

ANALISIS SENTIMEN MASYARAKAT INDONESIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE TENTANG PILPRES 2024

Nyongki Alexander Radja Bria, Arita Witanti

Program Studi Informatika, Fakultas Teknologi Informasi, Universitas Mercu Buana Yogyakarta
Jl. Jembatan Merah No. 84C Gejayan, Yogyakarta, Indonesia
nyongkyradja@gmail.com

ABSTRAK

Pemilihan Umum Presiden (Pilpres) merupakan salah satu momen politik terpenting di Indonesia yang terjadi setiap lima tahun sekali. Lebih dari sekadar proses memilih kepala negara, Pilpres juga mencerminkan aspirasi, harapan, dan pandangan masyarakat terhadap pemerintahan dan arah negara. Dalam konteks ini, pemahaman mengenai sentimen masyarakat menjelang Pilpres 2024 menjadi hal yang penting. Penelitian ini difokuskan pada analisis sentimen terkait Pemilihan Presiden 2024 apakah cenderung positif, negatif, atau netral, dengan memanfaatkan data dari platform Twitter. Metodologi yang dipergunakan pada penelitian ini meliputi tahap *crawling data*, pengolahan data, dan analisis sentimen. Data dikumpulkan menggunakan bahasa pemrograman Python pada aplikasi web Google Colab dengan menggunakan API Key Twitter. Metode klasifikasi SVM (*Support Vector Machine*) digunakan untuk mengklasifikasikan data. Dari hasil klasifikasi, didapatkan *accuracy* sebesar 65%, dengan kinerja yang lebih unggul dalam mengidentifikasi sentimen positif, mencapai *precision* 69%, *recall* 81%, dan *f1-score* 74%. Penelitian ini menggambarkan proses lengkap dari hasil pengumpulan data hingga klasifikasi sentimen menggunakan algoritma SVM. Hasilnya menunjukkan bahwa model cenderung lebih baik dalam mengidentifikasi sentimen positif dengan tingkat keakuratan yang cukup baik.

Kata kunci : Pilpres 2024, Twitter, Analisis Sentimen, Support Vector Machine

1. PENDAHULUAN

Pada saat ini, terdapat bermacam-macam platform media sosial seperti Twitter, Facebook, Instagram, dan lain sebagainya. Ketersediaan opsi yang beragam dalam media sosial memberikan kebebasan kepada masyarakat, terutama generasi muda, untuk memilih platform yang sesuai dengan kebutuhan mereka dalam hal komunikasi dan hiburan. Fenomena ini tercermin dari jumlah pengguna yang aktif di Indonesia cukup signifikan, mencapai 79 juta pengguna [1]. Jumlah pemakai media sosial yang ada di Indonesia juga memperlihatkan bahwa jejaring sosial tetap menjadi landasan masyarakat Indonesia untuk sarana komunikasi atau hiburan [2]. *Twitter* merupakan salah satu media sosial yang digunakan pada saat pemilihan umum. Di beberapa negara seperti Amerika Serikat, Jerman, dan Singapura, *twitter* telah digunakan dalam pemilihan umum [3]. *Twitter* saat ini menjadi media sosial yang sangat populer di seluruh dunia yang dapat dimanfaatkan oleh calon pasangan presiden sebagai sarana kampanye untuk menyampaikan citra pasangannya sendiri-sendiri kepada pendukungnya dan kepada calon pemilih.

Pada tahun 2019, calon untuk pemilihan presiden di benua Asia terlihat menggunakan *Twitter* dan menggunakan media sosial lainnya untuk berbagi kebijakan dan slogan, mengumpulkan pendukung, dan menggempur reputasi lawan mereka [4]. Hal itu juga berpengaruh di Negara Indonesia yang termasuk dalam lima besar pemakai social media di dunia, yang mana pengguna Internet di Indonesia mencapai 73,7% dari total penduduk, jumlah ini didapatkan dari hasil survei yang dilakukan oleh APJII (Asosiasi

Penyelenggara Jasa Internet Indonesia) periode 2019 sampai 2020 kuartal kedua [5].

Hasil survei dari Kementerian Kominfo, menunjukkan *Twitter* termasuk dalam lima media sosial yang populer di Indonesia, yakni sebanyak 19,5 juta pemakai [6]. Dilansir dari laman Statista di awal tahun 2022, terdapat sebanyak 18,45 juta pengguna aktif yang menggunakan *Twitter* di Indonesia [7]. Sebelum pemilu dan selama proses pemilu terkait penyelenggaraan pemilu banyak komentar positif dan negatif dari masyarakat.

Twitter menyediakan platform bagi orang-orang yang memiliki opini untuk menyampaikan opininya secara langsung yang menghasilkan banyak jumlahnya, sebagai contoh pada pilpres 2019 dari hasil riset dan *monitoring* Binokular yang dilakukan oleh PT Binokular Media Utama, yang notabene merupakan salah satu perusahaan media riset dan *monitoring* di Indonesia, dalam jangka waktu dari 1 Juli sampai 31 Desember 2018, dalam hal ini tagar #2019GantiPresiden memiliki jumlah tweet sebanyak 1.189.714, #2019PrabowoSandi memiliki 458.306 tweet, #Jokowilagi memiliki 370.409 tweet, #2019TetapJokowi memiliki 345.849 tweet dan #Prabowosandi sebanyak 258.695 tweet (Redaksi FN, 2019), total jumlah tweet dari ke lima tagar tersebut sebanyak 2.622.973 tweet [8].

Opini di *Twitter* tidak dapat langsung diidentifikasi sebagai negatif atau positif. Berbagai informasi yang didapatkan dari *Twitter*, jika ditafsirkan demikian tanpa memverifikasi sumbernya dan menyebarkannya secara andal, memiliki kecenderungan melahirkan berita palsu atau *hoax* justru mengarah pada *black campaign* terhadap calon

lawan dalam pilpres. Dalam sosial media Twitter, kita dapat menemukan komentar, opini atau aspirasi dari publik yang bisa dipergunakan untuk mewakili kejadian terkini terhadap isu terkait pemilihan presiden ini dengan berbagai tagar yang diawali dengan # (tagar atau *hashtag*). Agar opini tersebut bermanfaat dan dapat digunakan, diperlukan beragam cara agar memperoleh informasi yang penting, salah satunya dengan melakukan penelitian analisis sentimen. Analisis Sentimen merupakan metode yang digunakan untuk mengekstraksi, mengidentifikasi, dan memahami sentimen yang terdapat dalam teks, baik itu negatif, netral, atau positif. Dengan menganalisis sentimen masyarakat terhadap pilpres 2024, kita bisa mendapatkan pengertian yang lebih baik akan perasaan, pandangan, dan keyakinan yang mungkin muncul dari masyarakat terkait proses pemilihan tersebut.

Metodologi yang dipergunakan pada penelitian ini melingkupi tahap *crawling data*, pengolahan data, dan analisis sentimen. *Crawling data* dilakukan dengan menggunakan teknik pencarian pendapat masyarakat pada sosial media seperti *Twitter*. Setelah data terkumpul, selanjutnya melakukan proses pengolahan data seperti *preprocessing* dan *feature extraction* untuk mempersiapkan data sebelum dilakukan analisis sentiment [9]. Selanjutnya, dilakukan *Sentiment Analysis* dengan memakai algoritma *Support Vector Machine* untuk mengklasifikasikan opini masyarakat menjadi positif, negatif, atau netral [10].

Dalam konteks ini, penelitian berikut akan menganalisis opini publik terkait pilpres 2024, melalui pertimbangan postingan di *Twitter*. Pada penelitian ini dilakukan pengumpulan data melalui pengambilan dataset dari *Twitter* dengan memanfaatkan fasilitas API [11]. Dataset tersebut diambil menggunakan beberapa kata kunci terkait pilpres 2024 diantaranya “pilpres 2024”, “ganjar pranowo”, “prabowo subiyanto”, “anis baswedan”. Algoritma *Support Vector Machine* dipilih dikarenakan tingkat performa yang tinggi dalam melakukan klasifikasi sentiment analisis [12].

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Pada penelitian yang dilakukan oleh (Darwis, dkk, 2020). Yang membahas mengenai Penerapan Algoritma SVM Untuk Analisis Sentimen Pada Data *Twitter* Komisi Pemberantasan Korupsi Republik Indonesia. Penelitian menyimpulkan bahwa *Support Vector Machine* (SVM) dan ekstraksi fitur menggunakan TF-IDF. Dari 2000 data hasil *twitter crawling*, penelitian ini menghasilkan 1890 data dan 3846 term/kata dari hasil *preprocessing* lalu dihitung nilai dari kemunculan kata untuk labeling yang menghasilkan sentimen positif, negatif dan netral. Berdasarkan hasil pengujian yang dihasilkan, penerapan metode SVM menghasilkan nilai Akurasi sebesar 82% dan menghasilkan sentimen dengan label

negatif lebih besar dengan jumlah 77%, label positif 8% dan label netral 25% [13].

Pada penelitian yang dilakukan oleh (Hendra Tinambunan, dkk, 2022). Membahas tentang Klasifikasi Naive Bayes Dalam Memprediksi Tingkat Kepuasan Mahasiswa Terhadap Pengajaran Dosen Di Program Studi Matematika Fmipa Universitas Sam Ratulangi Manado. Penelitian ini menyimpulkan dari 200 data digunakan sebagai data training dan 100 data digunakan sebagai data testing. Dari 50 data training dengan klasifikasi BAIK sebanyak 46 data dan TIDAK BAIK sebanyak 4 data. Pengujian menggunakan metode Naive Bayes berhasil memprediksi klasifikasi dengan accuracy 0.96 atau 96% dan Sensitivity 100% [14].

Pada Penelitian yang dilakukan oleh (Tineges, dkk, 2020). Membahas tentang Analisis Sentimen Terhadap Layanan Indihome Berdasarkan *Twitter* Dengan Metode Klasifikasi *Support Vector Machine* (SVM). Melakukan *crawling data* dengan modul *tweepy* pada bahasa pemrograman Python dan *twitter API* dengan kata kunci *indihome* mendapatkan dataset sebanyak 10.000. kemudian melakukan *case folding*, *filtering*, *tokenizing* dan *stopwords remove* lalu melakukan pembobotan dengan teknik TF-IDF. Hasil penelitian dengan evaluasi menggunakan *Confusion Matrix*, didapat akurasi sebesar 87% dengan ketepatan antara hasil prediksi dengan data sebenarnya (*precision*) sebesar 86%, tingkat keberhasilan sistem dalam memprediksi sebuah data (*recall*) sebesar 95%, tingkat kesalahan semua data yang diprediksi (*error rate*) sebesar 13%, sedangkan untuk nilai perbandingan rata-rata *precision* dan *recall* (*f1-score*) adalah sebesar 90% [15].

Penelitian lain yang telah dilakukan oleh (Wafa, dkk, 2022). Mengenai Prediksi Penyakit Diabetes Menggunakan Algoritma klasifikasi *Support Vector Machine* (SVM) dengan kernel RBF. Dataset merupakan data dengan jenis wanita dengan keturunan indian *pima* yang memiliki 8 atribut dan sudah memiliki label. Evaluasi dengan teknik *Confusion Matrix* untuk pengujian model *support vector machine* (RBF) mampu memberikan hasil akurasi yang diperoleh sebesar 91.2% untuk accuracy, 93.0% untuk *precision*, 94.3% untuk *recall*, dan 93.7% untuk *f1-scorer*. Hasil pengujian yang dilakukan mendapat hasil akurasi yang sangat baik [16].

2.2. Analisis Sentimen

Analisis Sentimen merupakan teknik yang digunakan untuk label emosional selama konfigurasi komponen tertentu. Analisis sentimen didefinisikan sebagai tugas Ekstraksi Informasi dan *Natural Language Processing* (NLP). Informasi yang diekstraksi di Web selama pengumpulan opini melibatkan data dalam bentuk kata, kalimat, dan dokumen yang terkait dengan berbagai polaritas. Tugas utama analisis sentimen adalah menentukan polaritas dengan klasifikasi untuk kata, frasa, kalimat, dan dokumen tertentu, dll [17].

2.3. Text Mining

Text Mining memiliki tujuan dalam mendapatkan informasi dari sejumlah dokumen. Dengan melakukan pemrosesan, pengelompokan, serta menguraikan data yang bersifat tidak terstruktur dan jumlahnya banyak, *text mining* mampu mendapatkan informasi yang signifikan. Teknik *text mining* dipergunakan dalam menggali informasi berharga dari berbagai dokumen dengan data berupa teks yang tidak terstruktur. Proses ekstraksi informasi pada *text mining* memiliki kemampuan untuk melakukan analisis sentimen yang dapat secara emosional mengklasifikasikan pernyataan ke dalam kategori negatif atau positif. Sasaran dari *text mining* umumnya berupa dokumen yang bersifat semi terstruktur atau tidak terstruktur. *Text mining* secara efisien menghasilkan informasi yang penting dari sejumlah dokumen [18].

2.4. Support Vector Machine

Support Vector Machine merupakan algoritma yang berupaya untuk menemukan hyperplane “terbaik” yang berfungsi sebagai batas antara dua kelas. Hyperplane pada *Support Vector Machine* ini berdasarkan vektor pendukung dan margin. Vektor pendukung adalah semua vektor data yang paling dekat dengan *hyperplane*, sedangkan margin mewakili lebar split *hyperplane*. *Support vector machine* (SVM) dapat melakukan klasifikasi data diskrit dengan cara linier (linear separable) dan nonlinier/non-linearly separable [19].

SVM pada dasarnya dipergunakan dalam mengklasifikasikan data numerik, namun SVM juga sangat cepat dan efisien dalam memecahkan masalah dalam data tekstual. Data teks sangat bagus dalam klasifikasi menggunakan algoritma SVM karena bawaan teks yang cenderung berukuran tinggi, yang tidak terkait tetapi selalu berkaitan satu dengan yang lainnya, dan pada umumnya akan diurutkan secara linier berdasarkan kategori yang berbeda [20]. Teori *Support Vector Machine* secara umum dimasukkan kedalam persamaan (1).

$$y(x, w) = \sum_{i=0}^{m-1} \omega_i \phi_i(x) = w^T \phi(x) \quad (1)$$

Keterangan :

$x = (x_1, x_2, \dots, x_D)^T$ adalah variabel input

$w = (w_0, w_1, \dots, w_D)^T$ adalah parameter

$\phi(x)$ = fungsi basis, M adalah jumlah total parameter dari model [21].

2.5. Confusion Matrix

Confusion matrix merupakan tabel yang dipergunakan dalam evaluasi kinerja model klasifikasi dengan membandingkan perkiraan model terhadap nilai sesungguhnya pada data yang diuji. Melalui perhitungan *confusion matrix*, terdapat berbagai metrik evaluasi, antara lain *True Positive (TP)* yang mengindikasikan jumlah sampel pada kelas positif yang diklasifikasi dengan benar, dan *(TN) True Negative* yang menggambarkan total sampel pada

kelas negatif yang diklasifikasi secara tepat. Di sisi lain, *(FP) False Positive* adalah total sampel pada kelas negatif yang salah terklasifikasi menjadi kelas positif, sedangkan *False Negative (FN)* merujuk pada jumlah sampel pada kelas positif yang salah terklasifikasi menjadi kelas negatif. Hasil dari analisis *confusion matrix* memungkinkan perolehan nilai akurasi model [22]. Rumus *confusion matrix* ditampilkan pada persamaan (2) sampai (5).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$Presicion = \frac{TP}{TP+FP} \quad (4)$$

$$F1 - Score = \frac{2 * Recall * Presicion}{Recall + Presicion} \quad (5)$$

3. METODE PENELITIAN



Gambar 1. Alur penelitian

3.1. Pengumpulan Data

Pengumpulan data pada penelitian ini menggunakan teknik *crawling* data tweet menggunakan fasilitas (API) *Application Program Interface* yang ada pada Twitter untuk mendapatkan parameter yang ingin diteliti, selanjutnya dilakukan proses pengambilan data pada subjek yang ingin diteliti [23]. Dalam kasus ini peneliti mengambil data postingan pada sosial media Twitter yang memiliki kata kunci “pilpres 2024”, “capres 2024”, “ganjar pranowo”, “prabowo subiyanto”, “anis baswedan” yang akan dipergunakan untuk dataset. Jumlah data yang didapatkan sebanyak 5000 tweet yang di ambil pada bulan Oktober 2023.

3.2. Preprocessing

Setelah melakukan pengumpulan data, langkah selanjutnya adalah *preprocessing*. Hasil dari tahap ini akan menyediakan data yang siap digunakan untuk langkah-langkah berikutnya. *Preprocessing* data meliputi beberapa tahap yaitu :

- Cleaning* untuk menghilangkan variabel yang tidak diperlukan seperti URL, simbol dll.
- Case Folding* merubah semua huruf kapital atau huruf besar pada data menjadi huruf kecil agar terdapat keseragaman huruf.
- Tokenizing* untuk memisahkan kalimat menjadi beberapa bagian atau kata.

- d) *Stopword Removal* yaitu metode dalam menghilangkan kata yang tidak diperlukan, contohnya kata hubung pada, yang, di, akan dll.
- e) *Stemming* untuk menghapus kata sambung agar kata tersebut kembali menjadi kata dasar, contohnya kata “berdusta” diubah menjadi kata “dusta”.
- f) Labeling untuk memberi label positif, negatif dan netral.

3.3. TF-IDF

TFIDF merupakan gabungan antara (*TF*) *term frequency* dan (*IDF*) *inverse document frequency*. (*TF-IDF*) *Term Frequency-Inverse Document Frequency* digunakan untuk mengukur apakah sebuah kata atau term itu penting dalam dokumen tertentu. *TF-IDF* adalah metode umum yang dipergunakan dalam pengambilan informasi pemrosesan teks [24].

3.4. Klasifikasi

Pada tahap ini adalah melakukan klasifikasi dataset yang sudah di *preprocessing*. Pengujian dilakukan dengan bantuan ekstensi Google Colab untuk mengetahui *Accuracy*, *Precision* dan *recall*. Dengan pengujian tersebut maka akan mengetahui opini publik apakah cenderung ke positif, negative atau netral.

3.5. Evaluasi

Confusion matrix menggambarkan jumlah hasil prediksi yang benar dan salah dari model dalam bentuk matriks. *Confusion matrix* dipilih sebagai metode evaluasi yang digunakan untuk mengukur keakuratan metode *Support Vector Machine* guna menilai performanya.

4. HASIL DAN PEMBAHASAN

4.1. Crawling Data

Penelitian ini memakai dataset twitter sebagai bahan pengolahan analisis sentiment mengenai Pemilihan Presiden (PILPRES) 2024. Data tersebut diambil menggunakan Bahasa pemrograman *python* dalam web aplikasi *Google Colab*. Untuk proses *crawling* data dari twitter memerlukan API Key yang teregistrasi untuk dapat berinteraksi dengan twitter. Hasil *crawling* ditampilkan pada Gambar 2.

tweet_id	tweet_text	user_id	user_name	user_avatar	user_verified	tweet_created_at	tweet_retweet_count	tweet_reply_count	tweet_like_count
17324147	innalillahi wainnailaihi rojiuun	17324147	innalillahi wainnailaihi rojiuun	https://twitter.com/innalillahi	True	2023-12-07T10:00:00.000Z	0	0	0
17324148	selamatkan bangsa indonesia jangan pilih ya...	17324148	selamatkan bangsa indonesia jangan pilih ya...	https://twitter.com/selamatkanbangsa	True	2023-12-07T10:00:00.000Z	0	0	0
17324149	belum pernah nyoba tp disini dijelaskan ad...	17324149	belum pernah nyoba tp disini dijelaskan ad...	https://twitter.com/nyobaditini	True	2023-12-07T10:00:00.000Z	0	0	0
17324150	ya saya hanya mengikuti pola pikirmu yang ja...	17324150	ya saya hanya mengikuti pola pikirmu yang ja...	https://twitter.com/ya_saya_hanya	True	2023-12-07T10:00:00.000Z	0	0	0
17324151	kabupaten merata utang pusat dan bank dunia pi...	17324151	kabupaten merata utang pusat dan bank dunia pi...	https://twitter.com/kabupatenmerata	True	2023-12-07T10:00:00.000Z	0	0	0

Gambar 2. Hasil Pengambilan data twitter

4.2. Pembersihan Data

Setelah data terkumpul, selanjutnya yaitu Pembersihan Data. Proses Pembersihan data dilakukan untuk menghilangkan informasi yang tidak relevan atau tidak perlu untuk memastikan kualitas penggunaan data dalam analisis. Pada penelitian ini,

menggunakan data text. Setelah melalui serangkaian proses pembersihan data ini, data awal hasil *crawling* sebanyak 5000 tweet menyusut menjadi 4281 data tweet yang sudah bersih dan siap untuk tahap *pre-processing*. Hasil pembersihan data ditampilkan pada Gambar 3.

id	full_text
0	innalillahi wainnailaihi rojiuun
1	selamatkan bangsa indonesia jangan pilih ya...
2	belum pernah nyoba tp disini dijelaskan ad...
3	ya saya hanya mengikuti pola pikirmu yang ja...
4	kabupaten merata utang pusat dan bank dunia pi...
...	...
4276	indonesia maju bersama ganjar mahfud
4277	yg akan saya pilih nanti adalah ganjar pranow...
4278	katanya soekarno mengapa tidak polig...
4279	yg bener kmu puh sepuh
4280	saya juga mas sebagai warga solo saya kecewa ...
4281 rows	

Gambar 3. Hasil pembersihan data

4.3. Hasil Labeling dan Preprocessing

Langkah selanjutnya adalah melakukan pelabelan data menggunakan Pustaka *TextBlob*. *TextBlob* adalah Pustaka sederhana yang mendukung analisis dan operasi kompleks pada data tekstual. Hasilnya akan menampilkan tuple yang terdiri dari nilai sentimen (-0.1 hingga 0.1, dimana -0.1 adalah sentimen negatif, 0 sentimen netral, dan 0.1 sentimen positif).

Data dalam Bahasa Indonesia harus di translate ke Bahasa Inggris menggunakan layanan *Google Translate*. Setelah diterjemahkan, *TextBlob* kemudian digunakan untuk menganalisis sentimen teks yang telah diterjemahkan. Hasilnya akan menampilkan sentimen dari teks dalam bahasa Inggris beserta score sentimen. Setelah mendapatkan score sentiment, Langkah selanjutnya adalah mengklasifikasi score sentiment tersebut ke dalam kategori negatif, netral, atau positif. Hasilnya ditampilkan pada Gambar 4.

id	full_text	text_eng	Score	Sentiment
0	Innalillahi wainnailaihi rojiuun	Innalillahi wainnailaihi rojiuun	0.000000	Netral
1	Selamatkan bangsa Indonesia jangan pilih ya...	Save the Indonesian nation, don't choose the ...	0.000000	Netral
2	belum pernah nyoba tp disini dijelaskan ad...	I've never tried it, but here it's explained how	0.000000	Netral
3	Ya Saya hanya mengikuti pola pikirmu yang ja...	Yes, I'm just following your way of thinking ...	0.050000	Positif
4	Kabupaten Merata Utang Pusat dan Bank Dunia Pi...	Regency equalizes central debt and World Bank...	0.000000	Netral

Gambar 4. Hasil labeling data

Setelah dilakukan proses Labeling tercatat sentimen sebanyak 2300 positif, 581 negatif, dan 1400 netral.

Preprocessing Data adalah tahapan penting dalam analisis data yang bertujuan untuk membersihkan, merapikan, dan merubah data mentah menjadi bentuk yang mudah dipahami dan dapat diolah oleh algoritma pemrosesan data atau model yang akan digunakan dalam analisis selanjutnya. Adapun hasil dari tahap ini dapat di lihat pada Gambar 5.

	full_text	Sentiment
0	innalillahi wainnailaihi rojuun	Netral
1	selamat bangsa indonesia pilih nama cak imin	Netral
2	nyoba tp	Netral
3	ya ikut pola pikir logika barang suruh bukti k...	Positif
4	kabupaten rata utang pusat bank dunia pinjam b...	Netral

Gambar 5. Hasil *preprocessing* data

Selanjutnya, data teks diubah menjadi bentuk numerik untuk diproses oleh algoritma klasifikasi yang disebut sebagai *vectorizing*. Salah satu metode umumnya adalah pembobotan kata, di mana pada penelitian ini digunakan algoritma pembobotan kata. Dalam Python, perintahnya adalah `TfidfVectorizer()`. Hasil pembobotan kata ditampilkan pada Gambar 6.

	aamiinaamiin	aamiincakep	...	zholimim	zilenial	zimpatisan	zionis
0	0.0	0.0	...	0.0	0.0	0.0	0.0
1	0.0	0.0	...	0.0	0.0	0.0	0.0
2	0.0	0.0	...	0.0	0.0	0.0	0.0
3	0.0	0.0	...	0.0	0.0	0.0	0.0
4	0.0	0.0	...	0.0	0.0	0.0	0.0

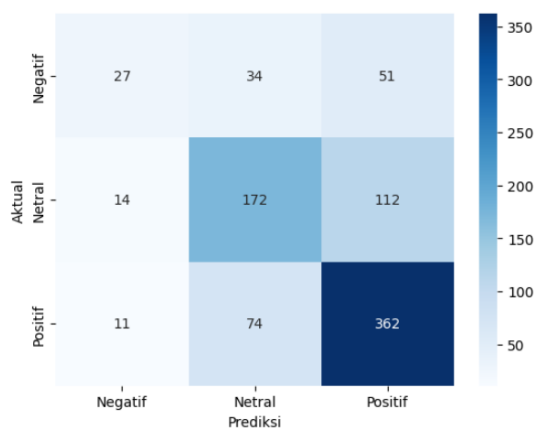
Gambar 6. Hasil pembobotan *tf-idf*

4.4. Hasil Klasifikasi dan Evaluasi

Penggunaan metode dalam penelitian ini adalah *Support Vector Machine* (SVM). Langkah awal dalam proses klasifikasi yaitu membagi data. Peneliti membagi data menjadi 20% data uji dan 80% data latih. Hasil klasifikasi dan Evaluasi ditampilkan pada Tabel 1 dan Gambar 7.

Tabel 1. Hasil klasifikasi

Sentiment	Precision	Recall	F1-Score
Negatif	0.52	0.24	0.33
Netral	0.61	0.58	0.60
Positif	0.69	0.81	0.74
Accuracy	0.65		

Gambar 7. Hasil evaluasi *convusion matrix*

Dari hasil klasifikasi, didapatkan *Accuracy* sebesar 65% dengan pendapat negatif berjumlah 112, netral 298, dan positif 447. Pada *Confusion Matrix*

didapatkan *True Negative* sebanyak 27, *True Neutral* 381, dan *True Positive* 362. Pada sentimen positif didapatkan *precision* sebesar 69%, *recall* 81%, dan *f1-score* 74%. Hal ini menunjukkan bahwa model cenderung lebih baik dalam mengidentifikasi dan memprediksi sentimen positif daripada negatif dan netral.

5. KESIMPULAN

Penerapan metode SVM dalam analisa sentimen tentang PILPRES 2024 berhasil mencapai tingkat *Accuracy* sebesar 65%, yang menunjukkan bahwa model cenderung memprediksi dengan benar sekitar 65% dari keseluruhan dataset. Model memiliki performa yang lebih baik dalam mengidentifikasi sampel positif, dengan *recall* 81% yang lebih tinggi daripada *precision* 69%. Hal ini menunjukkan bahwa model cenderung lebih baik dalam mengidentifikasi sampel yang sebenarnya positif, tetapi masih ada kesalahan prediksi yang termasuk dalam *false positive*. Pada penanganan sentimen positif, model memiliki *f1-score* sebesar 74%, menunjukkan keseimbangan antara *precision* dan *recall* dalam mengidentifikasi sentimen positif. Secara keseluruhan, hasil ini mengindikasikan bahwa model klasifikasi analisis sentimen menggunakan metode SVM memiliki performa yang relatif baik dalam mengenali dan mengklasifikasikan sentimen terkait PILPRES 2024 di Indonesia.

DAFTAR PUSTAKA

- [1] T. & N. A. Rizki Putra, "Pengaruh Media Sosial terhadap Partisipasi Pemilih Pemula pada Pemilihan Presiden 2019: Studi pada Mahasiswa FISIPOL UGM," *Jurnal PolGov*, pp. 2(1), 193–222, 2021.
- [2] W. D. H. T. & P. P. Prasetyo, "Partisipasi Politik Pemilih Pemula Dalam Pemilihan Umum 2019 Di Kelurahan Sumber Kecamatan Banjarsari Kota Surakarta," *Civics Education and Social Science Journal (Cessj)*, p. 1(1), 2019.
- [3] E. & B. Sang, "Predicting the 2011 dutch senate election results with twitter. Proceedings of the Workshop on Semantic Analysis in ...," *Association for Computing Machinery*, pp. 53, 53–60, 2012.
- [4] I. a. S. A. Kurniawan, "Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019," *Jurnal Eksplora Informatika*, pp. 1-10, 2019.
- [5] R. A. S. S. H. W. A. U. J. P. & M. D. Gunawan, "Adiksi Media Sosial dan Gadget bagi Pengguna Internet di Indonesia.," *Techno-Socio Ekonomika*, pp. 4(1), 1., 2021.
- [6] F. Anwar, "Perubahan dan Permasalahan Media Sosial.," *Jurnal Muara Ilmu Sosial, Humaniora, Dan Seni*, pp. 1(1), 137, 2017.

- [7] Statista., "Leading countries based on number of Twitter users as of January 2022," *Statista Research Department*, 2022.
- [8] O. & S. H. Fanny, "Analisis Sentimen Pengguna Media Sosial Terhadap Omnibus Law Berdasarkan Hashtag di Twitter Analysis of Social Media Users Sentiments against Omnibus Law Based on Hashtags on Twitter," *SISTEMASI: Jurnal Sistem Informasi*, pp. 2547-9719, 2022.
- [9] W. G. (. Hernawati Hernawati(1), "SENTIMEN ANALISIS OPERASI TANGKAP TANGAN KPK MENURUT MASYARAKAT MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE, NAÏVE BAYES, BERBASIS PARTICLE SWARM OPTIMIZATION," *journal.lppmunindra*, 2019.
- [10] A. A. R. U. E. Muhammad Diki Hendriyanto, "Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine," *INTECOMS: Journal of Information Technology and Computer Science*, 2022.
- [11] A. T. S. Ulfa Kurniasih, "Analisis Sentimen Terhadap Bantuan Subsidi Upah (BSU) pada Kenaikan Harga Bahan Bakar Minyak (BBM)," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, p. 2236, 2022.
- [12] I. J. J. E. Aditiya Hermawan, "Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine," *Jurnal Sains dan Teknologi*, vol. 12, p. 135, 2023.
- [13] D. P. E. S. & P. A. F. O. Darwis, "Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," *EduTic - Scientific Journal of Informatics Education*, vol. 7(1), p. 1–11, 2020.
- [14] R. T. J. & E. M. C. Hendra Tinambunan, "Klasifikasi naive bayes dalam memprediksi tingkat program studi matematika fmipa universitas sam ratulangi manado," *Prosiding seminar nasional sains dan terapan (SINTA) VI*, p. 141– 149, 2022.
- [15] R. T. A. & S. I. D. Tineges, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *Jurnal Media Informatika Budidarma*, vol. 4(3), p. 650, 2020.
- [16] H. S. H. A. I. U. F. R. T. J. S. J. S. K. C. & C. K. Wafa, "Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM)," *INFORMATICS AND DIGITAL EXPERT*, vol. 1, p. 40–45, 2022.
- [17] V. L. Shilpa Patil, "LIVE TWITTER SENTIMENT ANALYSIS USING STREAMLIT," *SSRN eLibrary*, 2022.
- [18] I. Sabatovych, "Do social media create revolutions? Using Twitter sentiment analysis for predicting the Maidan Revolution in Ukraine," *Sage Journals Home*, 2019.
- [19] D. N. W. A. P. I. E. K. J. S. K. P. U. S. T. D. & M. S. V. Batubara, "Analisis Prediksi Keterlambatan Pembayaran Listrik Menggunakan Komparasi Metode Klasifikasi Decision Tree dan Support Vector Machine," *Jurnal Riset Komputer*, pp. 102-108, 2022.
- [20] M. T. F. B. R. Jumerlyanti Mase, "Penerapan Algoritme Support Vector Machine (SVM) Pada Pengklasifikasian Penyakit Kucing," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, p. 3649, 2018.
- [21] D. r. n. H. Murfi, "123dok.com," 2017. [Online]. Available: <https://123dok.com/document/zkel54pz-course-machine-learning-dr-hendri-murfi-neural-networks.html>. [Accessed 2023].
- [22] D. G. H. D. G. I. Ni Made Ayu Juli Astari, "Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes," *JURNAL SISTEM DAN INFORMATIKA (JSI)*, p. 25, 2020.
- [23] J. E. B. S. R. F. F. F. Yumarlin MZ, "Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi Covid 19 Menggunakan Algoritma Naïve Bayes Classifier," *Electronic Journal Politeknik Harapan Bersama Tegal*, vol. 11, p. 440, 2022.
- [24] W. G. A. N. K. H. F. A. J. Ristyani Slamet, "Analisis Sentimen Twitter Terhadap Penggunaan Artis Korea Selatan Sebagai Brand Ambassador Produk Kecantikan Lokal," *INTECOMS: Journal of Information Technology and Computer Science*, vol. 5, 2022.