</i>
2	mari bersama kita cegah tersebarnya berita hoaks menjelang pemilu	<i>lets together with us prevent the spread of hoaks news before the election</i>
3	pn jakarta pusat perintahkan pemilu ditunda mahfud md sensasi berlebihan	<i>central jakarta district court ordered the election to be postponed mahfud md excessive sensation</i>
4	putusan pn jakpus adalah bagian dari drama penundaan pemilu banyak peran disana ada peran antagonis	<i>the decision of the central jakarta district court is part of the drama delayed elections many roles there is an antagonistic role</i>

Teks sudah diterjemahkan menjadi kalimat berbahasa inggris pada kolom *Translation* tetapi ada satu kata yang tidak berubah

seperti kata hoaks, dimana seharusnya menjadi *hoax*.

- d. *Labeling*, metode *VADER* yang memiliki leksikon tersendiri terhadap pemberian skor kata digunakan untuk menghitung nilai polaritas (*compound*).

Tabel 4. VADER Scoring

No	Translation	Scoring String
1	<i>successful elections there is no identity and sara politics</i>	<i>no successful (0.72)</i>
2	<i>lets together with us prevent the spread of hoaks news before the election</i>	<i>prevent (0.03)</i>
3	<i>central jakarta district court ordered the election to be postponed mahfud md excessive sensation</i>	<i>postponed (-0.21)</i>
4	<i>the decision of the central jakarta district court is part of the drama delayed elections many roles there is an antagonistic role</i>	<i>delayed (-0.23), antagonistic (-0.44)</i>

Hasil perhitungan scoring string dari keempat kalimat pada tabel 4 secara berturut-turut adalah 0.41, 0.03, -0.21, dan -0.67. Maka pemberian kelas positif dan negatif dari penggunaan metode *VADER* dapat dilihat pada tabel 5 berikut:

Tabel 5. VADER Labeling

No.	Translation	Kelas
1	<i>successful elections there is no identity and sara politics</i>	Positif
2	<i>lets together with us prevent the spread of hoaks news before the election</i>	Positif
3	<i>central jakarta district court ordered the election to be postponed mahfud md excessive sensation</i>	Negatif
4	<i>the decision of the central jakarta district court is part of the drama delayed elections</i>	Negatif

No.	Translation	Kelas
	many roles there is an antagonistic role	

Maka didapat data nomor satu dan dua dikategorikan sebagai sentimen positif, yaitu data 1 dan 2. Sedangkan, data 3 dan 4 termasuk sebagai kategori sentimen negatif.

- e. *Tokenization*, setiap kalimat diubah menjadi sekumpulan kata pembentuknya. Kata-kata ini terpisah menjadi atribut-atribut tersendiri.

Tabel 6. *Tokenization*

No.	Translation	Tokenization
1	successful elections there is no identity and sara politics	successful, elections, there, is, no, identity, and, sara, politics
2	lets together with us prevent the spread of hoaks news before the election	lets, together, with, us, prevent, the, spread, of, hoaks, news, before, the, election
3	central jakarta district court ordered the election to be postponed mahfud md excessive sensation	central, jakarta, district, court, ordered, the, election, to, be, postponed, mahfud, md, excessive, sensation
4	the decision of the central jakarta district court is part of the drama delayed elections many roles there is an antagonistic role	the, decision, of, the, central, jakarta, district, court, is, part, of, the, drama, delayed, elections, many, roles, there, is, an, antagonistic, role

Pemisah kata (atribut) pada kolom *Tokenization* adalah tanda koma (.). *Tokenization* diterapkan beberapa kali, yaitu saat proses labelisasi, klasifikasi *Support Vector Machine* optimasi dan *non-optimasi*, serta *WordCloud*.

- f. *Stopword*, Proses *stopword* melibatkan pencarian dan penghapusan kata-kata seperti konjungsi yang umum digunakan.

Tabel 7. *Stopword*

No.	Tokenization	Stopword
1	successful, elections, there, is, no, identity, and, sara, politics	there, is, no, and

No.	Tokenization	Stopword
2	lets, together, with, us, prevent, the, spread, of, hoaks, news, before, the, election	together, with, us, the, of, before
3	central, jakarta, district, court, ordered, the, election, to, be, postponed, mahfud, md, excessive, sensation	the, to, be
4	the, decision, of, the, central, jakarta, district, court, is, part, of, the, drama, delayed, elections, many, roles, there, is, an, antagonistic, role	the, of, is, many, there, an

Tabel 7 dengan kolom *Stopword* adalah kata-kata yang akan dihapus karena kurangnya manfaat terhadap ekstraksi informasi.

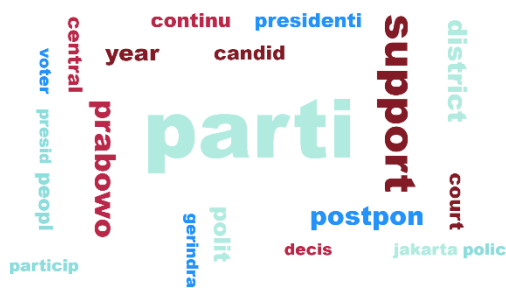
- g. *Stemming*, bertujuan untuk menghilangkan kata-kata yang mempunyai imbuhan. Contohnya dalam bahasa inggris adalah *re-*, *over-*, dan *non-* sebagai awalan kata serta *-ness*, *-less*, dan *-ment* pada akhiran kata.

Tabel 8. *Stemming*

No.	Tokenization	Stemming
1	successful, elections, identity, sara, politics	success, elect, ident, sara, polit sara, politics
2	lets, prevent, spread, hoaks, news, election	let, prevent, spread, hoak, new, elect
3	central, jakarta, district, court, ordered, election, postponed, mahfud, md, excessive, sensation	central, jakarta, district, court, order, elect, postpon, mahfud, md, excess, sensat sensation
4	decision, central, jakarta, district, court, part, drama, delayed, elections, roles, antagonistic, role	decis, central, jakarta, district, court, part, drama, delay, elect, role, antagonist, role

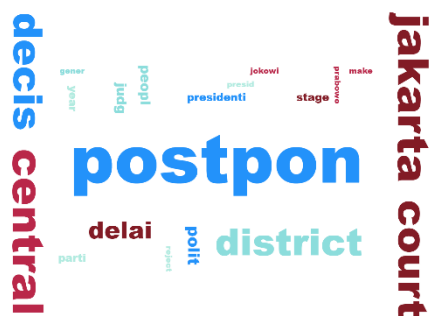
3.2 Visualisasi

- a. *WordCloud*



Gambar 10. WordCloud Sentimen Positif

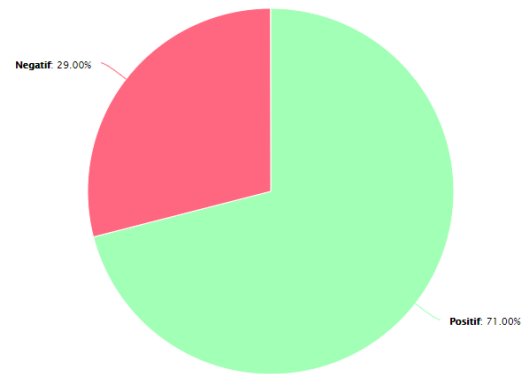
Sebanyak 20 kata dengan frekuensi tertinggi pada data sentimen positif diambil sebagai bentuk *WordCloud*. Terdapat lima perolehan kata tertinggi yaitu *partai* dengan total 278 kemunculan, *support* dengan total 149 kemunculan, *prabowo* dengan total 111 kemunculan, 94 total kemunculan dari kata *postpon*, dan 92 total kemunculan dari kata *district*.



Gambar 11. WordCloud Sentimen Negatif

Sama seperti *WordCloud* sebelumnya, 20 kata tertinggi diambil untuk divisualisasikan ke dalam bentuk *WordCloud* sentimen negatif. Kata-kata yang berada di posisi 5 teratas adalah *postpon* dengan total 156 kemunculan, *jakarta* 104 kemunculan, *central* 100 kemunculan, 98 kemunculan pada kata *court*, dan 96 kemunculan pada kata *decis*.

- b. Sentimen
Kelas dengan persentase tertinggi hasil pelabelan metode *VADER* adalah sentimen positif sebesar 71.43% (1000 data). Kemudian diikuti persentase 28.57% (400 data) pada sentimen negatif. Berikut adalah diagram visualisasi sentimen:



Gambar 12. Visualisasi Sentimen

3.3 Evaluasi Klasifikasi *Machine Learning*

Tabel 9. *Confusion Matrix SVM 30%*

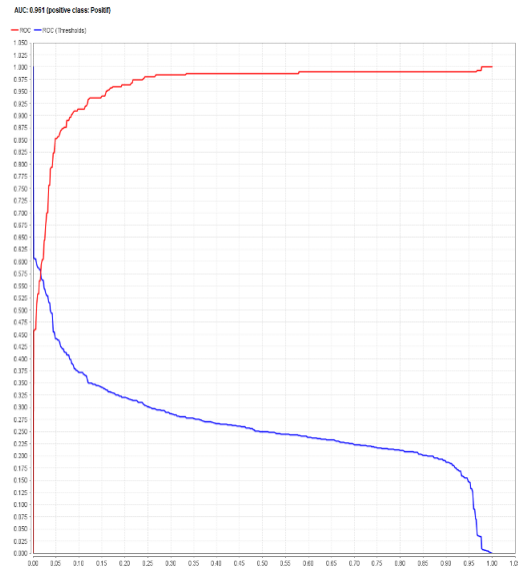
Aktual Sentimen	Prediksi Sentimen	
	Positif	Negatif
Positif (TP, FP)	235	11
Negatif (FN, TN)	65	289

Percobaan yang dilakukan terhadap 600 data pengujian ini memperoleh data prediksi benar dengan sentimen positif sebanyak 235 data dan data prediksi benar untuk data negatif adalah 289 data. Lalu, data misklasifikasi pada sentimen positif adalah sebanyak 11 data dan 65 data lain merupakan kesalahan klasifikasi dari sentimen negatif.

Tabel 10. *Confusion Matrix SVM (PSO) 30%*

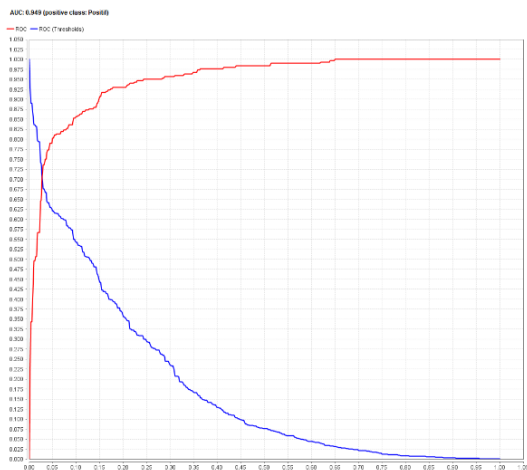
Aktual Sentimen	Prediksi Sentimen	
	Positif	Negatif
Positif (TP, FP)	263	38
Negatif (FN, TN)	37	262

Dilihat dari tabel 10 dengan 600 data uji, perolehan ketepatan klasifikasi sentimen positif maupun sentimen negatif secara berturut-turut adalah 263 dan 262 data. Sedangkan, 38 beserta 37 data lainnya adalah hasil kesalahan klasifikasi SVM pada sentimen positif dan sentimen negatif.



Gambar 13. Kurva ROC SVM 30%

Hasil evaluasi dari *AUC* (Area Under the Curve) pengujian SVM 600 data adalah sebesar 0.961.



Gambar 14. Kurva ROC SVM (PSO) 30%

Untuk evaluasi *AUC* dari pengujian SVM (PSO) adalah 0.949. Nilai dari *AUC* ini bisa dipakai untuk mengetahui apakah suatu model klasifikasi berjalan dengan baik. Berikut ini adalah aturan yang menentukan nilai tersebut berdasarkan penelitian (Que et al., 2020):

- Nilai *AUC* diantara 0.90 - 1.00 menandakan bahwa hasil klasifikasi sangat baik;
- Nilai *AUC* diantara 0.80 - 0.90 menandakan bahwa hasil klasifikasi baik;
- Nilai *AUC* diantara 0.70 - 0.80 membuktikan bahwa hasil klasifikasi cukup baik;

- Nilai *AUC* diantara 0.60 - 0.70 menunjukkan bahwa hasil klasifikasi buruk;
- Nilai *AUC* diantara 0.50 - 0.60 menunjukkan bahwa hasil klasifikasi sangat buruk.

Dengan perolehan model *AUC* 0.961 dan 0.949 pada kedua uji coba, maka hasil klasifikasi keduanya dapat dikatakan layak. Berikut ini adalah komparasi perolehan *accuracy*, *precision*, *recall*, dan *f-measure* antara algoritma SVM dan algoritma SVM (PSO):

Tabel 11. Komparasi SVM dan SVM (PSO) 30%

Parameter	SVM	SVM (PSO)
Accuracy	87.33%	87.50%
Precision	95.53%	87.38%
Recall	78.33%	87.67%
F-Measure	86.08%	87.52%
AUC	0.961	0.949

Dari kelima parameter, algoritma SVM (PSO) unggul dalam perolehan persentase *accuracy*, *precision*, dan *recall*. Persentase algoritma SVM lebih rendah pada *precision* maupun *AUC*. Selain itu, terdapat selisih *recall* yang tinggi dari kedua algoritma, yaitu sebesar 9.34%.

4. KESIMPULAN

Hasil klasifikasi sentimen positif dengan pelabelan otomatis *VADER* sebanyak 1000 data, jumlah ini lebih besar daripada klasifikasi sentimen negatif yang hanya berjumlah 400 data. Kemudian algoritma SVM mendapatkan hasil persentase akurasi data *split non-PSO* 70:30 adalah 87.33%. Sedangkan pada data *split* kombinasi PSO sebesar 87.50%. Maka hasil terbaik didapatkan oleh data *split* 70:30 kombinasi PSO dengan persentase 87,50%. Pemodelan SVM (PSO) mampu menyeleksi fitur lebih baik dibandingkan algoritma SVM tersendiri yang ditandai dari perolehan persentase akurasi. Di sisi lain, algoritma *machine learning* ini memerlukan sejumlah data latih pelabelan yang bisa berupa otomatis maupun secara manual.

Untuk *WordCloud* pada kedua data sentimen, ekstraksi informasi yang dapat digarisbawahi adalah terkait *postpon* (penundaan), Jakarta, *court* (pengadilan), *reject* (menolak), beberapa Figur Publik dan *parti* (Partai), serta *support* (dukungan). Dataset penelitian berkaitan erat dengan adanya isu penundaan pemilu. Dari jumlah sentimen positif yang mencapai 1000, maka bisa disimpulkan bahwa masyarakat mendukung keberadaan

pemilu yang kemudian ditandai juga oleh banyaknya kata “*reject*” atau penolakan penundaan pemilu pada sentimen negatif maupun kata “*support*” atau dukungan pada sentimen positif. Adapun informasi bahwa beberapa orang kurang menyukai figur publik atau partai tertentu.

Sebagai bentuk penelitian selanjutnya, *tuning parameter SVM* dan *PSO* seperti *C*, *gamma*, *local best*, dan *global best* bisa diatur sedemikian rupa tanpa menggunakan *SMOTE* dengan pengujian *Cross Validation* agar akurasi klasifikasi *machine learning* meningkat. Algoritma lain seperti *K-Nearest Neighbor*, *Long Short Term Memory* dan *Logistic Regression* juga dapat diterapkan.

UCAPAN TERIMA KASIH

Terima kasih kepada dosen pembimbing dan teman-teman mahasiswa di Universitas Mercu Buana yang telah memberikan ide, saran, dan dukungan dalam menyelesaikan penelitian ini.

DAFTAR PUSTAKA

- Arsi, P., Wahyudi, R., & Waluyo, R. (2021). Optimasi SVM Berbasis PSO pada Analisis Sentimen Wacana Pindah Ibu Kota Indonesia. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(2), 231–237. <https://doi.org/https://doi.org/10.29207/resti.v5i2.2698>
- Chauhan, P., Sharma, N., & Sikka, G. (2021). The emergence of social media data and sentiment analysis in election prediction. *Journal of Ambient Intelligence and Humanized Computing*, 12(2), 2601–2627. <https://doi.org/10.1007/s12652-020-02423-y>
- Chiny, M., Chihab, M., Bencharef, O., & Chihab, Y. (2021). LSTM, VADER and TF-IDF based Hybrid Sentiment Analysis Model. *International Journal of Advanced Computer Science and Applications*, 12(7), 265–275. <https://doi.org/10.14569/IJACSA.2021.0120730>
- Diniyanto, A. (2022). Penundaan Pemilihan Umum di Negara Hukum: Kajian Demokrasi Konstitusional (General Elections Postponement in Rule of Law: A Study of Constitutional Democracy). *Negara Hukum: Membangun Hukum Untuk Keadilan Dan Kesejahteraan*, 13(2), 227–245. <https://doi.org/https://doi.org/10.22212/jnh.v13i2.3365>
- Firdlous, D. A., Andrian, R., & Widodo, S. (2023). Sentiment Analysis Public Twitter on 2024 Election using the Long Short Term Memory Model. *SISTEMASI*, 12(1), 52–60. <https://doi.org/10.32520/stmsi.v12i1.2145>
- Hafidz, N., & Liliana, D. Y. (2021). Klasifikasi Sentimen pada Twitter Terhadap WHO Terkait Covid-19 Menggunakan SVM, N-Gram, PSO. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(2), 213–219. <https://doi.org/10.29207/resti.v5i2.2960>
- Juanita, S. (2020). Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 4(3), 552–558. <https://doi.org/https://doi.org/10.30865/mib.v4i3.2140>
- Kaburuan, E. R., & Setiawan, N. R. (2023). Sentimen Analisis Review Aplikasi Digital Korlantas Pada Google Play Store Menggunakan Metode SVM. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 12(1), 105–116. <https://doi.org/https://doi.org/10.32736/sisfokom.v12i1.1614>
- Margiansyah, D. (2019). Populisme di Indonesia Kontemporer: Transformasi Persaingan Populisme dan Konsekuensinya dalam Dinamika Kontestasi Politik Menjelang Pemilu 2019. *Jurnal Penelitian Politik*, 16(1), 47–68. <https://doi.org/https://doi.org/10.14203/jpp.v16i1.783>
- Nardilasari, A. P., Hananto, A. L., Hilabi, S. S., Tukino, & Priyatna, B. (2023). Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM Pada Media Sosial Twitter. *JOINTECS (Journal of Information Technology and Computer Science)*, 8(1), 11–18. <https://doi.org/https://doi.org/10.31328/jointecs.v8i1.4265>
- Obiedat, R., Qaddoura, R., Al-Zoubi, A. M., Al-Qaisi, L., Harfoushi, O., Alrefai, M., & Faris, H. (2022). Sentiment Analysis of Customers’ Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution. *IEEE Access*, 10, 22260–22273. <https://doi.org/https://doi.org/10.1109/ACCESS.2022.3149482>
- Que, V. K. S., Iriani, A., & Purnomo, H. D. (2020). Analisis Sentimen Transportasi Online

- Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 9(2), 162–170. <https://doi.org/https://doi.org/10.22146/jnteti.v9i2.102>
- Risnantoyo, R., Nugroho, A., & Mandara, K. (2020). Sentiment Analysis on Corona Virus Pandemic Using Machine Learning Algorithm. *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, 4(1), 86–96. <https://doi.org/https://doi.org/10.31289/jite.v4i1.3798>
- Rohim, M., & Wardana, A. (2019). Analisis Politik Milenial : Persepsi Siswa SMA Terhadap Dinamika Politik Pada PEMILU 2019 di Indonesia. *Jurnal Ilmu Pemerintahan : Kajian Ilmu Pemerintahan Dan Politik Daerah*, 4(1), 47–63. <https://doi.org/https://doi.org/10.24905/jip.v4i1.1212>
- Sautomo, S., Hafidz, N., Achyani, Y. E., & Gata, W. (2020). Sentiment Analysis Due to " Mudik" Prohibited of COVID-19 Through Twitter. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 6(1), 7–12. <https://doi.org/https://doi.org/10.33480/jitk.v6i1.1357>
- V. D, C. (2019). Hybrid approach: naive bayes and sentiment VADER for analyzing sentiment of mobile unboxing video comments. *International Journal of Electrical and Computer Engineering (IJECE)*, 9(5), 4452. <https://doi.org/10.11591/ijece.v9i5.pp4452-4459>
- Widi, S. (2022, December 15). *Data Pemilih Golput saat Pemilu, Turun Drastis pada 2019*. Dataindonesia.Id.
- Yusri, M., Amrizal, D., & Yusriati. (2019). *Pendidikan Politik Pemerintah Kontra Golput* (R. Franita, Ed.; 1st ed.). Lembaga Penelitian dan Penulisan Ilmiah AQLI. https://books.google.co.id/books?hl=id&lr=&id=_PbyDwAAQBAJ&oi=fnd&pg=PA1&dq=Golput&ots=0pHEAF1tln&sig=w13KkEQnJlGPgBORHscg4dinNzl&redir_esc=y#v=onepage&q=Golput&f=false