

PERBANDINGAN PENGGUNAAN *TFIDFVECTORIZER*, *COUNTVECTORIZER*, DAN *HASHINGVECTORIZER* DENGAN OPTIMALISASI PARAMETER PADA MACHINE LEARNING UNTUK ANALISIS SENTIMEN PEMILU 2024

Febriyanti Panjaitan

Teknik Informatika, Universitas Bina Darma
Jl. A. Yani No. 3 Plaju Palembang, Indonesia
febriyanti_panjaitan@binadarma.ac.id

ABSTRAK

Penggunaan Twitter sebagai platform media sosial telah menjadi semakin penting dalam memahami opini publik terkait berbagai isu, termasuk Pemilu 2024 di Indonesia. Penelitian ini bertujuan untuk menganalisis sentimen terkait Pemilu 2024 di Indonesia menggunakan berbagai teknik vektorisasi teks dan model machine learning yang dioptimalkan. Data yang dikumpulkan terdiri dari 744 tweet dan setelah preprocessing, tweet diklasifikasikan menjadi tiga sentimen: positif, negatif, dan netral, menggunakan library TextBlob. Mayoritas tweet (673) dikategorikan sebagai netral, sementara 37 tweet bersentimen positif dan 34 tweet bersentimen negatif. Untuk meningkatkan akurasi klasifikasi, digunakan model Support Vector Machine (SVM) dengan teknik vektorisasi *TfidfVectorizer*, *CountVectorizer*, dan *HashingVectorizer*, yang dioptimalkan menggunakan Grid Search. Hasil menunjukkan bahwa model SVM yang dioptimalkan mencapai akurasi, presisi, recall, dan F1-score sebesar 100% untuk ketiga teknik vektorisasi. Penelitian ini menunjukkan bahwa optimasi hyperparameter dan pemilihan teknik vektorisasi yang tepat sangat efektif dalam analisis sentimen, memberikan wawasan yang mendalam tentang opini publik terkait Pemilu 2024. Hasil ini dapat digunakan untuk mendukung proses demokrasi yang lebih transparan dan responsif. Hasil ini dapat digunakan untuk mendukung proses demokrasi yang lebih transparan dan responsif..

Kata kunci : Analisis Sentimen, Pemilu 2024, Machine Learning, *TfidfVectorizer*, *CountVectorizer*, *HashingVectorizer*, Support Vector Machine, Grid Sear

1. PENDAHULUAN

Pemilihan umum (pemilu) adalah momen penting dalam sistem demokrasi yang memungkinkan rakyat untuk memilih pemimpin dan wakil mereka. Pemilu 2024 di Indonesia diharapkan menjadi salah satu pemilu terbesar dan paling signifikan dalam sejarah bangsa [1]. Dalam era digital saat ini, media sosial dan platform online memainkan peran penting dalam membentuk opini publik dan persepsi terhadap kandidat serta partai politik. Oleh karena itu, analisis sentimen menjadi alat yang sangat berguna untuk memahami pandangan dan perasaan masyarakat terhadap pemilu [2].

Analisis sentimen adalah proses mengidentifikasi dan mengkategorikan opini yang dinyatakan dalam teks, terutama untuk menentukan apakah sikap penulis terhadap subjek tertentu adalah positif, negatif, atau netral [3]. Metode machine learning telah banyak digunakan dalam analisis sentimen karena kemampuannya untuk menangani volume data yang besar dan kompleks. Namun, kinerja model machine learning sangat bergantung pada representasi fitur teks dan pemilihan hyperparameter yang tepat [4].

Penelitian ini juga melakukan prediksi sentiment analisis dengan tiga metode yang umum digunakan untuk merepresentasikan teks dalam analisis sentimen adalah *TfidfVectorizer*, *CountVectorizer*, dan *HashingVectorizer* [5].

Meskipun masing-masing metode memiliki keunggulan tersendiri, kinerja model machine learning

sangat dipengaruhi oleh pemilihan dan optimasi hyperparameter. Optimasi hyperparameter adalah salah satu teknik yang dapat digunakan untuk meningkatkan performa model machine learning. Hyperparameter adalah parameter yang ditetapkan sebelum proses pelatihan model dan mempengaruhi kinerja model secara signifikan [6].

Dengan melakukan optimasi hyperparameter, model machine learning dapat disesuaikan secara lebih tepat, menghasilkan prediksi sentimen yang lebih akurat dan andal. Proses ini meningkatkan akurasi dan efisiensi dalam analisis sentimen.

Penelitian tentang analisis sentimen pemilu 2024 dengan optimalisasi hyperparameter pada model machine learning yang menggunakan *TfidfVectorizer*, *CountVectorizer*, dan *HashingVectorizer* bertujuan untuk mengevaluasi dan mengidentifikasi kombinasi fitur teks dan hyperparameter yang paling efektif. Dengan pendekatan ini, diharapkan dapat diperoleh model yang mampu menangkap sentimen publik secara akurat, memberikan wawasan yang mendalam tentang persepsi masyarakat selama periode pemilu, dan mendukung pengambilan keputusan yang lebih baik oleh politisi dan pembuat kebijakan.

Penelitian sebelumnya menunjukkan bahwa penggunaan teknik machine learning seperti *Support Vector Machine* (SVM)[7], *Naïve Bayes*[8] dan *K-Nearest Neighbors* (KNN) dalam analisis sentimen memiliki potensi besar dalam memahami dinamika sosial-politik [9], [10].

Namun, hanya sedikit penelitian yang secara khusus membandingkan penggunaan *TfidfVectorizer*, *CountVectorizer*, dan *HashingVectorizer* dengan fokus pada optimalisasi parameter machine learning, terutama pada metode SVM. Metode ini digunakan karena SVM memberikan hasil prediksi analisis sentimen dengan akurasi terbaik dibandingkan dengan ketiga metode machine learning lainnya [11], [12].

Oleh karena itu, penelitian ini diharapkan dapat memberikan wawasan yang lebih mendalam tentang cara-cara terbaik untuk mengelola dan menganalisis data teks terkait Pemilu 2024..

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian terdahulu pertama oleh Raditia Vindua dan rekan-rekan [13] membahas pentingnya Twitter sebagai platform utama untuk membangun sentimen dan preferensi politik masyarakat Indonesia menjelang Pemilu 2024. Penelitian ini menggunakan library TextBlob dalam bahasa pemrograman Python untuk menganalisis tweet-tweet terkait pemilu, yang kemudian diproses dengan proses cleaning dan stemming. Hasil analisis sentimen menunjukkan bahwa sekitar 40% tweet bersentimen positif, 52% netral, dan 8% negatif, mengindikasikan opininya cenderung netral hingga positif terhadap pemilihan umum mendatang.

Penelitian kedua, yang dilakukan oleh Salim Puad dan tim [9], menggunakan algoritma Naïve Bayes dengan metodologi KDD untuk mengeksplorasi pandangan masyarakat terhadap Pemilu 2024. Analisis sentimen ini menghasilkan 331 label positif, 261 label negatif, dan 825 label netral, dengan akurasi tertinggi pada model yang menggunakan pembagian data 90:10, menunjukkan kualitas klasifikasi yang baik dengan nilai AUC 0,71. Pada analisis text menggunakan *TfidfVectorizer*

Studi ketiga oleh Francis Matheos Sarimole dan kolega [14] memfokuskan pada analisis sentimen terhadap isu penundaan Pemilu 2024 di Twitter menggunakan metode Naïve Bayes dan SVM. Hasilnya menunjukkan akurasi sebesar 98,80% untuk SVM dan 91,61% untuk Naïve Bayes, dengan Naïve Bayes memberikan hasil prediksi yang lebih tinggi terhadap sentimen negatif terkait isu tersebut.

Penelitian-penelitian ini memberikan gambaran tentang beragam pendekatan dalam menganalisis sentimen publik terkait pemilu, mulai dari penggunaan algoritma klasifikasi hingga optimalisasi teknik preprocessing untuk meningkatkan akurasi dan interpretasi hasil..

2.2. Support Vector Machine

Support Vector Machine (SVM) adalah salah satu algoritma machine learning yang digunakan untuk klasifikasi dan regresi. Algoritma ini dikenal karena kemampuannya untuk bekerja dengan baik dalam kasus-kasus yang kompleks dan tidak terpisahkan secara linier [15]. Secara keseluruhan, SVM adalah

alat yang kuat untuk klasifikasi dan regresi, terutama ketika data memiliki batasan non-linier yang jelas. SVM menawarkan fleksibilitas dan presisi dalam menangani berbagai jenis data dan masalah klasifikasi.

2.3. *TfidfVectorizer*

TfidfVectorizer yaitu menghitung nilai *Term Frequency-Inverse Document Frequency* (TF-IDF) dari setiap kata dalam dokumen, memberikan bobot yang lebih tinggi pada kata-kata yang muncul jarang namun penting dalam dokumen tertentu [16].

2.4. *CountVectorizer*

CountVectorizer yaitu metode ini menghitung frekuensi kemunculan setiap kata dalam dokumen, menghasilkan representasi vektor yang menunjukkan jumlah kemunculan setiap kata [17].

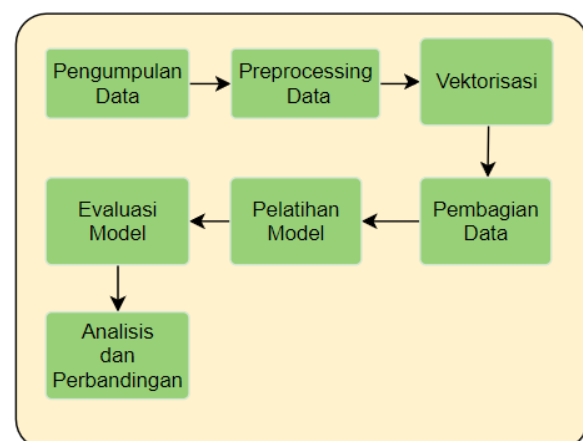
2.5. *HashingVectorizer*

HashingVectorizer adalah metode ini menggunakan fungsi hash untuk memetakan kata-kata ke dalam ruang vektor, yang membantu dalam pengurangan dimensionalitas dan meningkatkan efisiensi komputasi [18].

SVM dapat digunakan untuk mengklasifikasikan tweet setelah tweet diubah menjadi representasi numerik menggunakan *TfidfVectorizer*, *CountVectorizer* dan *HashingVectorizer*[19]. SVM digunakan untuk menentukan apakah tweet tersebut bersentimen positif, negatif, atau netral berdasarkan mayoritas sentimen dari tweet terdekat di data pelatihan.

3. METODE PENELITIAN

Pendekatan penelitian ini akan mencakup beberapa langkah utama, yaitu pengumpulan data, preprocessing, vektorisasi, pembagian data, pelatihan model, evaluasi model, dan analisis dan perbandingan.



Gambar 1. Metode pendekatan penelitian

Berikut adalah penjelasan dari metode pendekatan penelitian ini:

a. Pengumpulan Data

Sumber Data dikumpulkan dari platform media sosial seperti Twitter, yang berisi tweet terkait Pemilu

2024 dari Januari-Juni 2024. Teknik Pengumpulan menggunakan API Twitter atau alat scraping untuk mengumpulkan tweet dalam periode waktu tertentu.

b Preprocessing Data

Langkah-langkah pada tahapan ini meliputi pembersihan data dengan menghapus tautan, tagar, tanda baca, angka, dan karakter khusus. Kemudian melakukan lowercasing untuk mengubah semua teks menjadi huruf kecil untuk konsistensi. Selanjutnya melakukan penghapusan Stopwords dengan cara menghapus kata-kata umum yang tidak berkontribusi pada analisis sentimen yang terakhir melakukan Stemming/Lemmatization untuk mengubah kata ke bentuk dasarnya. Untuk analisis sentiment menggunakan Postext dan TextBlob.

c Vektorisasi

Metode Vektorisasi: Menggunakan tiga teknik berbeda untuk mengubah teks menjadi representasi vector yaitu *TfidfVectorizer*, *CountVectorizer*, dan *HashingVectorizer*

d Pembagian Data

Split Data: Membagi data menjadi set pelatihan (training set) dan set pengujian (testing set), dengan rasio 80:20.

e Pelatihan Model

Algoritma Machine Learning menggunakan beberapa algoritma klasifikasi SVM. Optimasi Hyperparameter: Melakukan tuning hyperparameter menggunakan teknik seperti Grid Search untuk menemukan kombinasi parameter terbaik. Optimasi Grid Search adalah metode untuk menemukan kombinasi hyperparameter terbaik untuk algoritma machine learning. Hyperparameter adalah parameter yang ditetapkan sebelum proses pelatihan model dimulai dan tidak dapat dipelajari dari data. Pemilihan hyperparameter yang tepat sangat penting karena dapat secara signifikan mempengaruhi kinerja model [6]. Hyperparameter yang digunakan pada metode SVM terdapat pada Tabel 1.

Tabel 1. Hyperparameter SVM

No	Metode	Hyperparameter
1	SVM	'C': [0.001, 0.01, 0.1, 1, 10, 100] 'kernel': ['linear', 'rbf', 'poly'], 'gamma': ['scale', 'auto']

f Evaluasi Model

Metrik Evaluasi menggunakan konfusi metrik seperti akurasi, presisi, recall, dan F1-score [20] untuk mengevaluasi kinerja model. K-Fold Cross-Validation untuk melakukan cross-validation untuk memastikan model tidak overfitting [21].

Berikut persamaan dalam metrik evaluasi untuk nilai mendapatkan nilai akurasi, presisi, recall, dan F1-score [20].

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FN + FP} \quad (1)$$

$$\text{Presisi} = \frac{TP}{TP + FP} \quad (2)$$

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

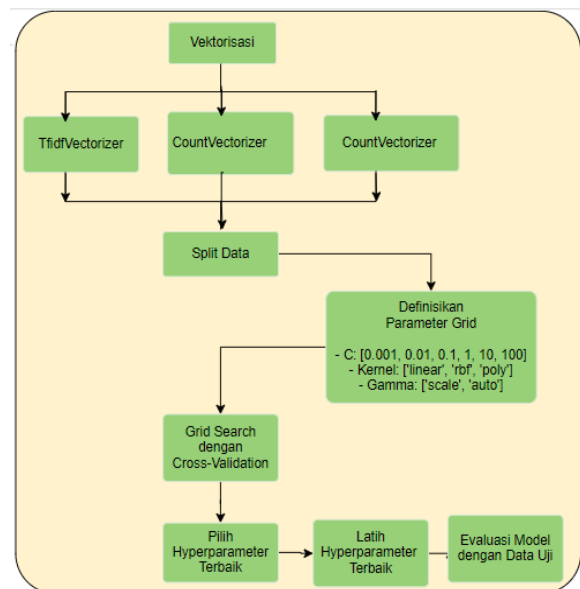
$$F1 - \text{score} = 2x \frac{\text{Presisi} \cdot \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (4)$$

Dimana:

- TP (*True Positive*): Jumlah data positif yang diprediksi benar oleh model. Ini adalah data positif yang sebenarnya dan juga diprediksi benar sebagai positif oleh model.
- TN (*True Negative*): Jumlah data negatif yang diprediksi benar oleh model. Ini adalah data negatif yang sebenarnya dan juga diprediksi benar sebagai negatif oleh model.
- FP (*False Positive*): Jumlah data negatif yang salah diprediksi sebagai positif oleh model. Ini adalah data negatif yang sebenarnya tetapi diprediksi salah sebagai positif oleh model.
- FN (*False Negative*): Jumlah data positif yang salah diprediksi sebagai negatif oleh model. Ini adalah data positif yang sebenarnya tetapi diprediksi salah sebagai negatif oleh model.

g Analisis dan Perbandingan

Analisis Hasil: Membandingkan hasil dari berbagai kombinasi vektorisasi dan model yang telah dioptimalkan.



Gambar 2. Flowchart metode SVM dengan optimasi hyperparameter

Gambar 2 merupakan flowchart tersebut menggambarkan langkah-langkah dalam menggunakan metode SVM dengan optimisasi hyperparameter untuk analisis sentimen. Proses dimulai dengan pengumpulan data tweet terkait Pemilu 2024. Setelah data terkumpul dan preprocessing selesai, dan melakukan vektorisasi, kemudian data dibagi menjadi dua set: data pelatihan dan data pengujian. Ini memungkinkan model untuk dilatih dan diuji secara terpisah, memastikan evaluasi kinerja yang lebih akurat. Kemudian, model SVM diinisialisasi sebagai model dasar yang akan dioptimalkan.

Langkah penting berikutnya adalah mendefinisikan grid hyperparameter yang akan digunakan dalam optimisasi. Dalam hal ini, parameter yang akan dioptimalkan seperti pada Tabel 1. Grid search dengan cross-validation kemudian dilakukan untuk mencari kombinasi terbaik dari hyperparameter ini.

Setelah hyperparameter terbaik dipilih berdasarkan hasil grid search, model SVM dilatih menggunakan kombinasi hyperparameter terbaik tersebut. Model yang telah dilatih ini kemudian dievaluasi menggunakan data pengujian untuk menentukan seberapa baik model dapat mengklasifikasikan tweet ke dalam sentimen positif, negatif, atau netral. Hasil evaluasi dianalisis untuk memahami kinerja model.

Langkah-langkah ini memastikan bahwa model SVM yang digunakan memiliki performa terbaik berkat optimisasi hyperparameter, sehingga dapat memberikan analisis sentimen yang lebih akurat dan andal terkait Pemilu 2024.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Data yang terkumpul terdiri dari 744 tweet terkait Pemilu 2024 yang diambil dari Twitter antara Januari hingga Juni 2024 yang di ambil pada tanggal 28 Mei 2024. Data hasil pengumpulan tweet terdapat pada Gambar 3.

id	id_tweet	created_at	favorite_count	full_text	location	reply_count	username
0	0	Mon Jun 24 17:51:34 +0000 2024	0	Bowo Sidik th menyampaikan ke penduduk KPU ter...	Indonesia	3	AndanarTama
1	1	Sat Feb 10 05:31:36 +0000 2024	36691	Pemilu! Panki jadi KPU? Tes app sifekap ngak...	RUSSIA	4401	UGM_PESS
2	2	Wed Jun 17 07:25:24 +0000 2024	28129	Akun akaakamisan di Twitter hilang. Padahal be...	Rusia	895	hahhhe
3	3	Tue Dec 19 07:41:24 +0000 2023	3802	Me cuti panjang Februari 2024 9 Februari: Lib...	Ring 1	805	idreokidre
4	4	Mon Feb 10 04:49:17 +0000 2024	7287	Update Rekap Nasional Real Count: Sabtu 10 Feb...	Jakarta Capital Region, Indone...	128	Pamitono
...	...	...	...	...	...	...	...
739	739	Thu Jun 13 06:57:51 +0000 2024	0	#TenangPemilu KPU Kabupaten Rembang Kuti Rap...	Rembang, Indonesia	0	KPUKabRembang
740	740	Thu Jun 13 14:47:28 +0000 2024	2	LEBIH TAHU! HUKUM PEMILU Kamis 13 Juni 2024 sal...	Surabaya, Jawa Timur	0	gfr_tasfren
741	741	Thu Jun 13 19:10:29 +0000 2024	0	Kepala Kejaksaan Negeri Bengkulu Resmi Catur...	Rusia	0	KapalBengkulu
742	742	Thu Jun 13 19:11:29 +0000 2024	0	Kepala Kejaksaan Negeri Bengkulu di Wakil Ol...	Rusia	0	KapalBengkulu
743	743	Thu Jun 13 11:44:28 +0000 2024	2	STOP! Indonesia pura pura pemilu? Pilkada pilbu...	Totoli Indonesia di bismayahan	0	PratiFiction

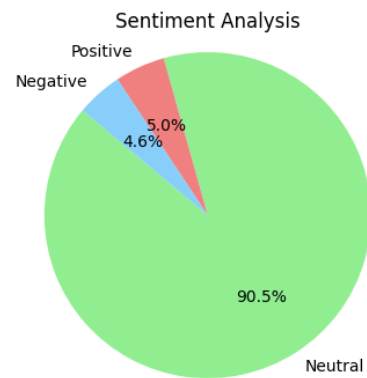
Gambar 3. Data awal Tweet

4.2. Preprocessing Data

Proses preprocessing melibatkan cleaning dan stemming teks tweet. Setelah preprocessing, tweet diklasifikasikan menjadi tiga sentimen: 37 tweet positif, 34 tweet negatif, dan 673 tweet netral. Hasil Crawling yang telah memiliki label terdapat pada Gambar 4 serta Hasil Presentase data terdapat pada Gambar 5, dimana terlihat bahwa sentiment dengan label netral memiliki nilai presentase tertinggi yaitu 90.5%.

tweet	POS tagged	Subjectivity	Polarity	sentiment_label
0 bowo sidik th menyampaikan ke penduduk kpk ter...	[[bowo, n], [sidik, n], [th, n], [menyampaikan...	0.0	0.000000	Neutral
1 pemilu! panki jadi kpps. tes app srekap ngak...	[[pemilu, n], [i, None], [panki, n], [jadi, n]...	0.0	0.000000	Neutral
2 akun akaakamisan di twitter hilang. padahal be...	[[akun, n], [akaakamisan, n], [di, n], [twitte...	0.0	0.000000	Neutral
3 ide cuti panjang februari 2024 9 februari: lib...	[[ide, n], [cuti, n], [panjang, n], [februari, ...	0.0	0.000000	Neutral
4 update rekap nasional real count. sabtu 19 feb...	[[update, a], [rekap, n], [nasional, a], [real...	0.3	0.200000	Positive
...	...	...	...	...
739 tenanpemilih kpu kabupaten rembang kuti rapat...	[[tenanpemilih, n], [kpu, n], [kabupaten, None...	1.0	0.500000	Positive
740 lebih tahu hukum pemilu kams 13 juni 2024 sal...	[[lebih, n], [tahu, n], [hukum, n], [pemilu, n...	0.0	0.000000	Neutral
741 kepala kejaksaan negeri bengalis bapak zainur...	[[kepala, n], [kejaksaan, n], [negeri, n], [be...	0.0	0.000000	Neutral
742 kepala kejaksaan negeri bengalis di wakil ol...	[[kepala, n], [kejaksaan, n], [negeri, n], [be...	0.0	0.000000	Neutral
743 stop indonesia pura pura pemilu? pilkada pilbu...	[[stop, n], [indonesia, n], [pura, n], [pura, ...	0.5	0.333333	Positive

Gambar 4. Hasil preprocessing data



Gambar 5. Presentase Data dengan Label

Hasil analisis menunjukkan bahwa data tweet yang bersentimen netral jauh lebih banyak dibandingkan dengan tweet bersentimen positif dan negatif. Hal ini dapat disebabkan oleh beberapa faktor. Pertama, sebagian besar pengguna media sosial mungkin cenderung menyampaikan informasi atau pendapat secara netral, terutama ketika menyangkut topik yang sensitif seperti Pemilu. Kedua, penggunaan alat seperti TextBlob untuk analisis sentimen sering kali mengklasifikasikan teks yang ambigu atau kurang emosional sebagai netral. Post-tagging (analisis gramatikal) dan TextBlob mungkin menemukan lebih banyak kata atau frasa yang tidak secara jelas menunjukkan sentimen positif atau negatif [23], sehingga lebih banyak tweet diklasifikasikan sebagai netral.

4.3. Vektorisasi

Dalam penelitian ini, penggunaan *TfidfVectorizer*, *CountVectorizer* dan *HashingVectorizer* dengan optimisasi hyperparameter menggunakan grid search menghasilkan parameter terbaik.

4.4. Pembagian Data

Data yang telah diproses dibagi menjadi data pelatihan dan data pengujian dengan proporsi 80:20. Dari 744 tweet yang tersedia, 595 tweet digunakan sebagai data pelatihan dan 149 tweet digunakan sebagai data pengujian. Pembagian data ini terdapat pada Tabel 2.

Tabel 2. Parameter terbaik masing-masing vektorisasi

Jenis Data	Netral	Positif	Negatif
Data pelatihan	538	31	26
Data pengujian	135	6	8
Jumlah	673	37	34

4.5. Inisialisasi Model SVM

Model SVM diinisialisasi dengan parameter dasar yang dengan meningkatkan akurasi analisis sentimen, dilakukan optimasi hyperparameter yang terdapat pada Tabel 1, dengan menggunakan metode Grid Search dan kross validaton pada algoritma SVM

pada tiga jenis vektorisasi teks: *TfidfVectorizer*, *CountVectorizer*, dan *HashingVectorizer*.

4.6. Optimasi Hyperparameter dengan Grid Search

Hasil dari optimasi hyperparameter menunjukkan parameter terbaik untuk masing-masing vektorisasi, hasil ini terlihat pada Tabel 3.

Tabel 3. Parameter terbaik masing-masing vektorisasi

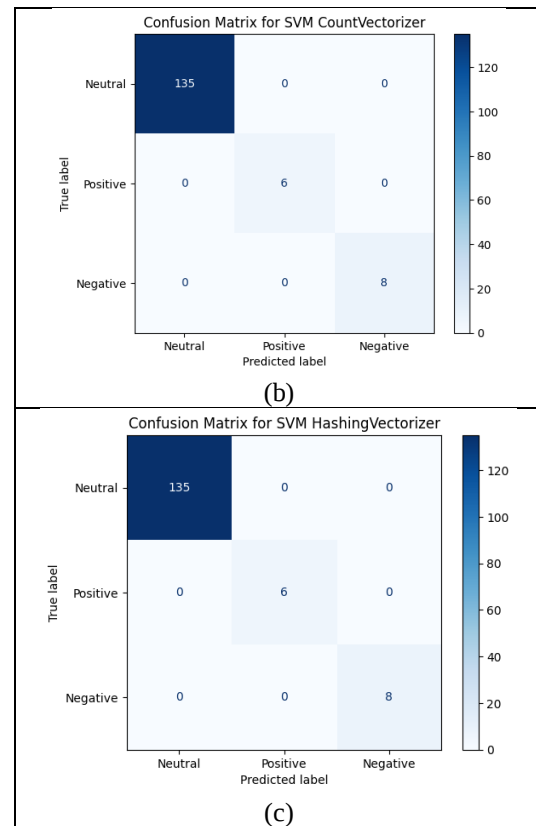
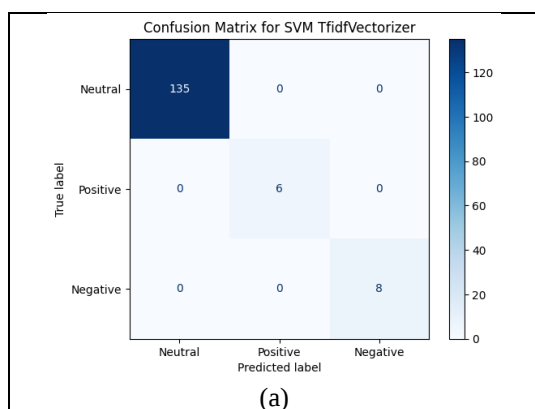
No	Vektrosisasi	Hyperparameter
1	<i>TfidfVectorizer</i>	'C': 10, 'gamma': 'scale', 'kernel': 'linear'
2	<i>CountVectorizer</i>	'C': 1, 'gamma': 'scale', 'kernel': 'linear'
3	<i>HashingVectorizer</i>	'C': 10, 'gamma': 'scale', 'kernel': 'linear'

4.7. Evaluasi Model pada Data Pengujian

Hasil dari confusion matrix menunjukkan performa klasifikasi model pada data pengujian. Dari 159 tweet dalam data pengujian, label yang terklasifikasi adalah sebagai berikut:

- Netral: 135 tweet, yang semuanya terklasifikasi dengan benar sebagai netral.
- Positif: 6 tweet, yang semuanya terklasifikasi dengan benar sebagai positif.
- Negatif: 8 tweet, yang semuanya terklasifikasi dengan benar sebagai negatif.

Hasil ini mengindikasikan bahwa model memiliki tingkat akurasi yang sangat tinggi, dimana setiap tweet dalam data pengujian berhasil diklasifikasikan sesuai dengan label aslinya. Konfusi matrik yang menunjukkan hasil ini dapat dilihat pada Gambar 6.



Gambar 6. Konfusi Matrik 3 Vektorisasi (a) *TfidfVectorizer*, (b) *CountVectorizer*, (c) *HashingVectorizer*

Untuk mengukur akurasi, presisi, recall, dan F1-score berdasarkan data dari confusion matrix, di mana ketiga teknik vektorisasi berhasil mengklasifikasi dengan benar untuk semua sentimen, yaitu 135 netral, 6 positif, dan 8 negatif, kita dapat menetapkan nilai TP, TN, FP, dan FN sebagai berikut:

- TP: Jumlah data positif yang diprediksi benar oleh model = 6
- TN: Jumlah data negatif yang diprediksi benar oleh model = 8
- FP: Jumlah data negatif yang salah diprediksi sebagai positif oleh model = 0 (karena tidak ada data negatif yang salah diprediksi sebagai positif)
- FN: Jumlah data positif yang salah diprediksi sebagai negatif oleh model = 0 (karena tidak ada data positif yang salah diprediksi sebagai negatif)

Berdasarkan nilai TP, TN, FP, dan FN serta menggunakan persamaan (1), (2), (3), dan (4), dalam kasus ini di mana semua data terklasifikasi dengan benar, berikut hasil dari nilai akurasi, presisi, recall, dan F1-Score:

$$Akurasi = \frac{135 + 6 + 8}{135 + 6 + 8} = 1.0 = 100\%$$

$$Presisi = \frac{6}{6} = 1.0 = 100\%$$

$$Recall = \frac{6}{6} = 1.0 = 100\%$$

$$F1 - Score = 2 \times \frac{1.0 \times 1.0}{1.0 + 1.0} = 2 \times \frac{1.0}{2.0} = 1.0 = 100\%$$

Setelah menghitung nilai berdasarkan persamaan matriks konfusi, diperoleh hasil evaluasi yang sangat baik dengan akurasi, presisi, recall, dan F1-score mencapai 100% untuk semua jenis vektorisasi. Hasil ini mengindikasikan bahwa model yang dikembangkan mampu secara efektif mengklasifikasi sentimen tweet terkait Pemilu 2024.

Tabel 4. Tabel Hasil Evaluasi

No	Vektorisasi	Hasil Konfusi Matrik (%)			
		Acc.	Pre.	Rec.	F1
1	<i>TfidfVectorizer</i>	100	100	100	100
2	<i>CountVectorizer</i>	100	100	100	100
3	<i>HashingVectorizer</i>	100	100	100	100

4.8. Analisis Hasil

Hasil penelitian ini menunjukkan bahwa model SVM yang dioptimalkan mampu menangani data tweet dengan sangat baik, tanpa adanya kesalahan klasifikasi pada data pengujian. Optimasi hyperparameter memungkinkan pemilihan parameter yang paling cocok untuk dataset tertentu, yang dalam kasus ini, membantu meningkatkan kinerja model secara signifikan. Dengan demikian, ketiga teknik vektorisasi teks yang digunakan bersama dengan SVM yang dioptimalkan menunjukkan efektivitas yang sangat tinggi dalam analisis sentimen tweet terkait Pemilu 2024.

Penelitian ini memberikan hasil yang signifikan dalam konteks Pemilu 2024, yaitu menyediakan analisis yang lebih akurat dan berbasis data tentang suasana politik pemilu. Secara keseluruhan, penelitian ini berkontribusi pada peningkatan pemahaman tentang opini publik dan menyediakan alat yang efektif untuk analisis sentimen dalam konteks pemilu, yang pada akhirnya mendukung proses demokrasi yang lebih transparan dan responsif.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengidentifikasi sentimen dari tweet terkait Pemilu 2024 dengan menggunakan tiga teknik vektorisasi teks (*TfidfVectorizer*, *CountVectorizer*, dan *HashingVectorizer*) dan model SVM yang dioptimalkan dengan Grid Search. Dari data yang dikumpulkan dan diproses, mayoritas tweet diklasifikasikan sebagai netral, sementara tweet dengan sentimen positif dan negatif lebih sedikit. Dengan optimasi hyperparameter, model SVM mencapai akurasi, presisi, recall, dan F1-score yang sempurna (100%) untuk ketiga teknik vektorisasi, menunjukkan kinerja yang sangat baik dalam klasifikasi sentimen.

Hasil ini menunjukkan bahwa penggunaan teknik vektorisasi teks yang tepat dan optimasi hyperparameter dapat meningkatkan efektivitas analisis sentimen, memberikan wawasan yang lebih mendalam tentang opini publik terkait Pemilu 2024. Untuk meningkatkan akurasi dan generalisasi model, disarankan untuk mengumpulkan data tweet dalam jumlah yang lebih besar dan dalam jangka waktu yang

lebih lama. Hal ini dapat mencakup periode sebelum, selama, dan setelah pemilu. Selain SVM, disarankan untuk mengeksplorasi dan membandingkan metode machine learning dan deep learning lainnya, seperti Random Forest, Gradient Boosting, atau model berbasis neural network seperti LSTM dan BERT, untuk melihat apakah ada peningkatan dalam kinerja model.

DAFTAR PUSTAKA

- [1] Y. Sukma *et al.*, "Peran Kelompok Kaliber Dalam Pengawasan Pemilu Tahun 2024 Untuk Mewujudkan Demokrasi Bersih," *J. Pengabd. Kpd. Masy. Nusantara*, vol. 5, no. 2, pp. 1795–1802, 2024.
- [2] I. Arifin, M. R. S. Ahmad, J. Arifin, A. D. M. P. Agustang, and A. Sadriani, "Peran Media Sosial Dalam Mempengaruhi Keputusan Pemilih Pemula Pada Pemilu 2024 di Indonesia," *Equilib. J. Pendidik.*, vol. 12, no. 2, pp. 201–208, 2024.
- [3] Y. Cahyono and S. Saprudin, "Analisis sentiment tweets berbahasa sunda menggunakan naive bayes classifier dengan seleksi feature chi squared statistic," *J. Inform. Univ. Pamulang*, vol. 4, no. 3, pp. 87–94, 2019.
- [4] R. A. Yudhanegara, N. A. Hana, S. Y. Mahfiridho, and A. R. Kardian, "Perbandingan Resident Set Size dan Virtual Memory Size Algoritma Machine Learning dalam Analisis Sentimen," *J. MEDIA Inform. BUDIDARMA*, vol. 8, no. 1, pp. 371–382, 2024.
- [5] N. U. Rahmanulloh and I. Santoso, "Delineation of The Early 2024 Election Map: Sentiment Analysis Approach to Twitter Data," *J. Online Inform.*, vol. 7, no. 2, pp. 226–235, 2022.
- [6] W. Nugraha and A. Sasongko, "Hyperparameter Tuning on Classification Algorithm with Grid Search," *Sist. J. Sist. Inf.*, vol. 11, no. 2, pp. 391–401, 2022.
- [7] W. Silalahi and A. Hartanto, "Klasifikasi Sentimen Support Vector Machine Berbasis Optimasi Menyambut Pemilu 2024," *JRST (Jurnal Ris. Sains dan Teknol.)*, vol. 7, no. 2, pp. 245–255, 2023.
- [8] E. Priansyah and T. Sutabri, "Analisis Sentimen Berbasis Naïve Bayes Pada Media Sosial Twitter Terhadap Hasil Pemilu Indonesia 2024," *IJM Indones. J. Multidiscip.*, vol. 2, no. 3, pp. 128–138, 2024.
- [9] S. Puad, G. Garono, and A. S. Y. Irawan, "Analisis Sentimen Masyarakat Pada Twitter Terhadap Pemilihan Umum 2024 Menggunakan Algoritma Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 3, pp. 1560–1566, 2023.
- [10] W. Septian and F. M. Sarimole, "Komparasi Analisis Sentimen Masyarakat Terhadap Isu Penundaan Pemilu 2024 Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine," *J. Sains dan Teknol.*, vol. 5, no. 3, pp. 890–899, 2024.

- [11] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan metode naïve bayes dan support vector machine pada analisis sentimen twitter," *SMATIKA J. STIKI Inform. J.*, vol. 10, no. 02, pp. 71–76, 2020.
- [12] W. W. H. Cholil, F. Panjaitan, F. Ferdiansyah, A. Arista, R. Astriratma, and T. Rahayu, "Comparison of Machine Learning Methods in Sentiment Analysis PeduliLindungi Applications," in *2022 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, 2022, pp. 276–280.
- [13] R. Vindua and A. U. Zailani, "Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python," *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 2, pp. 479–487, 2023.
- [14] U. R. H. Baba, "Analisa Sentimen Menjelang Komparasi Analisis Sentimen Masyarakat Terhadap Isu Penundaan Pemilu 2024 Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine," *Innov. J. Soc. Sci. Res.*, vol. 4, no. 3, pp. 11972–11990, 2024.
- [15] R. F. Putra *et al.*, *Data Mining: Algoritma dan Penerapannya*. PT. Sonpedia Publishing Indonesia, 2023.
- [16] A. Rahman, W. Wiranto, and A. Doewes, "Online news classification using multinomial naïve bayes," *ITSMART J. Teknol. dan Inf.*, vol. 6, no. 1, pp. 32–38, 2017.
- [17] R. Vincent, I. Maulana, and O. Komarudin, "Perbandingan Klasifikasi Naive Bayes Dan Support Vector Machine Dalam Analisis Sentimen Dengan Multiclass Di Twitter," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 4, pp. 2496–2505, 2023.
- [18] R. Roshan, I. A. Bhacho, and S. Zai, "Comparative Analysis of TF-IDF and Hashing Vectorizer for Fake News Detection in Sindhi: A Machine Learning and Deep Learning Approach," *Eng. Proc.*, vol. 46, no. 1, p. 5, 2023.
- [19] A. Alex, "Machine Learning Based System Health Check Analyzer For Energy Components." University of Stavanger, Norway, 2018.
- [20] D. Ananda and R. R. Suryono, "Analisis Sentimen Publik Terhadap Pengungsi Rohingya di Indonesia dengan Metode Support Vector Machine dan Naïve Bayes," *J. MEDIA Inform. BUDIDARMA*, vol. 8, no. 2, pp. 748–757, 2024.
- [21] F. Azzahra, N. Suarna, and Y. A. Wijaya, "Penerapan Algoritma Random Forest Dan Cross Validation Untuk Prediksi Data Stunting," *KOPERTIP Sci. J. Informatics Manag. Comput.*, vol. 8, no. 1, pp. 1–6, 2024.