

Perbandingan Metode Naïve Bayes Classifier dan Support Vector Machine dalam Analisis Sentimen Terhadap Pemilihan Presiden 2024

Alga Prananda, Elin Haerani, Muhammad Fikry, Febi Yanto
Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri
Sultan Syarif Kasim Riau, Pekanbaru, Indonesia
*E-mail koresponden: apdevs18@gmail.com

Diserahkan 30 Oktober 2023; Direview 23 November 2023; Dipublikasikan 17 Desember 2023

Abstrak

Salah satu sarana masyarakat untuk memberikan pendapat atau opini adalah menggunakan media sosial, khususnya youtube. Pada penelitian ini berfokus melakukan analisis sentimen terhadap Pemilihan Presiden 2024 dengan tiga kelas dan 2000 data opini, mendapatkan 875 kelas positif, 577 negatif, dan 548 netral. Tahapan penelitian melibatkan pengumpulan data, *pre-processing* (*case folding, tokenizing, filtering, stemming*), klasifikasi, pengujian, dan evaluasi. Juga melakukan perbandingan antara metode *Naive Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM), menunjukkan bahwa SVM mendapat akurasi lebih baik dari NBC di setiap tipe pembagian kelas. Selain itu, hasil analisis sentimen menggunakan empat kata kunci menunjukkan dominasi sentimen positif terhadap Anies Baswedan (80.54%), Prabowo Subianto (64.76%), Calon Presiden secara umum (33.91%), dan Ganjar Pranowo (36.17%). Sentimen negatif cenderung tinggi untuk Ganjar Pranowo (51.42%) dan Prabowo Subianto (25.99%), sementara Anies Baswedan dan Calon Presiden memiliki tingkat sentimen negatif yang lebih rendah (16.53% dan 25.22%). Sentimen netral tercatat pada Prabowo Subianto (9.25%), Ganjar Pranowo (12.41%), Calon Presiden secara umum (40.87%), dan Anies Baswedan (2.93%).

Kata kunci: *Naïve Bayes Classifier; Pemilihan Presiden; Support Vector Machine*

Abstract

One way for people to express their opinions is through social media, particularly YouTube. This study focuses on sentiment analysis for the 2024 Presidential Election using three classes and 2000 opinion data, resulting in 875 positive, 577 negative, and 548 neutral classes. The research stages included data collection, pre-processing (case folding, tokenizing, filtering, and stemming), classification, testing, and evaluation. Comparison between the Naive Bayes Classifier (NBC) and Support Vector Machine (SVM) methods shows that SVM achieved the best accuracy with 0.905 (90:10), 0.900 (80:20), and 0.867 (70:30), while NBC obtained an accuracy of 0.840 (90:10), 0.850 (80:20), and 0.831 (70:30). In addition, the results of sentiment analysis using four keywords show the dominance of positive sentiments towards Anies Baswedan (80.54%), Prabowo Subianto (64.76%), presidential candidates in general (33.91%), and Ganjar Pranowo (36.17%). Negative sentiment tended to be higher for Ganjar Pranowo (51.42%) and Prabowo Subianto (25.99%), whereas Anies Baswedan and Presidential candidates had lower levels of negative sentiment (16.53% and 25.22%, respectively). Neutral sentiments were recorded for Prabowo Subianto (9.25%), Ganjar Pranowo (12.41%), presidential candidates (40.87%), and Anies Baswedan (2.93%).

Keywords: *Naïve Bayes Classifier; Presidential Election; Support Vector Machine*

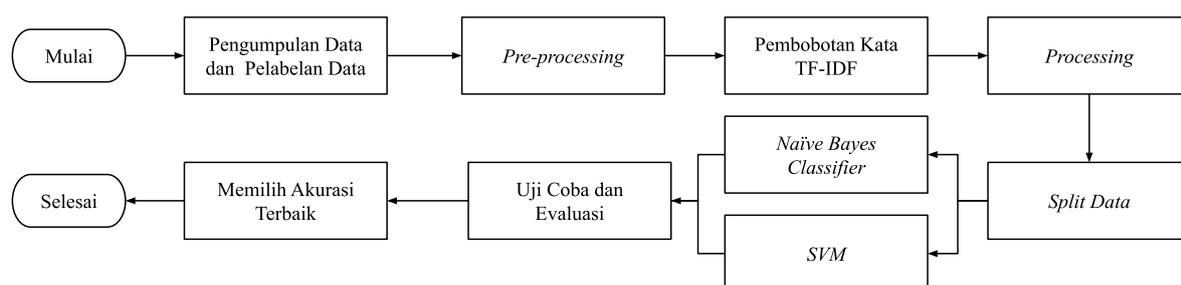
PENDAHULUAN

Analisis sentimen adalah proses untuk memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini [1]. Salah satu metode klasifikasi yang sering digunakan dalam analisis sentimen adalah *Naive Bayes Classifier*. Pendekatan metode ini termasuk dalam kategori pembelajaran probabilistik berdasarkan teorema *Bayes* yang sering digunakan dalam *Natural Language Processing* (NLP). Metode ini berfokus pada konsep frekuensi kata, yaitu seberapa sering kata-kata yang terkait topik diskusi Pemilihan Presiden 2024 muncul dalam suatu dokumen [2]. Tidak hanya itu, metode klasifikasi yang paling umum juga digunakan dalam analisis sentimen yaitu *Support Vector Machine* (SVM), Metode SVM melakukan klasifikasi dengan mencari *hyperplane* (bidang pemisah) yang terbaik untuk memisahkan data ke dalam berbagai kelas sentiment [3]. Dalam penelitian terkait melakukan analisis sentimen terhadap Calon Presiden 2019 menggunakan *Naive Bayes*. Penelitian tersebut mencapai akurasi 64,6% untuk Paslon 01 dan 58% untuk Paslon 02 dalam pengujian dua kelas, serta 77,7% untuk Paslon 01 dan 84% untuk Paslon 02 dalam pengujian dua kelas [4]. Penelitian lainnya membandingkan metode Klasifikasi *Naive Bayes* dan *Support Vector Machine* untuk melakukan analisis sentimen dalam layanan komplain mahasiswa [5]. Hasil evaluasi menunjukkan bahwa SVM mencapai akurasi 84,45%, sementara *Naive Bayes* memiliki akurasi 69,75%. Penelitian klasifikasi berita *hoax* menggunakan dua metode yaitu *Naive Bayes Classifier* dan *Support Vector Machine* dengan empat model berbeda. Dataset terdiri dari 287 berita (200 valid, 87 *hoax*). Evaluasi model menunjukkan bahwa metode *Naive Bayes* memiliki performa tertinggi dengan *precision* 95.4%, *recall* 95.4%, *f1-score* 95.4%, dan akurasi 93.1%. Pada algoritma SVM dengan *kernel Sigmoid*, *precision* mencapai 95.6%, *recall* 100%, *f1-score* 97.7%, dan akurasi 96.5% yang menyatakan metode SVM lebih baik dari *Naive Bayes* [6]. Tujuan penelitian ini untuk memahami pandangan publik terhadap berbagai aspek pemilihan presiden, termasuk pengaruh hasil pemilihan dan dukungan terhadap calon presiden tertentu. Selain itu, penelitian ini berupaya mendeteksi isu-isu sensitif yang dapat mempengaruhi opini masyarakat, dengan harapan dapat memberikan wawasan kepada tim kampanye untuk merumuskan strategi yang efektif dan merespons secara cepat terhadap isu-isu yang mungkin muncul selama pemilihan. Manfaat dari penelitian ini diharapkan dapat memberikan pemahaman mendalam tentang informasi yang berharga bagi tim kampanye dalam mengambil keputusan dan responsif terhadap dinamika opini masyarakat seiring berjalannya proses pemilihan presiden.

METODE PENELITIAN

Perancangan Sistem

Ada beberapa langkah yang terlibat dalam menganalisis sentimen seperti yang ditunjukkan pada Gambar 1.



Gambar 1 Tahapan perancangan sistem

Analisis sentimen dilakukan beberapa tahapan dalam perancangan sistem yaitu:

Pengumpulan dan Pelabelan Data

Penelitian ini menggunakan *crawling* data untuk mengumpulkan data yang telah disediakan menggunakan bantuan sistem pada *website* yaitu situs web *Netlytic* untuk mempermudah dalam pengambilan data pada media sosial. Proses pengambilan data dari komentar youtube, jumlah maksimal data yang diambil menggunakan web *Netlytic* adalah 2000 data. Pengumpulan data diambil secara acak dengan kriteria banyak ditonton oleh masyarakat menggunakan kata kunci yang relevan yaitu Pemilihan Presiden 2024 sehingga menghasilkan data di dalam dokumen. Pelabelan dilakukan dan divalidasi oleh guru Bahasa Indonesia di SMAN 1 Sungai Apit dengan pembagian menjadi tiga kelas yaitu positif, negatif dan netral. Data positif adalah umpan balik dari pengguna seperti arti kata makna setuju, sedangkan negatif adalah komentar yang mengandung hal-hal yang tidak baik dan terakhir netral adalah komentar yang tidak relevan dengan pembahasan dalam penelitian.

Data yang digunakan dalam penelitian ini menggunakan data dari teks komentar youtube yang dibagi menjadi tiga kelas sentimen positif, negatif dan netral seperti yang ditampilkan pada Tabel 1.

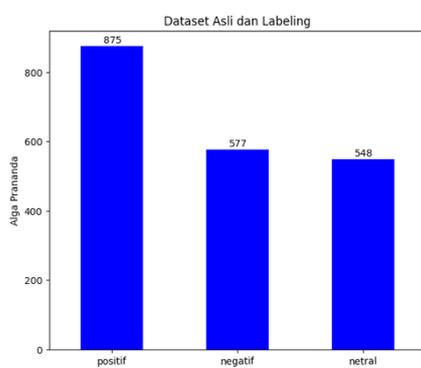
Tabel 1 Jumlah data per kelas

No	Kelas	Jumlah
1	Positif	875
2	Negatif	577
3	Netral	548
Total		2000

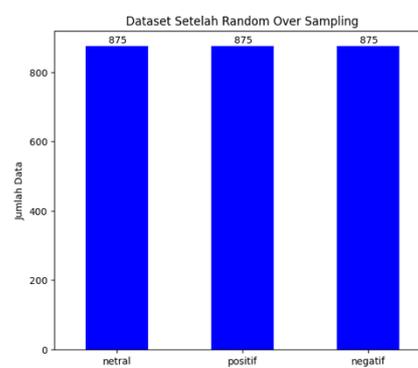
Pada Tabel 1 menunjukkan bahwa kelas positif mendapatkan 875 data, kelas negatif mendapatkan 577 data, dan kelas netral mendapatkan 548 data.

Balancing

Tahapan *balancing* pada penelitian ini menggunakan teknik *Random Over Sampling* untuk menjadikan data seimbang. *Random over sampling* digunakan untuk menyeimbangkan data minoritas menjadi sebanyak data mayoritas. Pada Gambar 2 menjelaskan keseluruhan data dari setiap kelas yang tidak seimbang sehingga membutuhkan teknik *balancing* data untuk membuat data tersebut menjadi seimbang dengan hasil pada Gambar 3 menjadi 875 data yang telah disesuaikan dengan data yang paling banyak yaitu pada kelas positif.



Gambar 2 Sebelum *Random Over Sampling*



Gambar 3 Setelah *Random Over Sampling*

Pre-Processing

Pre-processing adalah langkah dimana data diubah sehingga sesuai dengan format yang benar dan dapat diproses. Ada beberapa langkah *pre-processing* teks dalam melakukan analisis

sentimen, tergantung dalam kebutuhan. Pada penelitian ini menggunakan beberapa langkah, yaitu *cleaning text*, *case folding*, *tokenizing*, *filtering*, dan *stemming* [7]. *Cleaning* merupakan tahap yang diperlukan untuk menghapus tanda baca, angka, simbol, tautan URL, dan nama pengguna yang terdapat dalam teks. *Case folding* adalah proses mengubah yang terdapat huruf besar menjadi huruf kecil. Pada tahap ini menggunakan *case folding* untuk membuat teks komentar menjadi konsisten dan teratur. Hasil perubahan dapat dilihat pada Tabel 2.

Tabel 2. Hasil *case folding*

Sebelum	Sesudah	Kelas
Sudah seharusnya Indonesia di Pimpin oleh Tokoh Muslim yg Amanah dan Terbaik karena Sudah Terbukti Kinerja dan Prestasi Pak Anies selama jadi Gubernur DKI Jakarta. Semoga Indonesia ini lebih Baik dan lebih ber Matabat dan Adil dari sebelum nya	sudah seharusnya indonesia di pimpin oleh tokoh muslim yg amanah dan terbaik karena sudah terbukti kinerja dan prestasi pak anies selama jadi gubernur dki jakarta. semoga indonesia ini lebih baik dan lebih ber matabat dan adil dari sebelum nya	Positif
Jadi ingat kampret dan cebong, gedek ma pendukung kampret yang mengkafirkan cebong, akhirnya kampret gigit jari, dan prabowo jadi menteri, mirip tebakan anak pesantren itu, jadi inget mbah amin mau jalan kaki jogja jakarta, sudahlah ... Mimpi nggak usah tinggi2 yang pasti2 saja. 😊😊😊	jadi ingat kampret dan cebong, gedek ma pendukung kampret yang mengkafirkan cebong, akhirnya kampret gigit jari, dan prabowo jadi menteri, mirip tebakan anak pesantren itu, jadi inget mbah amin mau jalan kaki jogja jakarta, sudahlah ... mimpi nggak usah tinggi2 yang pasti2 saja. 😊😊😊	Negatif
Orang klu ketemu pendukung si A pasti bilang saya orangnya si A. Klu ketemu pendukung si B pasti bilang saya orangnya si B Itu sdh biasa sejak dulu dong 😊😊. Ini orang kegedean pikiranya .	orang klu ketemu pendukung si a pasti bilang saya orangnya si a. klu ketemu pendukung si b pasti bilang saya orangnya si b itu sdh biasa sejak dulu dong 😊😊. ini orang kegedean pikiranya .	Netral

Tokenizing adalah proses pemecahan kalimat perkata/token [8]. Pada tahap ini setelah melakukan *lower case* pada tahap *case folding*, selanjutnya melakukan tahap *tokenizing* untuk melakukan pemotongan *string input* berdasarkan tiap kata yang menyusunnya agar bisa masuk ke tahapan *filtering* untuk dianalisis lebih lanjut. Hasil perubahan dapat dilihat pada Tabel 3.

Tabel 3. Hasil *tokenizing*

Sebelum	Sesudah	Kelas
sudah seharusnya indonesia di pimpin oleh tokoh muslim yg amanah dan terbaik karena sudah terbukti kinerja dan prestasi pak anies selama jadi gubernur dki jakarta. semoga indonesia ini lebih baik dan lebih ber matabat dan adil dari sebelum nya	['harus', 'nya', 'indonesia', 'pimpin', 'tokoh', 'muslim', 'yg', 'amanah', 'baik', 'bukti', 'kerja', 'prestasi', 'anies', 'gubernur', 'dki', 'jakarta', 'moga', 'indonesia', 'ber', 'matabat', 'adil', 'nya']	Positif
jadi ingat kampret dan cebong, gedek ma pendukung kampret yang mengkafirkan cebong, akhirnya kampret gigit jari, dan prabowo jadi menteri, mirip tebakan anak pesantren itu, jadi inget mbah amin mau jalan kaki jogja jakarta, sudahlah ... mimpi nggak usah tinggi2 yang pasti2 saja. 😊😊😊	['kampret', 'cebong', 'gedek', 'ma', 'dukung', 'kampret', 'kafir', 'cebong', 'kampret', 'gigit', 'jari', 'prabowo', 'menteri', 'tebak', 'anak', 'pesantren', 'inget', 'mbah', 'amin', 'jalan', 'kaki', 'jogja', 'jakarta', 'mimpi', 'nggak', ' ', ' ', '']	Negatif
orang klu ketemu pendukung si a pasti bilang saya orangnya si a. klu ketemu pendukung si b pasti bilang saya orangnya si b itu sdh biasa sejak dulu dong 😊😊. ini orang kegedean pikiranya .	['orang', 'klu', 'ketemu', 'dukung', 'si', 'a', 'bilang', 'orang', 'si', 'a', 'klu', 'ketemu', 'dukung', 'si', 'b', 'bilang', 'orang', 'si', 'b', 'sdh', ' ', ' ', 'orang', 'gede', 'pikiranya']	Netral

Filtering merupakan tahapan *preprocessing* untuk memfilter kata-kata dari hasil *term*. Pada penelitian ini menggunakan algoritma *Stopword*, yaitu kata-kata *non-deskriptif* yang dapat dibuang saat mendekati *bag-of-word*. *Stopword removal* secara otomatis menghilangkan kata-kata yang dianggap tidak penting dalam teks [9]. Beberapa kata akan dihapus menggunakan

library stopwords, seperti 'bisa', 'tegas', 'mampu', dan 'tidak' serta *stopword* lain yaitu “dan”, “di”, “dari” dan sebagainya. Hasil *filtering* dapat dilihat pada Tabel 4.

Tabel 4. Hasil *filtering*

Sebelum	Sesudah	Kelas
sudah seharusnya indonesia di pimpin oleh tokoh muslim yg amanah dan terbaik karena sudah terbukti kinerja dan prestasi pak anies selama jadi gubernur DKI Jakarta. semoga indonesia ini lebih baik dan lebih bermartabat dan adil dari sebelumnya	harus nya indonesia dipimpin tokoh muslim yg amanah baik bukti kerja prestasi Anies Gubernur DKI Jakarta semoga indonesia bermartabat adil nya	Positif
jadi ingat kampret dan cebong, gedek ma pendukung kampret yang mengkafirkan cebong, akhirnya kampret gigit jari, dan Prabowo jadi menteri, mirip tebakan anak pesantren itu, jadi inget mbah amin mau jalan kaki jogja Jakarta, sudahlah ... mimpi nggak usah tinggi2 yang pasti2 saja. 😂😂😂	kampret cebong gedek ma dukung kampret kafir cebong kampret gigit jari Prabowo menteri tebak anak pesantren inget mbah amin jalan kaki jogja Jakarta mimpi nggak	Negatif
orang klu ketemu pendukung si A pasti bilang saya orangnya si A. klu ketemu pendukung si B pasti bilang saya orangnya si B itu sdh biasa sejak dulu dong 😂😂. ini orang kegedean pikirannya .	orang klu ketemu dukung si A bilang orang si A klu ketemu dukung si B bilang orang si B sdh orang gede pikirannya	Netral

Selanjutnya, dilakukan proses *stemming* menggunakan *library Sastrawi* untuk merubah kata-kata ke dalam bentuk kata dasar yang tidak memiliki kata imbuhan. Misalnya, kata “dipimpin” menjadi “pimpin”. Hasil *stemming* dapat dilihat pada Tabel 5.

Tabel 5. Hasil *stemming*

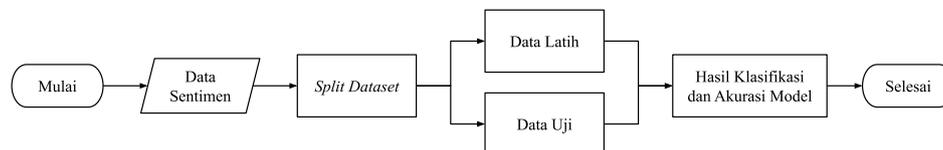
Sebelum	Sesudah	Kelas
sudah seharusnya indonesia di pimpin oleh tokoh muslim yg amanah dan terbaik karena sudah terbukti kinerja dan prestasi pak anies selama jadi gubernur DKI Jakarta. semoga indonesia ini lebih baik dan lebih bermartabat dan adil dari sebelumnya	harus nya indonesia dipimpin tokoh muslim yg amanah baik bukti kerja prestasi Anies Gubernur DKI Jakarta semoga indonesia bermartabat adil nya	Positif
jadi ingat kampret dan cebong, gedek ma pendukung kampret yang mengkafirkan cebong, akhirnya kampret gigit jari, dan Prabowo jadi menteri, mirip tebakan anak pesantren itu, jadi inget mbah amin mau jalan kaki jogja Jakarta, sudahlah ... mimpi nggak usah tinggi2 yang pasti2 saja. 😂😂😂	kampret cebong gedek ma dukung kampret kafir cebong kampret gigit jari Prabowo menteri tebak anak pesantren inget mbah amin jalan kaki jogja Jakarta mimpi nggak	Negatif
orang klu ketemu pendukung si A pasti bilang saya orangnya si A. klu ketemu pendukung si B pasti bilang saya orangnya si B itu sdh biasa sejak dulu dong 😂😂. ini orang kegedean pikirannya .	orang klu ketemu dukung si A bilang orang si A klu ketemu dukung si B bilang orang si B sdh orang gede pikirannya	Netral

Pembobotan Kata TF-IDF

Berdasarkan hasil *tokenizing* dalam tahap *pre-processing* yang telah dilakukan sebelumnya, ditemukan frekuensi kata yang sering muncul dalam dataset. Token per kata yang berasal dari tahap *pre-processing* tersebut kemudian diubah menjadi representasi vektor dan dihitung menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF merupakan suatu pendekatan yang digunakan untuk menilai sejauh mana hubungan antara kata dan dokumen dengan memberikan nilai bobot kepada setiap kata. Pendekatan TF-IDF menggunakan dua konsep, yakni seberapa sering suatu kata muncul dalam dokumen dan seberapa jarang kata tersebut muncul dalam dokumen [10-12]. Dalam penelitian ini, proses pembuatan vektor kata dan pemberian bobot kata dilakukan dengan *library TfidfVectorizer*. *TfidfVectorizer* adalah sebuah fungsi dalam *library text feature extraction* yang dikembangkan oleh *Scikit-learn*. Fungsinya adalah untuk melakukan proses TF-IDF pada teks.

Processing

Pada Gambar 4 adalah tahap klasifikasi model, pada penelitian ini dataset dibagi menjadi data pelatihan dan data pengujian dengan proporsi 90:10, 80:20, 70:30. Selain itu, untuk mendapatkan kombinasi terbaik dari data yang telah dibagi sebelumnya, diterapkan juga *k-Fold Cross Validation* dengan jumlah 10 pengujian. Tujuan dari proses pelatihan dan pengujian ini adalah membangun model klasifikasi dan menghitung tingkat kinerja dari kedua metode pada aspek akurasi saat memprediksi data uji.



Gambar 4. Flowchart klasifikasi

Uji Coba dan Evaluasi

Setelah melakukan pembagian data pada tahap *processing*, uji coba model klasifikasi dilakukan menggunakan metode *Confusion Matrix* dari data uji. Tahap evaluasi dilakukan dengan *k-Fold Cross Validation* untuk mengukur hasil dari model klasifikasi. Hasil dari klasifikasi ditampilkan ke dalam bentuk *confusion matrix* untuk membandingkan kelas *predicted* dan kelas *actual*. *Confusion matrix* untuk tiga kelas sentimen dapat dilihat pada Tabel 6.

Tabel 6. *Confusion Matrix 3x3*

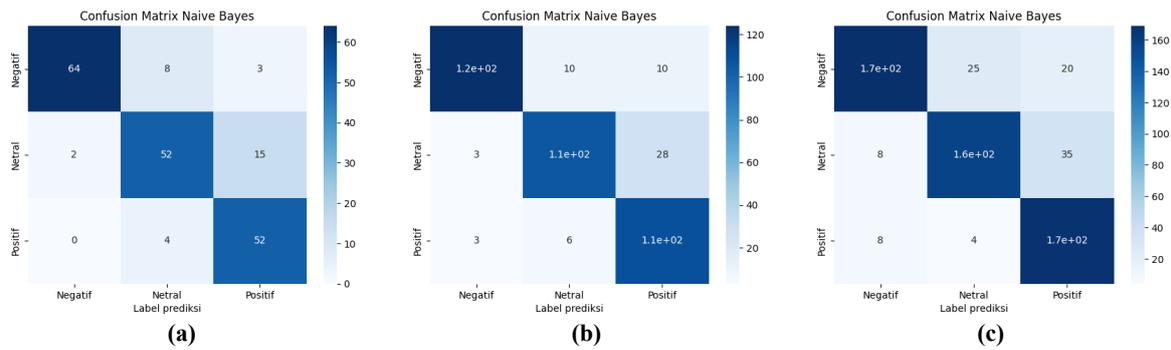
<i>Actual</i>	<i>Prediction</i>		
	Negatif	Netral	Negatif
Negatif	<i>True Negatif</i> (TN)	<i>False Negatif</i> (FN)	<i>False Negatif</i> (FN)
Netral	<i>False Netral</i> (FNR)	<i>True Netral</i> (TNR)	<i>False Netral</i> (FNR)
Positif	<i>False Positif</i> (FP)	<i>False Positif</i> (FP)	<i>True Positif</i> (TP)

Matriks *accuracy*, *precision*, *recall*, dan *f1-score* akan didapatkan berdasarkan Tabel 6 *confusion matrix* untuk mengukur kinerja model klasifikasi.

HASIL DAN PEMBAHASAN

Naïve Bayes Classifier

Gambar 5 menampilkan hasil pengujian *confusion matrix* yang terdapat kelas positif, negatif dan netral yang masih tersebar ke dalam kelas lain sehingga menghasilkan nilai akurasi sebesar 0.840 dengan pembagian data 90:10, akurasi 0.850 pada 80:20, dan akurasi 0.831 pada 70:30, dapat dilihat Gambar 5.



Gambar 5 Confusion Matrix model Naive Bayes Classification dengan pembagian data latih dan uji (a) 90 :10 (b) 80:20 dan (c) 70:30

Hasil matrik kinerja model dalam penelitian ini didapatkan menggunakan *library Multinomial Naive Bayes* dapat dilihat pada Tabel 7.

Tabel 7 Metode Naive Bayes dengan pembagian data

Split Data	Kelas	Precision	Recall	F1-Score
90:10	Negatif	0.97	0.85	0.91
	Netral	0.81	0.75	0.78
	Positif	0.74	0.93	0.83
	AVG	0.84	0.84	0.84
80:20	Negatif	0.95	0.86	0.91
	Netral	0.87	0.77	0.82
	Positif	0.74	0.92	0.82
	AVG	0.85	0.85	0.85
70:30	Negatif	0.91	0.79	0.85
	Netral	0.84	0.79	0.81
	Positif	0.75	0.93	0.83
	AVG	0.83	0.83	0.83

Tabel 7 menjelaskan hasil kinerja model dengan pembagian data 90:10, 80:20, dan 70:30, dapat melihat bahwa nilai *precision*, *recall*, dan *F1-Score* untuk setiap kelas mengalami variasi. Variasi ini memberikan gambaran tentang sejauh mana model mampu berkinerja dalam situasi yang berbeda. Selain itu, nilai support juga memberikan informasi tentang seberapa kuat representasi setiap kelas dalam data. Hasil ini memberikan isyarat bahwa model dapat menangani distribusi kelas dengan baik pada kondisi tertentu.

Selanjutnya, evaluasi model NBC dilakukan melalui *k-Fold Cross Validation*. Data dibagi menjadi beberapa bagian dengan rasio 90:10, 80:20, dan 70:30 dengan melakukan pelatihan pada data dan pengujian terhadap data, dilakukan eksperimen dengan pengujian 10 *k-Fold Cross Validation*, dan hasilnya dapat dilihat pada Tabel 8.

Tabel 8 Rasio 10 k-Fold Naive Bayes Classifier

k-Fold	NBC 90:10	NBC 80:20	NBC 70:30
1	0.868	0.860	0.906
2	0.868	0.892	0.877
3	0.880	0.887	0.862
4	0.909	0.892	0.886
5	0.897	0.910	0.901
6	0.884	0.860	0.876
7	0.909	0.896	0.876
8	0.880	0.887	0.881
9	0.900	0.900	0.931
10	0.925	0.864	0.876
AVG	0.892	0.885	0.887

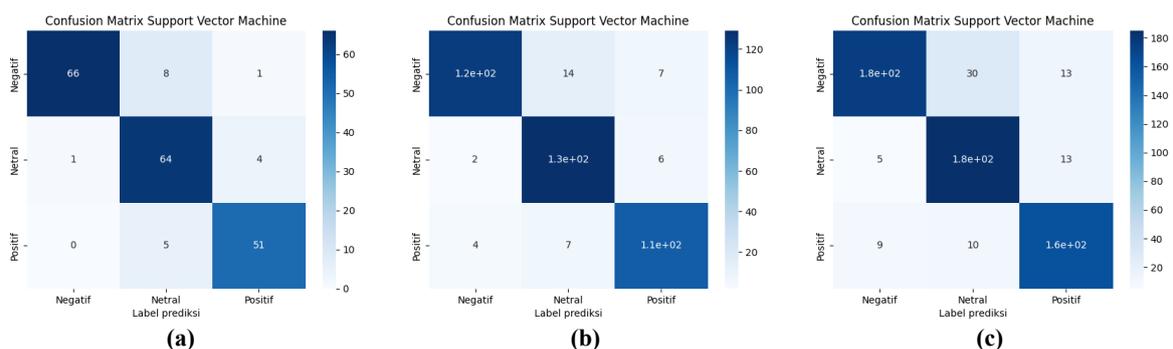
Tabel 10 Rasio 10 k-Fold Support Vector Machine

k-Fold	SVM 90:10	SVM 80:20	SVM 70:30
1	0.897	0.883	0.906
2	0.888	0.923	0.886
3	0.921	0.896	0.926
4	0.905	0.892	0.886
5	0.909	0.923	0.940
6	0.909	0.901	0.925
7	0.921	0.909	0.896
8	0.921	0.905	0.905
9	0.938	0.923	0.930
10	0.925	0.882	0.900
AVG	0.913	0.904	0.910

Pada Tabel 8 menunjukkan hasil analisis yang terdapat beberapa rasio yang berbeda dalam melakukan pengujian 10 *k-Fold* untuk mendapatkan nilai rata-rata menjelaskan bahwa pada NBC 90:10 mendapatkan akurasi rata-rata sebesar 0.892 dengan nilai tertinggi 0,925 dan terendah 0,868, sedangkan NBC 80:20 akurasi rata-rata sebesar 0.885 dengan nilai tertinggi 0,910 dan terendah 0,860, dan NBC 70:30 akurasi rata-rata sebesar 0.887 dengan nilai tertinggi 0,931 dan terendah 0,876.

Hasil *Support Vector Machine*

Gambar 6 menampilkan hasil pengujian *confusion matrix* yang terdapat kelas positif, negatif dan netral yang masih tersebar ke dalam kelas lain sehingga menghasilkan nilai akurasi sebesar 0.905 dengan pembagian data 90:10, akurasi 0.900 pada 80:20, dan akurasi 0.867 pada 70:30.



Gambar 6 *Confusion Matrix* model *Support Vector Machine* dengan pembagian data latihan dan uji (a) 90 :10 (b) 80:20 dan (c) 70:30

Pada penelitian ini menggunakan sebuah fitur yang dikenal sebagai *kernel function* atau trik *kernel*. Konsep dari *kernel* ini adalah memanfaatkan ruang berdimensi lebih tinggi untuk membentuk *hyperplane*. Jenis fungsi *kernel* yang diterapkan dalam penelitian ini menggunakan *kernel Radial Basis Function* (RBF). Dalam *kernel* ini, terdapat *hyperparameter* yang memegang peran penting dalam menentukan kinerja klasifikasi yang dihasilkan oleh SVM. Pada RBF *kernel* membutuhkan dua parameter *kernel* yaitu *C* dan *γ*. Parameter *C* dan *γ* nilainya ditentukan sebelum pelatihan model. Parameter *C* pada SVM mengontrol *trade-off* antara memaksimalkan margin dan mengurangi kesalahan klasifikasi. Jika *C* terlalu besar, model mengalami *overfit* pada data pelatihan. Jika *C* terlalu kecil, model mengalami *underfit* dan gagal menangkap pola. Parameter *γ* atau *gamma* mengontrol bentuk fungsi *kernel* RBF. *Gamma* yang tinggi dapat menyebabkan *overfitting*, terutama jika dataset tidak terlalu besar. Nilai yang rendah dari *gamma* dapat menyebabkan *underfitting*. Pada penelitian ini menggunakan parameter *C* dengan skala *array* [0.1, 1, 10] dan parameter *gamma* dengan skala *array* [0.01, 0.1, 1] pada data uji dan latihan disesuaikan pada *array* ke-1 dan ke-2 dengan proporsi 90:10, 80:20, dan 70:30. Kinerja matrik model yang dihasilkan menggunakan data uji dapat dilihat pada Tabel 9.

Tabel 9 Metode *Support Vector Machine* dengan pembagian data

<i>Split Data</i>	Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
90:10	Negatif	0.99	0.88	0.93
	Netral	0.83	0.93	0.88
	Positif	0.91	0.91	0.91
80:20	AVG	0.91	0.91	0.91
	Negatif	0.95	0.85	0.90

	Netral	0.86	0.94	0.90
	Positif	0.89	0.91	0.90
AVG		0.90	0.90	0.90
	Negatif	0.93	0.80	0.86
70:30	Netral	0.82	0.91	0.86
	Positif	0.86	0.90	0.88
AVG		0.87	0.87	0.87

Selanjutnya, evaluasi model SVM dilakukan melalui *k-Fold Cross Validation*. Data dibagi menjadi beberapa bagian dengan rasio 90:10, 80:20, dan 70:30 dengan melakukan pelatihan pada data dan pengujian terhadap data, dilakukan eksperimen dengan pengujian 10 *k-Fold Cross Validation*, dan hasilnya dapat dilihat pada Tabel 10.

Pada Tabel 10 menunjukkan hasil analisis yang terdapat beberapa rasio yang berbeda dalam melakukan pengujian 10 *k-Fold* untuk mendapatkan nilai rata-rata menjelaskan bahwa pada SVM 90:10 mendapatkan akurasi rata-rata sebesar 0.913 dengan nilai tertinggi 0,938 dan terendah 0,888, sedangkan SVM 80:20 akurasi rata-rata sebesar 0.904 dengan nilai tertinggi 0,923 dan terendah 0,882, dan SVM 70:30 akurasi rata-rata sebesar 0.910 dengan nilai tertinggi 0,940 dan terendah 0,886.

Hasil Perbandingan

Hasil perbandingan dari kedua metode dapat dilihat pada Tabel 11.

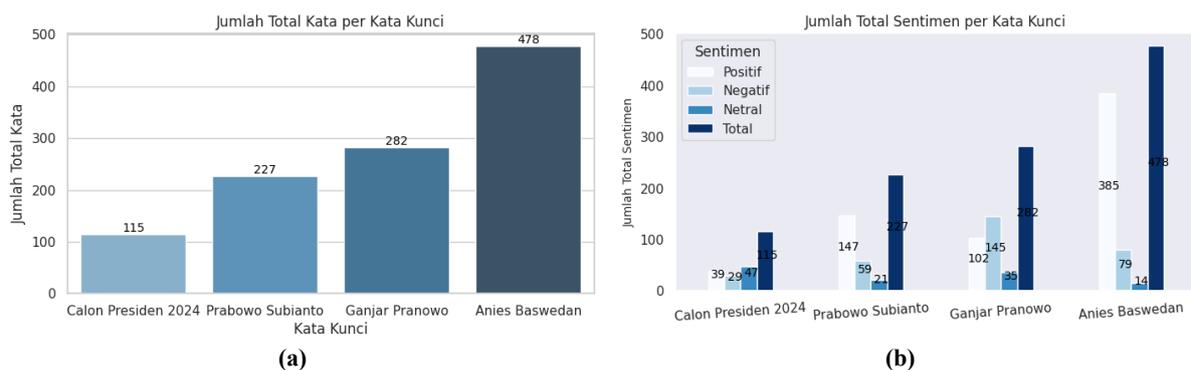
Tabel 11 Hasil perbandingan

No	Metode	Akurasi 90:10	Akurasi 80:20	Akurasi 70:30
1	<i>Naïve Bayes Classifier</i>	0.840	0.850	0.831
2	<i>Support Vector Machine</i>	0.905	0.900	0.867

Pada Tabel 11 menunjukkan hasil perbandingan dalam penelitian ini menggunakan rasio 90:10, 80:20, dan 70:30, bahwa metode *Support Vector Machine* memiliki tingkat akurasi yang baik dengan nilai sebesar 0.905, 0.900, dan 0.867 dibandingkan dengan metode *Naïve Bayes Classifier* yang menghasilkan tingkat akurasi sebesar 0.840, 0.850, dan 0.831.

Hasil Analisis Sentimen

Setelah melakukan perbandingan dari kedua metode tersebut, langkah berikutnya adalah melakukan analisis sentimen masyarakat terhadap pemilihan presiden tahun 2024 pada data komentar youtube yang telah dibersihkan. Proses ini dilakukan menggunakan empat kata kunci yang relevan, yaitu Calon Presiden 2024, Prabowo Subianto, Ganjar Pranowo, dan Anies Baswedan, menghasilkan 1110 data kata kunci dengan rincian Gambar 7.



Gambar 7 Perhitungan tiap kata kunci (a) Total kata per kata kunci, (b) Total sentimen per kata kunci

Pada Gambar 7 menjelaskan hasil berupa data kata kunci yang telah diklasifikasikan ke dalam kelas yaitu positif, negatif, dan netral. Selanjutnya, dilakukan perhitungan persentase sentimen dari keseluruhan data yang telah berhasil dikumpulkan dengan cara melakukan iterasi pada kamus sentimen dan menghitung total sentimen untuk setiap kelas. Dalam setiap kelas, total sentimen dihitung dengan menjumlahkan jumlah sentimen positif, negatif, dan netral. Langkah berikutnya ialah melakukan perhitungan persentase untuk masing-masing sentimen terhadap total sentiment menggunakan Persamaan 1 yang dapat dilihat pada Tabel 12 dan Gambar 8.

$$\text{Persentase} = \left(\frac{\text{Jumlah Setiap Jenis Sentimen}}{\text{Total Sentimen}} \right) \times 100 \quad (1)$$

Kata Kunci : Calon Presiden 2024

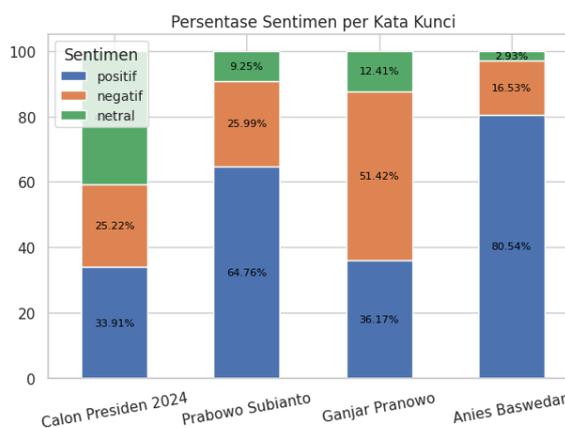
$$\text{Persentase}_{\text{positif}} = \frac{39}{115} \times 100\% = 33.91\%$$

$$\text{Persentase}_{\text{negatif}} = \frac{29}{115} \times 100\% = 25.22\%$$

$$\text{Persentase}_{\text{netral}} = \frac{47}{115} \times 100\% = 40.87\%$$

Tabel 12 Persentase sentimen

Kata Kunci	Positif	Negatif	Netral
Calon Presiden 2024	33.91%	25.22%	40.87%
Prabowo Subianto	64.76%	25.99%	9.25%
Ganjar Pranowo	36.17%	51.42%	12.41%
Anies Baswedan	80.54%	16.53%	2.93%



Gambar 8 Persentase Sentimen per Kata Kunci

Pada Tabel 12 dan Gambar 8 menjelaskan hasil persentase analisis sentimen masyarakat terhadap pemilihan presiden 2024, menggunakan empat kata kunci yang relevan, menunjukkan persepsi masyarakat yang positif, negatif, serta netral terhadap masing-masing calon presiden.

KESIMPULAN

Dari hasil penelitian ini melakukan perbandingan antara metode *Naïve Bayes* (NBC) dan *Support Vector Machine* (SVM) dalam menganalisis sentimen masyarakat terhadap Pemilihan Presiden 2024 di komentar youtube, menjelaskan bahwa metode SVM mendapatkan hasil akurasi terbaik. Pada pembagian data dengan rasio 90:10, SVM mencapai nilai akurasi sebesar 0.905, rasio 80:20 dengan akurasi 0.900, dan rasio 70:30 dengan akurasi 0.867. Sementara itu, NBC memperoleh akurasi 0.840 pada rasio 90:10, 0.850 pada rasio 80:20, dan 0.831 pada rasio 70:30. Selain itu, hasil analisis sentimen terhadap pemilihan presiden 2024 menggunakan empat kata kunci. Secara keseluruhan, sentimen positif mendominasi terhadap Anies Baswedan (80.54%), Prabowo Subianto (64.76%), Ganjar Pranowo (36.17%), dan Calon Presiden secara umum (33.91%). Di sisi lain, sentimen negatif lebih tinggi untuk Ganjar Pranowo (51.42%) dan Prabowo Subianto (25.99%), sementara Anies Baswedan dan Calon Presiden secara umum memiliki tingkat sentimen negatif yang lebih rendah (16.53% dan 25.22% secara berturut-

turut). Sentimen netral tertinggi tercatat pada Calon Presiden secara umum (40.87%), diikuti oleh Ganjar Pranowo (12.41%), diikuti Prabowo Subianto (9.25%), dan Anies Baswedan (2.93%). Hasil kesimpulan ini memberikan gambaran secara menyeluruh tentang pola sentimen masyarakat terhadap masing-masing kandidat dan pemilihan presiden secara umum.

DAFTAR PUSTAKA

- [1] T. T. Widowati and M. Sadikin, "ANALISIS SENTIMEN TWITTER TERHADAP TOKOH PUBLIK DENGAN ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR MACHINE," *Jurnal SIMETRIS*, vol. 11, no. 2, 2020, [Online]. Available: <https://t.co/Xzf91zHK41>
- [2] Yuyun, Nurul Hidayah, and Supriadi Sahibu, "Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 820–826, Aug. 2021, doi: 10.29207/resti.v5i4.3146.
- [3] I. P. Dedy, W. Darmawan, G. Aditra Pradnyana, I. Bagus, and N. Pascima, "Optimasi Parameter Support Vector Machine Dengan Algoritma Genetika Untuk Analisis Sentimen Pada Media Sosial Instagram," *SCIENCE AND INFORMATION TECHNOLOGY*, vol. 6, no. 1, pp. 58–67, 2023, [Online]. Available: <https://doi.org/10.31598>
- [4] S. Nurul, N. Safriadi, and E. Esyudha, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," *Jurnal Edukasi dan Penelitian Informatika*, 2019, [Online]. Available: <http://dev.twitter.com>.
- [5] Hermanto, A. Mustopa, and A. Y. Kuntoro, "ALGORITMA KLASIFIKASI NAIVE BAYES DAN SUPPORT VECTOR MACHINE DALAM LAYANAN KOMPLAIN MAHASISWA," *JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER*, vol. 5, no. 2, 2020, [Online]. Available: www.bsi.ac.id
- [6] R. Rakhmat Sani, Y. Ayu Pratiwi, S. Winarno, E. Devi Udayanti, and dan Farrikh Al Zami, "Analisis Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Hoax pada Berita Online Indonesia," *Jurnal Masyarakat Informatika*, vol. 13, no. 2, pp. 2777–0648, 2022.
- [7] S. U. Masruroh and I. Subchi, *Fast Text Dan Word2vec Pada Query Kesamaan Semantik Sistem Temu Kembali Informasi*. Deepublish, 2023. [Online]. Available: <https://books.google.co.id/books?id=sDjCEAAQBAJ>
- [8] S. Shevira, I. Made, A. D. Suarjaya, and P. Wira Buana, "Pengaruh Kombinasi dan Urutan Pre-Processing pada Tweets Bahasa Indonesia," *JITTER-Jurnal Ilmiah Teknologi dan Komputer*, vol. 3, no. 2, 2022.
- [9] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, p. 406, Apr. 2021, doi: 10.30865/mib.v5i2.2835.
- [10] M. P. Simatupang and D. P. Utomo, "ANALISA TESTIMONIAL DENGAN MENGGUNAKAN ALGORITMA TEXT MINING DAN TERM FREQUENCY- INVERSE DOCUMENT FREQUENCE (TF-IDF) PADA TOKO ALLMEEART," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, Dec. 2019, doi: 10.30865/komik.v3i1.1697.
- [11] Darmawan, B., Laksito, A. D., Yudianto, M. R. A., & Sidauruk, A. (2023). Analisis Perbandingan Ekstraksi Fitur Teks pada Sentimen Analisis Kenaikan Harga BBM. *Krea-TIF: Jurnal Teknik Informatika*, 11(1), 53–63. <https://doi.org/10.32832/krea-tif.v11i1.13819>
- [12] Habyba, A. N., Djatna, T., & Anggraeni, E. (2021). Positioning E-commerce Produk UKM berdasarkan Kebutuhan Afektif Pengguna. *Krea-TIF: Jurnal Teknik Informatika*, 9(1), 21–28. <https://doi.org/10.32832/kreatif.v9i1.3590>