

Implementasi Metode SVM Pada Sentimen Analisis Terhadap Pemilihan Presiden (Pilpres) 2024 Di Twitter

Jenny Anggraini , Debby Alita

Sistem Informasi, Universitas Teknokrat Indonesia, Labuhan ratu, Bandar Lampung, 35141, Indonesia

Info Artikel

Riwayat Artikel:

Received: 12-02-2024

Final Revision: 07-08-2024

Accepted: 07-08-2024

Abstract – The focus of the research is the use of Twitter as a platform to express the political opinions of the Indonesian people regarding the 2024 Presidential Election. By utilizing sentiment analysis using the Support Vector Machine (SVM) method, this research aims to evaluate the accuracy of SVM in classifying tweets and compare the performance of four types of SVM kernels. Visualizations of positive and negative sentiments are also generated to provide a clearer picture. The stages of the research involve Twitter data collection, and pre-processing with steps such as data cleansing, case folding, tokenizing, stemming, and filtering. Labeling is done to identify sentiment, then feature extraction using TF-IDF. SVM implementation with linear, polynomial, RBF, and sigmoid kernels is performed, followed by model evaluation using precision, recall, F-measure, and accuracy metrics. The study used SVM to analyze the sentiment of the 2024 presidential election on Twitter data. As a result, out of 3938 tweets, 1575 were positive and 2363 were negative. The SVM model achieved 95.05% accuracy, superior in predicting negative sentiment. Comparison of SVM kernels shows the highest accuracy in the linear kernel 95.43%. Sentiment analysis on tweets shows a majority of positive support for Ganjar 54.9%, while Anies and Prabowo have support levels of 15.8% and 29.3% respectively.

Keywords: 2024 presidential election, Analysis sentiment, Kernel SVM, Support Vector Machine (SVM), Twitter.

Corresponding Author:

Debby Alita

Sistem Informasi, Universitas Teknokrat Indonesia, Labuhan ratu, Bandar Lampung, 35141, Indonesia.

Email: debbyalita@teknokrat.ac.id



This is an open access article under the [CC BY 4.0](#) license.

Abstrak – Penggunaan Twitter sebagai platform untuk menyampaikan pendapat politik masyarakat Indonesia terkait Pilpres 2024 menjadi fokus penelitian. Dengan memanfaatkan analisis sentimen menggunakan metode Support Vector Machine (SVM), penelitian ini bertujuan mengevaluasi akurasi SVM dalam mengklasifikasikan tweet dan membandingkan performa empat jenis kernel SVM. Visualisasi sentimen positif dan negatif juga dihasilkan untuk memberikan gambaran lebih jelas. Tahapan penelitian melibatkan pengumpulan data Twitter, pre-processing dengan langkah-langkah seperti data cleansing, case folding, tokenizing, stemming, dan filtering. Labelling dilakukan untuk mengidentifikasi sentimen, kemudian ekstraksi fitur menggunakan TF-IDF. Implementasi SVM dengan kernel linier, polynomial, RBF, dan sigmoid dilakukan, diikuti oleh evaluasi model menggunakan metrik precision, recall, F-measure, dan akurasi. Penelitian menggunakan SVM untuk menganalisis sentimen Pilpres 2024 pada data Twitter. Hasilnya, dari 3938 tweet, 1575 positif dan 2363 negatif. Model SVM mencapai akurasi 95.05%, lebih unggul dalam memprediksi sentimen negatif. Perbandingan kernel SVM menunjukkan akurasi tertinggi pada kernel linier 95.43%. Analisis sentimen pada tweet menunjukkan mayoritas dukungan positif untuk Ganjar 54.9%, sementara Anies dan Prabowo memiliki tingkat dukungan masing-masing 15.8% dan 29.3%.

Kata Kunci: Analisis sentimen, Kernel SVM, Pilpres 2024, Support Vector Machine (SVM), Twitter.

I. PENDAHULUAN

Teknologi informasi dan komunikasi berkembang dengan cepat akhir-akhir ini. Informasi bisa didapat dengan cepat dan mudah terhadap berbagai sumber daya informasi dan komunikasi secara *online*, dengan adanya media sosial. Perkembangan ini tidak hanya mempengaruhi cara kita memperoleh informasi, tetapi juga memengaruhi cara kita berkomunikasi dan menyuarakan pendapat, terutama melalui media sosial seperti Twitter [1].

Twitter adalah platform jejaring sosial yang dapat digunakan untuk menyampaikan pendapat, saran atau kritik, di mana orang dapat berbagi pemikiran mereka tentang berbagai subjek, baik pribadi maupun publik. Pengguna Twitter mempunyai kemampuan untuk mengirimkan opini mereka, dan pengguna lain dapat melihatnya. Postingan yang dibuat oleh pengguna Twitter biasanya disebut dengan tweet [2]. Hasil survei dari www.digital-2023-indonesia jumlah penduduk Indonesia yang menggunakan jumlah pengguna media sosial yang aktif mencapai 167 juta, setara dengan 60,4% dari jumlah penduduk keseluruhan. Indonesia salah satunya adalah Twitter [3]. Masyarakat dapat memanfaatkan basis pengguna Twitter yang sangat besar untuk mengetahui pendapat masyarakat umum tentang Pemilihan presiden Indonesia [4].

Pemilihan Presiden yang akan diadakan pada tahun 2024 sudah mulai dirasakan kehadirannya, terutama di media sosial yang menjadi wadah untuk mengungkapkan pandangan, sentimen, dan preferensi terhadap tokoh politik yang seringkali mendapatkan penilaian positif dari lembaga survei[5]. Topik yang sedang dibahas saat ini adalah Pilpres 2024, yang mendapat sorotan dari opini publik di Twitter dengan jumlah tweet yang melimpah. Hal

ini menciptakan beragam pandangan, baik yang positif maupun negatif. Kumpulan data tweet tersebut dapat dijadikan sebagai sumber data yang dapat diolah atau dianalisis salah satunya melalui analisis sentimen.

Analisis sentimen dalam penelitian penggalian text yang digunakan untuk mengkategorikan kepercayaan, perasaan, dan sikap orang terhadap berbagai entitas. Dalam konteks penelitian ini, analisis sentimen digunakan untuk mengidentifikasi apakah tweet memiliki sentimen positif atau negatif terkait dengan calon presiden dalam Pemilihan Presiden 2024. serta Identifikasi sentimen ada dalam sebuah opini dengan analisis sentimen ini menganalisis sekumpulan data opini dengan menginterpretasikan dan menganalisis data teksual secara otomatis. Salah satu komponen dari penelitian penggalian teks, yang juga dikenal sebagai penelitian komputasi, adalah analisis sentimen, yang digunakan untuk mengkategorikan kepercayaan, perasaan, dan sikap orang terhadap berbagai entitas [6].

Melalui perbandingan tiga metode pembelajaran analisis sentimen yakni, Klasifikasi *Naïve Bayes* (NBC), *Maximum Entropy*, dan *Support Vector Machine* (SVM). [7]. SVM memiliki akurasi tertinggi sebesar 95,05% sebagai hasilnya, sehingga salah satu metode yang dapat digunakan untuk masalah regresi dan klasifikasi adalah *Support Vector Machine* (SVM). [8]. Dalam hal klasifikasi yang dapat dipisahkan secara linier, SVM dan pengklasifikasi linier memiliki ide dasar yang sama. Meskipun demikian, SVM telah ditingkatkan agar dapat berfungsi pada masalah *non-linear* dengan membawa konsep kernel ke dalam ruang kerja berdimensi tinggi.

Tujuan dari penelitian ini adalah mengevaluasi akurasi SVM dalam mengidentifikasi tweet dan mengamati hasil analisis sentimen, untuk mengukur sentimen publik secara akurat terkait calon presiden dan isu-isu terkait pilpres 2024 di platform twitter. Serta menentukan perbandingan kernel terbaik antara empat kernel SVM yaitu linear, polynomial, RBF dan sigmoid, adapun tahapan penelitian yaitu pengumpulan data, pre-processing, labelling, ekstraksi fitur, implementasi SVM dan evaluasi model. Adapun manfaat dari penelitian ini dapat memprediksi awal tentang hasil pemilihan calon presiden, Meskipun tidak bersifat final, informasi ini dapat menjadi tambahan bagi analis politik untuk memahami dinamika dan tren yang memengaruhi hasil pemilihan calon presiden.

Penelitian ini dilakukan oleh H. Suharyo, dkk [9] menggunakan teknik pembelajaran mesin, yaitu metode *Support Vector Machine* (SVM) untuk menganalisis opini (sentimen) berdasarkan polaritas dan subjektivitas. Menggunakan kernel linear dengan F-Measure of Subjectivity 73%, 77%, dan 75% serta Skor F-Measure untuk polaritas adalah 68%, dengan angka yang sama sebesar 68%, dan 63% untuk masing-masing figur Anies Baswedan, Joko Widodo, dan Prabowo Subianto menghasilkan hasil analisis sentimen yang lebih ideal.

Pada penelitian yang dilakukan oleh H. April, dkk [10] membandingkan tiga algoritma klasifikasi: support vector machine (SVM), K-Nearest Neighbor (K-NN), dan *Naïve Bayes* (NB) untuk meneliti opini publik mengenai tren calon presiden. Hasil pengujian menunjukkan bahwa algoritma SVM mengungguli K-NN dan *Naïve Bayes*, dengan tingkat akurasi sebesar 79,57%. Hal ini berarti metode SVM merupakan algoritma yang paling efisien dan efektif untuk mengidentifikasi komentar positif dan negatif terkait tren calon presiden 2024. Danar, dkk [11] melakukan Analisis sentimen terhadap calon presiden dan wakil presiden Indonesia pada pemilihan umum tahun 2019 dipelajari oleh penulis studi ini. Akurasi terbaik sebesar 92,5% diperoleh dengan menggunakan 800 dataset, di mana 80% di antaranya merupakan data training dan 20% data testing. Nilai presisi dan recall untuk setiap kelas berkisar antara 85,7% hingga 97,2% dan 78,2% hingga 93,5%. I.

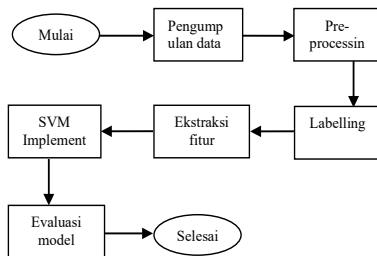
Bayu, dkk [12]. menentukan apakah judul berita mengandung sikap positif atau negatif, kami mengembangkan sistem analisis sentimen dalam penelitian ini dengan menggunakan data dari media berita online dan pendekatan *Support Vector Machine* (SVM). Hasil yang diperoleh adalah FastText mampu mengungguli TF-IDF untuk ekstraksi fitur. Serta kombinasi nilai Kernel, C, dan gamma yang memberikan skor akurasi terbaik masing-masing rbf, 1, dan auto pada rasio data uji 90:10, dengan skor akurasi 99%. Sedangkan F.H. Alberi, dkk [13] melakukan pendekatan *Support Vector Machine* dengan pemberian bobot TF-IDF untuk mengetahui mood politik pengguna Twitter terkait pemilihan presiden. Akurasi dan confusion matrix menjadi dasar untuk mengukur kinerja sistem. Nilai akurasi terbesar dari penelitian ini adalah 62.88% dengan menggunakan TF-IDF ketika kata Unigram, Bigram, dan Trigram digabungkan.

Pada penelitian B. Ghulam, dkk [14] bertujuan untuk menggunakan dataset Twitter untuk melakukan analisis sentimen terhadap calon presiden Indonesia tahun 2019. Penelitian ini menunjukkan bahwa nilai akurasi terbesar yaitu 79,02% diperoleh ketika metode machine learning SVM dan tokenisasi alfabetis digabungkan. Namun, kombinasi algoritma pembelajaran mesin NBC dengan tokenisasi N-gram menghasilkan nilai akurasi terendah dalam sampel ini, yaitu 44,94%.

II. METODE

Cara termudah untuk mengikuti aturan format halaman paper ini adalah menggunakan format dalam dokumen ini. Simpanlah file ini dengan nama lainnya, lalu ketikkan isi makalah anda ke dalamnya.

SVM adalah metode penelitian yang diterapkan dalam penelitian ini. Tahapan-tahapan penelitian berikut ini ditampilkan pada Gambar 1.



Gambar.1 Flowchart tahapan penelitian

A. Pengumpulan data

Peneliti menggunakan perayapan mengakses data melalui *Application Programming Interface* (API) untuk mengambil data dari Twitter. Jumlah data yang digunakan pada penelitian ini adalah 3938 tweet dengan total positif 1575 tweet dan 2363 tweet negatif yang diambil dari tweet masyarakat kepada calon presiden 2024 dengan kata kunci “pilpres2024”, “debatcapres” dan “capres”. Berikut pada Tabel I, adalah contoh tweet yang diperoleh:

TABEL I
TWEET HASIL DARI CRAWLING

| Username | Full_text |
|------------------|---|
| @detikcom | KPU: “Debat Pilpres Ketiga Tahun 2024 Akan Berlangsung di Istora Senayan Oleh karena itu, debat keempat dan terakhir akan berlangsung pada tanggal 7 Januari 2023 di Istora, tepatnya di Teater Senayan”. Kata Komisioner KPU, August Mellaz.> > https://t.co/6m0O8fwT71 #detikPemilu #DebatCapres” |
| @VISHALSing75552 | Debat Capres kali ini Pak Ganjar awesome banget dari tadi jawabannya.. Lanjutgann pakk!!! ☺ Mahfud Hebat #L3bihbaik #MahfudLebihBaik3 #GanjarMahfud2024 Buruh Naik Kelas |
| @YaswanthKumarJ1 | antaranews: Ganjar Pranowo mengatakan dirinya senang dan sudah siap menjalani debat perdana calon presiden-calon wakil presiden yang akan dilaksanakan malam ini. 7iGSH2JtaE Mahfud Hebat #L3bihbaik #MahfudLebihBaik3 #GanjarMahfud2024 sikat KKN |

B. Pre-processing

Pre-processing data merupakan langkah dalam proses data mining yang mengubah data mentah menjadi data berkualitas tinggi atau input yang berguna sehingga langkah selanjutnya dapat dilakukan [15]. Pembersihan data, pelipatan huruf, *tokenizing*, *stemming*, dan penyaringan adalah prosedur pra-pemrosesan yang dilakukan. Python digunakan untuk langkah pra-pemrosesan, dan pustaka NLTK disediakan.

C. Data Cleansing

Data *cleansing* merupakan tahap awal dari *pre-processing*. Pemilihan data dari Twitter didasarkan pada keterkaitannya dengan topik yang telah ditentukan, yaitu persepsi publik terhadap pemilihan presiden 2024. Berikut adalah hasil data yang telah di *cleaning* pada Tabel II

TABEL II
TWEET PROSES CLEANSING

| Tweet Sebelum Cleansing | Tweet Setelah Cleansing |
|---|---|
| KPU: "Debat Pilpres Ketiga Tahun 2024 Akan Berlangsung di Istora Senayan Oleh karena itu, debat keempat dan terakhir akan berlangsung pada tanggal 7 Januari 2023 di Istora, tepatnya di Teater Senayan". Kata Komisioner KPU, August Mellaz.> > https://t.co/6m0O8fwT71 #detikPemilu #DebatCapres" | KPU Istora Senayan akan menjadi tuan rumah Debat Ketiga Pilpres 2024. Menurut Komisioner KPU August Mellaz, debat ketiga akan berlangsung pada 7 Januari 2023 di Istora Senayan. gt gt detikDebat PilpresCapres |

D. Case Folding

Merupakan proses menyamaratakan semua huruf sehingga tidak ada perbedaan antar huruf satu dengan huruf lainnya seperti merubahnya menjadi huruf yang kecil [16]. Prosedur ini dilakukan karena isi tweet tidak selalu menggunakan huruf secara konsisten. Proses *Case Folding* dapat dilihat pada Tabel III.

TABEL III
TWEET SETELAH PROSES CASE FOLDING

| Tweet Sebelum Case Folding | Tweet Setelah Case Folding |
|---|--|
| KPU Istora Senayan akan menjadi tuan rumah Debat Ketiga Pilpres 2024. Menurut Komisioner KPU August Mellaz, debat ketiga akan berlangsung pada 7 Januari 2023 di Istora Senayan. gt gt detikDebat PilpresCapres | [kpu,debat,ketiga,pilpres,2024,a kan,digelar,di,istora, senayan,jadi ,untuk,debat,ketiga,itu,akan,dila ksanakan,pada,7,januari,tahun,2 023,bertempat,di,istora,tempatn ya,di,istora,senayan,kata,komisi oner,kpu,august,mellaz,gt,gt, detikpemilu,debatcapres] |

E. Tokenizing

Pada tahap *tokenizing* ini proses memecah kalimat atau teks menjadi kata-kata individual. Dengan melakukan tokenizing lebih mudah menghitung frekuensi kemunculan setiap kata dalam teks tersebut [17]. yang disebut token dan mungkin menghapus beberapa karakter (seperti tanda baca) pada saat yang bersamaan. [18]. Seperti yang ditujuhan pada Tabel IV

TABEL IV
TWEET SETELAH PROSES TOKENIZING

| Tweet Sebelum Tokenizing | Tweet Setelah Tokenizing |
|--|--|
| [kpu,debat,ketiga,pilpres,2024,a kan,digelar,di,istora, senayan,jadi ,untuk,debat,ketiga,itu,akan,dila ksanakan,pada,7,januari,tahun,2 023,bertempat,di,istora,tempatn ya,di,istora,senayan,kata,komisi oner,kpu,august,mellaz,gt,gt, detikpemilu,debatcapres] | [KPU,Debat,Ketiga,Pilpres,2024, Akan,Digelar,di,Istora, Senayan,Jadi,untuk,debat,ketiga,it u,akan,dilaksanakan, pada,7,januari,tahun,2023,bertem pat,di,Istora,tempatnya, di,Istora,Senayan,kata,Komisione r,KPU,August,Mellaz, gt,gt,detikPemilu,DebatCapres] |

F. Stemming

Tindakan terakhir Untuk menemukan akar kata dalam sebuah kalimat, salah satu tekniknya disebut *stemming*. Proses ini melibatkan pemisahan setiap kata dari bentuk dasarnya dan awalan atau akhiran apa pun[19]. Seperti pada Tabel V

TABEL V
TWEET SETELAH PROSES STEMMING

| Tweet Sebelum Stemming | Tweet Setelah Stemming |
|---|---|
| [KPU,Debat,Ketiga,Pilpres,2024, Akan,Digelar,di,Istora,Senayan,Ja di,untuk,debat,ketiga,itu,akan,dila ksanakan,pada,7,januari,tahun,20 23,bertempat,di,Istora,tempatnya, di,Istora,Senayan,kata,Komisioner ,KPU,August,Mellaz,gt,gt,detikPe milu,DebatCapres] | [kpu,debat,ketiga,pilpr,2024,akan, digelar,di,istora,senayan,jadi,untuk ,debat,ketiga,itu,akan, dilaksanakan,pada,7,januari,tahun, 2023,bertempat,di,istora,tempatnya ,di,istora,senayan,kata,komision,kp u,august,mellaz,gt,gt,detikpemilu, debatcapr] |

G. Filtering

Filtering merupakan tahapan yang bertujuan untuk menghilangkan kata umum yang biasa sering muncul dalam jumlah banyak[20]. Mengeluarkan kata-kata penting dari keluaran proses token atau menghilangkan *stopwords* juga dapat dianggap sebagai penyaringan[21]. Seperti pada Tabel VI.

TABEL VI
TWEET SETELAH PROSES FILTERING

| Tweet Sebelum Lemmatization | Tweet Setelah Lemmatization |
|--|--|
| [kpu,debat,ketiga,pilpr,2024,akan,digela r,di,istora,senayan,jadi,untuk,debat,keti ga,itu,akan,dilaksanakan,pada,7,januari, tahun,2023,bertempat,di,Istora,tempatny a,di,istora,senayan,kata,komision,kpu,a ugust,mellaz,gt,gt,detikpemilu, debatcapr] | [KPU,Debat,Ketiga,Pilpres,2024,Ak an,Digelar,di,Istora, Senayan,Jadi,untuk,debat,ketiga,itu,a kan,dilaksanakan, pada,7,Januari,tahun,2023,bertempat, di,Istora, tempatnya,di,Istora,Senayan,kata,Ko misioner, KPU, August,Mellaz,gt,gt,detikPemilu, DebatCapres] |

H. Labelling

Pemberian label data ini bertujuan untuk menentukan data yang akan diklasifikasikan kedalam beberapa kelas klasifikasi, label yang telah ditentukan yaitu Sentimen Positif dan Sentimen Negatif[22]. Oleh karena itu, peneliti melakukan labbeling data menggunakan bahasa pemrograman seperti python. Hasil dari labelling ditunjukkan pada Gambar 2 berikut ini:

| cleaned_text | tweets_lem | category_id | compound | sentiment |
|---|---|-------------|----------|-----------|
| KPU Debat Ketiga Pilpres 2024 Akan Digelar di ... | KPU Debat Ketiga Pilpres 2024 Akan Digelar di ... | 0 | 0.4939 | positive |
| Debat Capres kali ini Pak Ganjar awesome bange... | Debat Capres kali ini Pak Ganjar awesome bange... | 1 | 0.6249 | positive |
| antaranews Ganjar Pranowo mengatakan dirinya s... | antaranews Ganjar Pranowo mengatakan dirinya s... | 2 | 0.0000 | negative |
| Mau liat livestream debat kudu join member fc ... | Mau liat livestream debat kudu join member fc ... | 3 | 0.2960 | positive |
| Baru kali ini gue sepakat milik calon capres s... | Baru kali ini gue sepakat milik calon capres s... | 4 | 0.0000 | negative |

Gambar. 2 Hasil labelling data

I. Ekstraksi Fitur

Salah satu langkah preprocessing yang sangat membantu dalam memperkecil Dalam konteks data, prosesnya adalah membuang data yang tidak relevan atau tidak diperlukan dan meningkatkan hasil akurasi adalah seleksi fitur. Dalam rangka meningkatkan efektivitas dan akurasi proses klasifikasi, seleksi fitur berusaha untuk mengecilkan dimensi dari kumpulan teks dengan menghilangkan kata-kata yang dianggap tidak perlu atau tidak cukup menggambarkan isi dokumen. Langkah mengonversi kata menjadi fitur menggunakan pendekatan *Term Frequency-Inverse Document Frequency* (TF-IDF) dikenal dengan istilah seleksi fitur.

Teknik *Term Frequency-Inverse Document Frequency* (TF-IDF) menentukan tingkat keterkaitan sebuah kata dengan dokumen dengan memberikan bobot pada setiap kata. TF-IDF menghitung frekuensi kebalikan dari dokumen yang mengandung sebuah kata dan frekuensi kemunculannya dalam sebuah dokumen dengan menggabungkan konsep *term frequency* (TF) dan *inverse document frequency* (IDF).

J. SVM Implementasi

Tahap berikut dari implementasi model *Support Vector Machine* Berdasarkan data pelatihan, SVM berusaha membangun *hyperplane* optimal dalam ruang fitur untuk membagi dua kelas. Baik data linier maupun *non-linier* dapat diklasifikasikan menggunakan algoritma SVM. SVM dengan kernel linier dapat bekerja dengan baik pada data linier. Di sisi lain, SVM dapat dimodifikasi menggunakan fungsi kernel jika datanya *non-linear*.

Data ditransformasikan menggunakan fungsi kernel dalam dimensi fitur yang lebih tinggi untuk memfasilitasi pemisahan linier dapat berjalan lebih mudah. SVM dapat mengelola korelasi non-linear antara fitur-fitur dalam data berkat fungsi kernel. Kernel polinomial, RBF (*Radial Basis Function*), dan sigmoid adalah beberapa jenis fungsi kernel yang digunakan dalam SVM.

K. Evaluasi Model

Tahap akhir dalam penelitian ini melibatkan penilaian kinerja algoritma melalui proses evaluasi SVM merupakan pemroses teks yang efektif sehingga mampu mencapai tingkat akurasi yang tinggi dalam menganalisis sentimen masyarakat terhadap Pilpres 2024. Untuk mengevaluasi keakuratan dari hasil studi yang dilakukan, diperlukan pengujian. Perhitungan nilai akurasi, presisi, recall, dan f-measure merupakan bagian dari metodologi pengujian data yang dimanfaatkan dalam penelitian ini. Pada fase evaluasi, kriteria-kriteria berikut ini yang dicari:

1. *Rasio* jumlah data bernilai positif yang sebenarnya terhadap semua data yang benar bernilai dilabeli sebagai positif dikenal sebagai presisi. *Precision*. Dapat dilihat pada Persamaan (1) untuk menghitung nilai persisi [23].

$$\text{Precision} = \frac{TP}{TP+FP} \quad (1)$$

2. *Recall* adalah proporsi data yang positif dan akurat terhadap seluruh data yang positif, benar, dan tidak akurat dikategorikan sebagai negatif. Persamaan (2) dapat menghitung nilai *recall* [23].

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2)$$

3. *F-measure* adalah metrik yang menggabungkan *recall* dan *precision* untuk menilai seberapa baik informasi diambil. Hasil perkalian antara akurasi dan *recall*, dibagi dengan jumlah *precision* dan *recall*, lalu dikalikan dua untuk mendapatkan nilai *F-Measure*. Dapat dihitung pada persamaan (3) untuk mencari nilai *F-measure* [23].

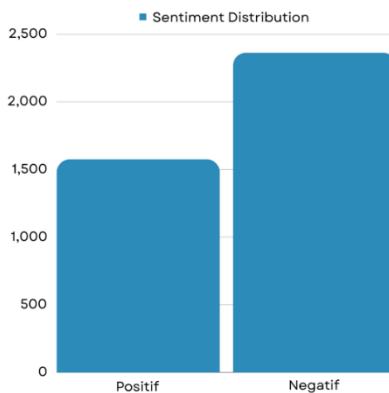
$$F\text{-Measure} = \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

4. *Accuracy* digunakan untuk mengukur performa klasifikasi berdasarkan keakuratan dari metode klasifikasi yang digunakan [24]. perhitungan akurasi, di mana semakin tinggi nilai akurasi, maka metode tersebut dianggap semakin baik. Formula untuk menghitung akurasi dapat dilihat pada persamaan (4) sebagai berikut:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

III. HASIL DAN PEMBAHASAN

Hasil dan analisis kategorisasi algoritma SVM (*Support Vector Machine*) terhadap pemilihan presiden 2024 disajikan di bawah ini. Sebanyak 3938 tweet positif dan negatif digunakan sebagai sumber data penelitian ini kepada calon presiden 2024 dengan kata kunci “pilpres2024”, “debatcapres” dan “capres”.



Gambar. 3 Hasil diagram sentimen positif dan negatif

Pada Gambar 3. Sikap positif diwakili oleh 1575 tweet, sedangkan sentimen negatif terdiri dari 2363 tweet. Dengan menggunakan teknik Support Vector Machine, klasifikasi dilakukan setelah tahap pelabelan.

Kumpulan data yang diperoleh dibagi menjadi data pelatihan dan data uji sebelum model SVM digunakan, dan ekstraksi fitur kemudian dilakukan. Langkah selanjutnya adalah menggunakan pendekatan TF-IDF untuk menghasilkan vektor fitur dengan mempertimbangkan bigram, trigram, dan unigram. Model *Support Vector Machine* (SVM) dengan kernel linear dilatih menggunakan data latih dan vektor fiturnya. Setelah pelatihan, model digunakan untuk melakukan prediksi pada data uji. Evaluasi kinerja model dilakukan dengan menggunakan metrik akurasi dan laporan klasifikasi, yang memberikan wawasan tentang seberapa baik model dapat memprediksi sentimen dengan benar. Hasil evaluasi, seperti akurasi dan laporan klasifikasi, ditampilkan untuk memberikan gambaran yang komprehensif tentang kinerja model dalam memahami dan memprediksi sentimen dari teks.

| Akurasi: 0.950507614213198 | | | | |
|----------------------------|-----------|--------|----------|---------|
| Laporan Klasifikasi: | | | | |
| | precision | recall | f1-score | support |
| negative | 0.95 | 1.00 | 0.97 | 743 |
| positive | 1.00 | 0.13 | 0.24 | 45 |
| accuracy | | | 0.95 | 788 |
| macro avg | 0.98 | 0.57 | 0.60 | 788 |
| weighted avg | 0.95 | 0.95 | 0.93 | 788 |

Gambar. 4 Laporan Klasifikasi Metode SVM

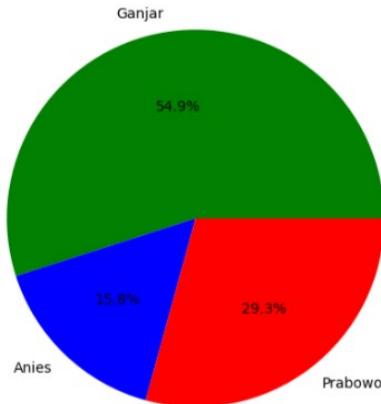
Pada Gambar. 4, Laporan klasifikasi yang diberikan menunjukkan bahwa model pembelajaran mesin berhasil memprediksi sentimen dengan akurasi 95,05%. Terlihat bahwa akurasi data yang mengandung sentimen negatif mencapai 95%, sedangkan 100% data memiliki sentimen yang baik dan akurat. *Recall* untuk data dengan sentimen negatif adalah 100%, tetapi hanya 13% yang dicapai untuk data dengan emosi positif. Data dengan sentimen negatif memiliki *F1-score* sebesar 97%, namun data dengan sentimen baik hanya memiliki *F1-score* sebesar 24%. Data dengan sentimen negatif memiliki 743 dukungan, sedangkan data dengan sikap yang baik memiliki 45 dukungan. Singkatnya, banyak metrik evaluasi yang menunjukkan bahwa model atau analisis berkinerja lebih baik pada sentimen negatif daripada sentimen positif.

```
Accuracy with linear kernel: 0.9543147208121827
Accuracy with poly kernel: 0.950507614213198
Accuracy with rbf kernel: 0.949238578680203
Accuracy with sigmoid kernel: 0.9530456852791879
```

Gambar. 5 Hasil perbandingan kernel

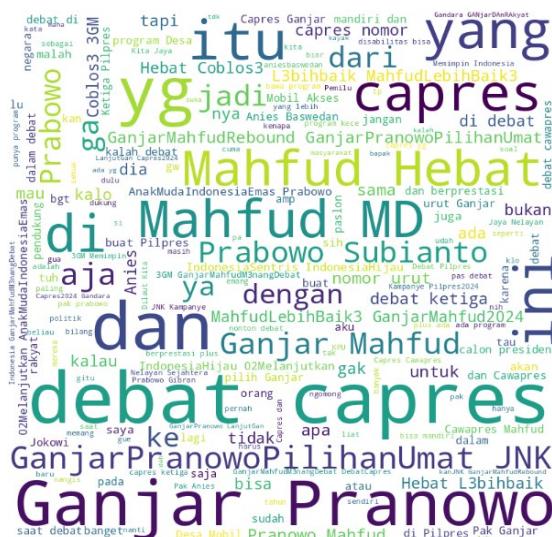
Dalam klasifikasi sentimen pada Gambar. 5 di atas, berbagai jenis kernel Support Vector Machine (SVM) telah diuji, dan hasilnya mengungkapkan tingkat akurasi rata-rata yang mengesankan. SVM dengan kernel linier mencapai akurasi rata-rata sebesar 95,43%, menunjukkan kemampuannya dalam menangani hubungan linier antara fitur-fitur dalam data teks. Sementara itu, SVM dengan kernel RBF (Radial Basis Function) memberikan akurasi

rata-rata sebesar 94.92%, memperlihatkan kemampuannya dalam menangani pola non-linier yang kompleks. Selanjutnya, SVM dengan kernel polynomial menunjukkan performa yang baik dengan akurasi rata-rata sebesar 95.05%, menggambarkan kemampuannya dalam menangkap pola data yang dapat diaproksimasi dengan fungsi polinomial. Di sisi lain, SVM dengan kernel sigmoid mencapai akurasi rata-rata sebesar 95.30%, menunjukkan keefektifannya dalam menangani data yang memiliki pola kompleks dan non-linier. Dengan demikian, hasil eksperimen ini memberikan Gambaran komprehensif tentang keunggulan dan kelemahan masing-masing kernel SVM dalam konteks analisis sentimen, di mana SVM dengan kernel linier menjadi pilihan yang baik dengan akurasi tertinggi.



Gambar. 6 Hasil pie chart sentimen positif terhadap paslon

Berdasarkan analisis sentimen dari data yang diperoleh, terlihat pada Gambar. 6, dengan jelas dalam diagram lingkaran (pie chart) bahwa mayoritas publik menyatakan dukungan positif terhadap Ganjar dengan persentase sebesar 54.9%. Di sisi lain, Anies mendapatkan tingkat dukungan sekitar 15.8%, sementara Prabowo memperoleh dukungan sekitar 29.3%. Diagram ini mencerminkan persepsi dan respons emosional dari masyarakat terhadap masing-masing figur, dengan Ganjar mendominasi perasaan positif yang diungkapkan.



Gambar. 7 Wordcloud kata populer tiga kandidat

Berikut ini adalah pada Gambar. 7 wordcloud istilah yang sering digunakan untuk topik Presiden 2024 di Twitter yang dibuat berdasarkan hasil sentimen dari data yang dikumpulkan dengan kata kunci “pilpres2024”, “debatcapres” dan “capres”, dan hasilnya seperti yang ditunjukkan pada Gambar. 7, yang muncul di dalam wordcloud di atas kata yang mendominasi, Kata-kata seperti “ganjar pranowo”, “Anies”, dan “prabowo subianto”, “debat”, “capres” dan seterusnya.

IV. SIMPULAN

Berdasarkan penelitian yang sudah dilakukan dapat disimpulkan bahwa analisis sentimen menggunakan algoritma SVM dapat diimplementasikan dalam klasifikasi sentimen terhadap calon presiden 2024. Dari 3938 data

tweet yang dianalisis terkait Pilpres 2024, terdapat 2363 tweet dengan sentimen negatif dan 1575 tweet dengan sentimen positif. Hasil dari penelitian ini sangat ideal untuk menerapkan pendekatan SVM karena akurasi yang dicapai oleh analisis sentimen dengan menggunakan metode SVM menghasilkan tingkat akurasi sebesar 95,05%. mayoritas publik memberikan dukungan positif pada Ganjar 54.9%, sementara Anies mendapat 15.8%, dan Prabowo 29.3%. Ganjar mendominasi perasaan positif dari masyarakat. Dari hasil evaluasi model dengan berbagai kernel, Kernel linier memiliki kinerja terbaik dalam evaluasi model dengan akurasi 95.45% dan 95.30% kernel sigmoid sedangkan Kernel polinomial juga memiliki akurasi tinggi 95.05%, tetapi kernel Radial Basis Function (RBF) menunjukkan penurunan performa dengan akurasi 94.92%. Oleh karena itu, kernel linier dianggap lebih baik untuk tugas klasifikasi, namun perlu mempertimbangkan metrik evaluasi lainnya untuk evaluasi yang lebih komprehensif.

Untuk penelitian selanjutnya disaranakan untuk perbandingan antara berbagai algoritma, seperti Maximum Entropy dan Naive Bayes Classifier (NBC). Ini diperlukan untuk memperoleh pemahaman yang lebih tentang performa relatif mereka serta wawasan yang lebih terkait kelebihan dan kecocokan masing-masing pendekatan dalam menangani kompleksitas masalah klasifikasi.

UCAPAN TERIMAKASIH

Terimakasih kepada Universitas Teknokrat Indonesia, Dosen Pembimbing yang telah mendukung dan membimbing dalam penelitian ini. Serta orangtua dan keluarga yang selalu memberikan semangat dan do'a sampai menyelesaikan penelitian ini.

DAFTAR PUSTAKA

- [1] A. Ahmad, "77415-ID-perkembangan-teknologi-komunikasi-dan-in," *Dakwah Tabligh*, vol. 13, pp. 137–149, 2012.
- [2] G. Nugroho, D. Murdiansyah, and K. Lhaksmana, "Analisis Sentimen Pemilihan Presiden Amerika 2020 di Twitter Menggunakan Naïve Bayes dan Support Vector Machine," *e-Proceeding Eng.*, vol. 8, no. 5, pp. 10106–10115, 2021, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15727/15440>
- [3] "Digital-2023-Indonesia-February-2023-V01_Compressed.Pdf." 2023. [Online]. Available: <https://s.id/digital-2023-indonesia>
- [4] S. N. J. Fitriyyah, N. Safriadi, and E. E. Pratama, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 3, p. 279, 2019, doi: 10.26418/jp.v5i3.34368.
- [5] M. I. Khatami, "The Existence Of 'Baliho 2024' In The Electricity Fight: Image Analysis Of Political Figures," *J. Imu Sos. dan Ilmu Polit. Univ. Jambi*, vol. 5, no. 2, pp. 14–24, 2021.
- [6] C. S. Hudaya, H. Fakhrurroja, and A. Alamsyah, "Analisis Persepsi Konsumen Terhadap Brand Go-Jek Pada Media Sosial Twitter Menggunakan Metode Sentiment Analysis Dan Topic Modelling," *J. Mitra Manaj.*, vol. 3, no. 6, pp. 664–673, 2019, doi: 10.52160/ejmm.v3i6.244.
- [7] B. Pang, L. Lee, and S. Jose, "Techniques," 2002.
- [8] B. Santosa, "Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis," *Graha Ilmu Yogyakarta.*, 2007.
- [9] S. Herwasto, I. M. I. Subroto, and B. Assegaf, "Sentiment Analysis of Indonesian Figure using Support Vector Machine," *J. Telemat. Informatics*, vol. 6, no. 3, pp. 230–237, 2018.
- [10] A. L. Hananto, A. P. Nardilasari, A. Fauzi, A. Hananto, B. Priyatna, and A. Y. Rahman, "Best Algorithm in Sentiment Analysis of Presidential Election in Indonesia on Twitter," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 6s, pp. 473–481, 2023.
- [11] D. W. Seno and A. Wibowo, "Analisis Sentimen Data Twitter Tentang Pasangan Capres-Cawapres Pemilu 2019 Dengan Metode Lexicon Based Dan Support Vector Machine," *J. Ilm. FIFO*, vol. 11, no. 2, p. 144, 2019, doi: 10.22441/fifo.2019.v11i2.004.
- [12] B. M. Iqbal, K. M. Lhaksmana, and E. B. Setiawan, "2024 Presidential Election Sentiment Analysis in News Media Using Support Vector Machine," *J. Comput. Syst. Informatics*, vol. 4, no. 2, pp. 397–404, 2023, doi: 10.47065/josyc.v4i2.3051.
- [13] A. M. F. Hulu and K. M. Lhaksmana, "Analisis Sentimen Politik pada Twitter Menggunakan Metode Support Vector Machine (Studi Kasus : Pilpres 2019)," *e-Proceeding Eng.*, vol. 6, no. 2, pp. 2017–2020, 2019.
- [14] G. A. BUNTORO, R. ARIFIN, G. N. SYAIFUDDIIN, A. SELAMAT, O. KREJCAR, and H. FUJITA, "Implementation of a Machine Learning Algorithm for Sentiment Analysis of Indonesia's 2019 Presidential Election," *IIUM Eng. J.*, vol. 22, no. 1, pp. 78–92, 2021, doi: 10.31436/IIUMEJ.V22I1.1532.
- [15] N. Saputra, "(Sentiment Analisys With Lexicon Preprocessing)," *Din. Inform.*, vol. 7, no. 1, pp. 45–57, 2019.
- [16] D. Alita, "Multiclass SVM Algorithm for Sarcasm Text in Twitter," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 1, pp. 118–128, 2021, doi: 10.35957/jatisi.v8i1.646.
- [17] P. Elisa and A. Rahman Isnain, "Comparison of Random Forest, Support Vector Machine and Naive Bayes Algorithms to Analyze Sentiment Towards Mental Health Stigma," *J. Tek. Inform.*, vol. 5, no. 1, pp. 321–329, 2024, [Online]. Available: <https://doi.org/10.52436/1.jutif.2024.5.1.1817>
- [18] I. D. Kurniati et al., *Buku Ajar*. 2015.
- [19] D. Wahyudi, T. Susyanto, and D. Nugroho, "Implementasi Dan Analisis Algoritma Stemming Nazief & Adriani Dan Porter Pada Dokumen Berbahasa Indonesia," *J. Ilm. SINUS*, vol. 15, no. 2, pp. 49–56, 2017, doi: 10.30646/sinus.v15i2.305.
- [20] S. Styawati, N. Hendrastuty, and A. R. Isnain, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, pp. 150–155, 2021, doi: 10.30591/jpit.v6i3.2870.
- [21] N. L. Ratniasih, M. Sudarma, and N. Gunantara, "Penerapan Text Mining Dalam Spam Filtering Untuk Aplikasi Chat," *Maj. Ilm. Teknol. Elektro*, vol. 16, no. 3, p. 13, 2017, doi: 10.24843/mite.2017.v16i03p03.
- [22] D. Alita and R. A. Shodiqin, "Sentimen Analisis Vaksin Covid-19 Menggunakan Naive Bayes Dan Support Vector Machine," *J. Artif. Intell. Technol. Inf.*, vol. 1, no. 1, pp. 1–12, 2023, doi: 10.58602/jaiti.v1i1.20.
- [23] W. P. Anggraini and M. S. Utami, "Klasifikasi Sentimen Masyarakat Terhadap Kebijakan Kartu Pekerja Di Indonesia," *Fakt. Exacta*, vol. 13, no. 4, p. 255, 2021, doi: 10.30998/faktorexacta.v13i4.7964.
- [24] D. Alita, Y. Fernando, and H. Sulistiani, "Implementasi Algoritma Multiclass Svm Pada Opini Publik Berbahasa Indonesia Di Twitter," *J. Tekno Kompak*, vol. 14, no. 2, p. 86, 2020, doi: 10.33365/jtk.v14i2.792.