

Analisis Sentimen Popularitas Capres dan Pilpres pada Media Sosial Twitter: Perbandingan Algoritma SVM, KNN, dan Naïve Bayes

Sentiment Analysis of Presidential Candidates and Presidential Elections Popularity on Twitter Social Media: Comparison of SVM, KNN, and Naïve Bayes Algorithms

Rojakul¹, Sumardianto², Arief Wibowo³

^{1,2,3}Program Studi Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta Selatan

E-mail: ¹2211600511@student.budiluhur.ac.id, ²221160222@student.budiluhur.ac.id, ³arief.wibowo@budiluhur.ac.id

Abstrak

Popularitas tokoh publik semakin tumbuh pesat di era media sosial, khususnya para pemangku kepentingan yang sering menggunakan analisis sentimen untuk mengukur tingkat opini mereka tentang suatu kegiatan atau penilaian atas tokoh publik tersebut. Dalam konteks ini, analisis sentimen menjadi alat yang kuat untuk memahami bagaimana tokoh publik dipersepsikan dan direspon oleh masyarakat. Penelitian ini bertujuan untuk mempertahankan ruang diskusi yang terbuka dan inklusif tentang opini masyarakat terkait dengan politik, serta memberikan kesempatan masyarakat untuk terlibat dalam diskusi serta meningkatkan partisipasi politik masyarakat. Beberapa tujuan dari metode yang diterapkan dalam penelitian ini adalah untuk membentuk klasifikasi *tweet* emosional kedalam kategori positif dan negatif, serta menggunakan algoritma pembagian terstruktur yaitu *Support Vector Machines* (SVM), *Naive Bayes* (NB) dan *K-Nearest Neighbor*. Pengujian yang telah dilakukan memberikan informasi bahwa algoritma NB mengungguli SVM dan K-NN dalam menyelesaikan kasus kepercayaan ini dengan tingkat akurasi 96,39%. Presisi mencapai 100% sementara tingkat *recall* mencapai 92,78%.

Kata kunci: Analisis Sentimen, SVM, Naive Bayes, K-NN, Capres, Pilpres

Abstract

The popularity of public figures is rapidly growing in the era of social media, particularly among stakeholders who often utilize sentiment analysis to gauge public opinion on certain activities or assessments of these public figures. In this context, sentiment analysis serves as a powerful tool to comprehend how public figures are perceived and responded to by the community. This research aims to address these challenges and provide valuable insights into how the public responds to the current discussions on presidential and vice-presidential candidates on social media, particularly on Twitter, and its impact on overall public opinion. Several objectives of the methods applied in this study are to categorize emotional tweets into positive and negative categories, using structured classification algorithms such as Support Vector Machines (SVM), Naive Bayes (NB), and K-Nearest Neighbor (KNN). The conducted tests provide information that the NB algorithm outperforms SVM and K-NN in resolving this trust issue with an accuracy rate of 96.36%. Precision reaches 100%, while the recall rate reaches 92.78%.

Keywords: Sentiment Analysis, SVM, Naive Bayes, K-NN, Presidential Candidates, Presidential Election

1. PENDAHULUAN

Dalam era digital, opini publik dalam media sosial, khususnya Twitter, telah menjadi kekuatan yang kuat dalam proses pemilihan presiden, karena Twitter merupakan platform jejaring sosial yang cukup populer dikalangan masyarakat, dimana terdapat 500 juta kicauan atau *tweet* per hari dari pengguna ataupun dari siaran pers Twitter. Keterlibatan masyarakat untuk ikut dalam partisipasi politik online juga mengalami pertumbuhan pesat, dan perubahan sosial ini mengakibatkan meningkatnya pandangan politik serta perdebatan pada platform tersebut. Opini-opini yang terbentuk di Twitter dapat mempengaruhi keputusan pemilih, dan potensi dampaknya terhadap proses demokrasi dan stabilitas politik sangatlah signifikan. Walaupun pesta demokrasi sudah selesai dilaksanakan, penelitian ini bertujuan untuk mempertahankan ruang diskusi yang terbuka dan inklusif tentang opini masyarakat terkait dengan politik, serta memberikan kesempatan masyarakat untuk terlibat dalam diskusi serta meningkatkan partisipasi politik masyarakat. Dengan adanya fitur untuk memberikan pendapat dan komentar, Twitter menyediakan sumber berita yang dapat dimanfaatkan untuk menganalisis opini serta pandangan masyarakat baik dari sisi individu maupun institusi. Dengan melihat opini yang terdapat dalam *tweet*, kita dapat memperoleh pemahaman tentang bagaimana sentimen terhadap suatu opini berkembang. Salah satunya yang memperbincangkan Capres dan Pilpres [1] khususnya dalam bidang politik.

Banyak orang mengungkapkan pendapatnya di jejaring sosial, terutama di Twitter [2]. Ini meningkatkan kemampuan analisa kami untuk memahami setiap pandangan politik komunitas dari pengguna jejaring sosial. Banyaknya pemilik akun jejaring sosial memberikan kesempatan kepada peneliti untuk mempersiapkan dan menganalisis posisi Capres dan Pilpres sebagai data opini. Analisis pendapat terhadap pemrosesan status atau kicauan pengguna Twitter menunjukkan pendapat umum tentang Capres dan Pilpres.

Analisis sentimen berguna untuk memperluas solusi untuk menganalisis, mengidentifikasi, dan merepresentasikan tampilan pada teks. Ini juga merupakan metode identifikasi opini, atau pendapat yang berasal dari isi materi atau *tweet* berupa teks yang berkaitan dengan informasi atau peristiwa yang mempunyai kecenderungan bersifat opini positif serta negatif [3][4]. Definisi Sentimen adalah penilaian atau sudut pandang yang didasarkan pada emosi yang sangat kuat terhadap suatu hal, sesuai dengan definisi yang tercantum dalam Kamus Besar Bahasa Indonesia (KBBI) [8]. Melakukan pengklasifikasian untuk membentuk analisa opini positif atau negatif secara manual memang memungkinkan, tetapi semakin bertambahnya jumlah sumber opini, tentunya membutuhkan energi yang cukup besar dalam proses klasifikasi. Dengan demikian, penggunaan metode pembelajaran mesin diajukan untuk membentuk klasifikasi polaritas opini dari beragam sumber data dengan menggunakan teknik penambangan teks.

Melalui blog, forum, media umum, dan situs ulasan, informasi dalam bentuk teks kini tersebar secara luas di internet [5], tidak hanya dalam ranah politik, bahkan analisis sentimen juga digunakan untuk menganalisa pendapat pelanggan mengenai keluhan terhadap produk makanan [6], dan juga dalam bidang kesehatan dilakukan untuk menganalisis sentimen masyarakat terhadap vaksin [7].

Penelitian klasifikasi juga pernah dilakukan sebelumnya dengan menganalisis opini calon presiden Indonesia tahun 2024 dengan metode algoritma *Naive Bayes* (NB). Penelitian tersebut bisa mengevaluasi dengan tingkat akurasi sebesar 73,68%. Pada contoh penelitian ini, dapat disimpulkan bahwa prosedur pemecahan dengan algoritman NB masih menunjukkan tingkat akurasi yang agak rendah [9].

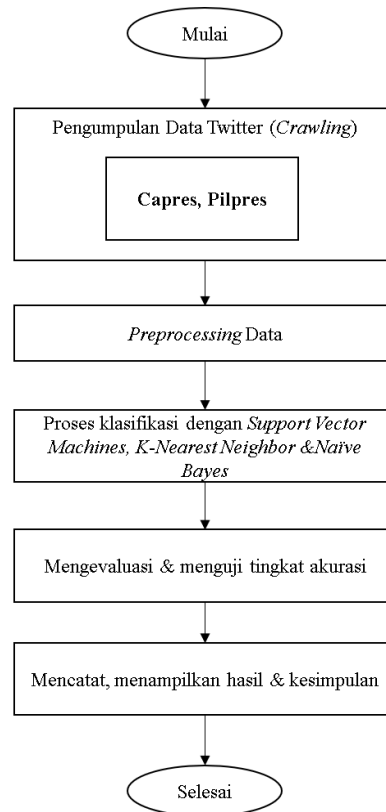
Studi sebelumnya yang berhubungan dengan analisis sentimen, menyimpulkan juga bahwa menggunakan algoritma *Naive Bayes* tanpa menggunakan fitur tambahan memiliki tingkat akurasi sebesar 96,44% dalam membentuk klasifikasi sentimen. Namun, tingkat akurasi dapat dapat meningkat menjadi 98% dengan penambahan fitur tambahan seperti pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) dan transformasi simbol sentimen [10].

Pada studi sebelumnya yang dilakukan oleh Didin Muhidin, performa algoritma *Support Vector Machine* (SVM) serta *K-Nearest Neighbor* (K-NN) dalam menganalisis opini tentang *New Normal* juga dibandingkan. Penelitian tersebut melibatkan analisis opini dengan menggunakan algoritma klasifikasi K-NN serta SVM. Hasil pengujiannya menunjukkan bahwa algoritma K-NN memiliki tingkat akurasi yang lebih tinggi, mencapai 72,96% [11].

Dalam mengklasifikasikan data berbasis teks, tiga (3) algoritma klasifikasi seperti *Support Vector Machine*, *Naive Bayes*, dan *K-Nearest Neighbor* memiliki kelebihan dan kekurangan tersendiri. Pada penelitian ini, ketiga algoritma tersebut diterapkan untuk menganalisis popularitas Capres dan Pilpres di Twitter. Perbandingan tersebut dilakukan untuk menentukan algoritma yang memiliki akurasi paling tinggi dalam analisis sentimen.

2. METODE PENELITIAN

Seperti ditunjukkan pada Gambar 1, penelitian ini dimulai dengan tahapan pengumpulan data dan dilakukan dengan cara mengindeks data pada platform Twitter, setelah itu dilanjutkan dengan melakukan tahap *preprocessing* data. Langkah langkah tersebut terdiri dari *data cleansing*, *case folding*, *tokenization* dan *stopword filter*. Kemudian hasil klasifikasi dan klasifikasi teks ditampilkan.



Gambar 1 Tahapan penelitian.

2.1 Metode pengumpulan data

Dalam penelitian ini, data opini dikumpulkan dari *tweet* dan komentar *retweet* yang diposting oleh pengguna platform Twitter. Informasi *tweet* diperoleh dengan melakukan indeksasi di Twitter, dengan menggunakan kata kunci Capres, Pilpres. Pengumpulan data tweet dilakukan selama rentang waktu antara tanggal 1 Juni 2023 sampai dengan 30 September 2023 dengan jumlah 5000 *tweet*

2.2 Tahap pra-pemrosesan teks atau preprocessing

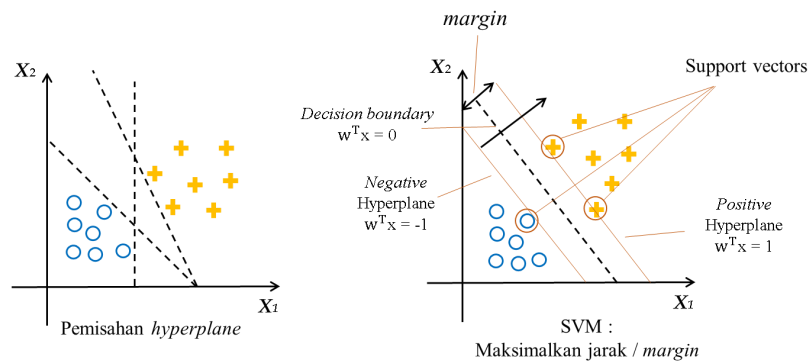
Langkah pertama dalam tahapan *preprocessing* adalah melakukan pembersihan data tekstual dengan menghilangkan informasi yang tidak diperlukan. Membersihkan fungsi tanpa karakter yang tidak perlu seperti menghilangkan karakter URL, *hashtag* (#), angka, *username* (@), *emoticon*, tanda baca (*punctuation*) dan simbol. Selain itu, juga dilakukan *case folding* untuk mengubah susunan kata suatu karakter sehingga menjadi seragam.

Langkah berikutnya adalah melakukan tokenisasi, yaitu memecah teks menjadi komponen-komponen terpisah. Tokenisasi merupakan tahap awal dalam pra-pemrosesan teks, di mana teks dipecah menjadi unit-unit seperti kalimat dan kata. *Transform cases* adalah proses mengubah gaya penulisan atau pengejaan suatu teks dari satu format ke format yang lain. Dalam pra-pemrosesan teks, *Filter StopWord* kita gunakan juga untuk mengeliminasi kata-kata yang tidak relevan dalam bahasa Indonesia. Selanjutnya, terdapat juga *Filter Token by length* yang berguna untuk menyaring atau menghapus token berdasarkan panjang karakter token dalam teks

2.3 Pengelompokan sentimen

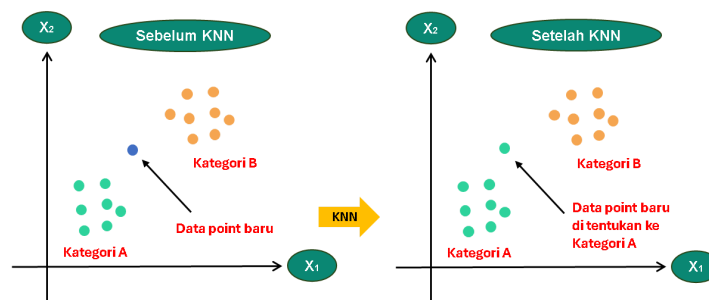
Dalam penelitian ini, proses pemodelan terstruktur menggunakan aplikasi Rapidminer. Metode pemodelan terstruktur yang digunakan melibatkan metode *Support Vector Machine* (SVM), *K-Nearest Neighbor* (KNN), dan *Naive Bayes* (NB).

Support Vector Machine (SVM) merupakan metode pembelajaran terbimbing yang umum digunakan untuk membentuk klasifikasi serta analisis regresi dalam analisis data dan identifikasi pola [12]. Representasi dari *hyperplane* SVM ditunjukkan pada Gambar 2 :



Gambar 2 menunjukkan representasi hyperplane pada SVM.

Proses selanjutnya adalah pemodelan menggunakan Algoritma K-NN, sebuah algoritma pembelajaran terbimbing yang digunakan untuk tugas klasifikasi dan regresi. Algoritma ini bekerja dengan mencari titik data K yang memiliki jarak terdekat dengan sampel yang diberikan, lalu menggunakan label kelas atau nilai dari titik data tersebut untuk menghasilkan prediksi tentang label kelas atau nilai sampel [13]. Hasil dari model algoritma K-NN ditunjukkan pada Gambar 3 :



Gambar 3 menunjukkan representasi algoritma KNN.

Algoritma *Naive Bayes* dipilih sebagai langkah berikutnya, yang merupakan metode klasifikasi yang didasarkan pada teorema probabilitas *Bayes* dengan asumsi yang sederhana dan naif yang dikenal sebagai "*naive*". Metode ini banyak digunakan dalam pengenalan pola dan analisis teks untuk memprediksi klasifikasi atau label dari suatu data berdasarkan fitur-fitur yang ada. Bentuk umum *teorema Bayes* adalah sebagai berikut:

$$P(H|X) = (1)$$

Informasi:

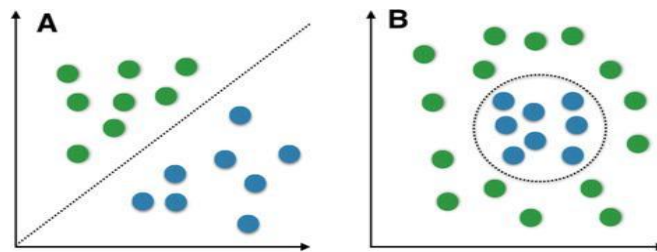
X = Data kelas tidak diketahui

H = Data hipotesis X adalah kategori khusus

$P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi X (probabilitas posterior)

$P(H)$ = Probabilitas hipotesis H (probabilitas awal).

Representasi hasil dari algoritma *Naive Bayes* dapat ditunjukkan pada Gambar 4 :



Gambar 4 menunjukkan representasi algoritma *Naive Bayes*.

2.4 Mengevaluasi dan menguji tingkat akurasi

Evaluasi klasifikasi dapat dilakukan dengan memanfaatkan data uji yang memiliki nilai tertentu dan tidak termasuk dalam data pelatihan. Pemodelan jenis klasifikasi bertujuan untuk menghasilkan deskripsi tentang urutan data, yang nilai akhirnya memberikan prediksi mengenai objek atau kategori data. Klasifikasi yang menghasilkan dua kelas disebut sebagai klasifikasi biner. Kedua kategori dijelaskan di bawah {positif, negatif}. Di bawah ini adalah empat kemungkinan proses klasifikasi untuk urutan data positif yang sebenarnya. Negatif Sejati; alarm palsu; Negatif Palsu.

Dalam evaluasi model klasifikasi, ukuran yang digunakan adalah tingkat keakuratan (*Accuracy*). Tingkat keakuratan mengindikasikan jumlah prediksi yang akurat. Skor ini berguna dalam mengukur tingkat presisi melalui penggunaan validasi silang K-fold. Matriks kebingungan digunakan dalam proses penilaian. Proses skoring ini digunakan untuk melihat kinerja model klasifikasi yang sedang dipertimbangkan dan untuk menentukan akurasinya.

Di bawah ini adalah rumus untuk setiap proses kategorisasi emosi. *Accuracy* adalah sebagian kecil dari prediksi sebenarnya untuk semua data. Perhitungan akurasi ditunjukkan dalam Formula 1. Akurasi adalah ukuran keakuratan hasil model. Persamaan tersebut merupakan perbandingan data positif dan total yang benar dengan label positif, perhitungan yang tepat ditunjukkan di Formula 2. *Recall*, yang juga dikenal sebagai ukuran kelengkapan model, dihitung dengan membandingkan jumlah sampel yang terdeteksi benar positif dengan total jumlah sampel positif yang seharusnya terdeteksi. Rumus untuk menghitung *recall* dapat ditemukan pada rumus 3 [14].

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

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

Dalam penjelasan model klasifikasi di atas, *True Positive* (TP) menunjukkan jumlah data yang selesai diprediksi dengan benar sebagai data positif; sebaliknya, *False Positive* (FP) menunjukkan jumlah data yang sebenarnya adalah data negatif, tetapi diprediksi sebagai data positif secara salah, yang menyebabkan kesalahan.

2.5 Perbandingan hasil pengujian

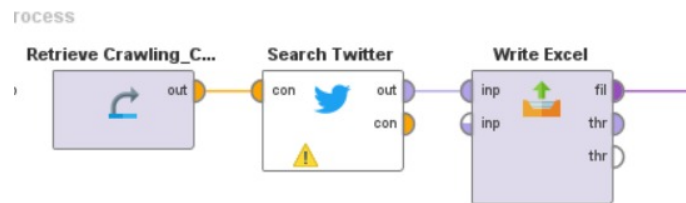
Setelah tahap evaluasi selesai, langkah selanjutnya adalah membandingkan tingkat akurasi antara metode SVM, K-NN, dan NB untuk mengetahui metode mana yang memiliki tingkat akurasi terbaik.

3. HASIL DAN PEMBAHASAN

Pada fase pengumpulan data, Rapidminer digunakan untuk men-*crawling* data Twitter dengan kata kunci Capres dan Pilpres. Pengumpulan data ini mencakup 5.000 *tweet* dari 1 Juni 2023 hingga 30 September 2023.

Tabel 1 Menunjukkan kata kunci dan jumlah data *tweet*.

Dataset	Jumlah Data
Capres, Pilpres	5000



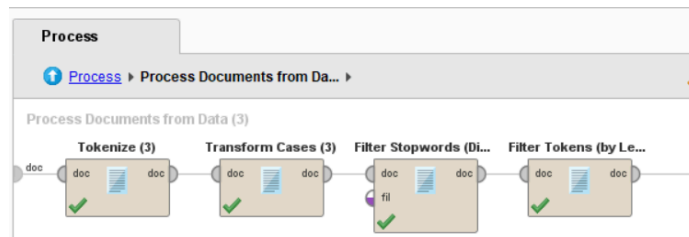
Gambar 5 Proses *Crawling* Data

Tabel 2 Data Hasil Proses *Crawling*

Output Data <i>Crawling</i>	
No	<i>Tweet</i>
1	RT @Ghurem2: Bergabungnya Partai Perindo menjadi salah satu pendukung Capres @ganjarpranowo akan semakin menambah kekuatan untuk menghadapi...
2	Kecerdasan dan kemampuan Airlangga Hartarto @airlangga_hrt dalam mengantisipasi situasi politik yang dinamis menjelang pemilu dan pilpres 2024. Airlangga akan menjadi penentu ritme dalam membangun koalisi partai politik yang akan mengusung capres cawapres nantinya. Terbaik! https://t.co/DeuE6LI3QV
3	In my prediction, ini adalah lawan sepadan untuk Pak Prabowo. Gerindra sebaiknya merangkul Demokrat agar bisa menang Pilpres. PKS mungkin membuat koalisi baru dengan partai kecil untuk mengusung Anies sebagai Capres ðŸ‘ðŸ¼

3.1 Preprocessing Data

Langkah selanjutnya. Pada titik ini, dilakukan beberapa prosedur, termasuk *cleansing*, tokenisasi, *case folding*, dan *stop word filter*. Gambar 6 berikut menunjukkan penjelasan tentang proses:



Gambar 6: Persiapan data

Pada tahap ini, data dibersihkan sesuai dengan prosedur yang telah dijelaskan sebelumnya dalam metodologi penelitian. Hasil pembersihan dari tahap ini ditampilkan dalam Tabel 3, yang menjelaskan langkah-langkah pembersihan dari awal hingga koreksi ejaan.

Tabel 3 Menunjukkan Data Setelah Pembersihan

Pembersihan data		
No	<i>Tweet</i>	<i>Cleansing</i>
1	RT @Ghurem2: Bergabungnya Partai Perindo menjadi salah satu pendukung Capres @ganjarpranowo akan semakin menambah kekuatan untuk menghadapi...	Bergabungnya Partai Perindo menjadi salah satu pendukung Capres ganjarpranowo akan semakin menambah kekuatan untuk menghadapi
2	Kecerdasan dan kemampuan Airlangga Hartarto @airlangga_hrt dalam mengantisipasi situasi politik yang dinamis menjelang pemilu dan pilpres 2024. Airlangga akan menjadi penentu ritme dalam membangun koalisi partai politik yang akan mengusung	Kecerdasan dan kemampuan Airlangga Hartarto hrt dalam mengantisipasi situasi politik yang dinamis menjelang pemilu dan pilpres Airlangga akan menjadi penentu ritme dalam membangun koalisi partai politik yang akan mengusung capres cawapres nantinya Terbaik

Pembersihan data		
No	Tweet	Cleansing
	capres cawapres nantinya. Terbaik! https://t.co/DeuE6LI3QV	
3	In my prediction, ini adalah lawan sepadan untuk Pak Prabowo. Gerindra sebaiknya merangkul Demokrat agar bisa menang Pilpres. PKS mungkin membuat koalisi baru dengan partai kecil untuk mengusung Anies sebagai Capres	In my prediction ini adalah lawan sepadan untuk Pak Prabowo Gerindra sebaiknya merangkul Demokrat agar bisa menang Pilpres PKS mungkin membuat koalisi baru dengan partai kecil untuk mengusung Anies sebagai Capres

Mengubah kata *tweet* menjadi huruf besar atau kecil adalah langkah berikutnya, yang dikenal sebagai *scotch-fold*. Dalam kasus ini, hasil langkah *case folding* disajikan dalam bentuk tabel. Model data yang mengikuti proses *case folding* disajikan dalam Tabel 4.

Tabel 4 Hasil Data *Case Folding*

Proses Case Folding		
No	Tweet	Case Folding
1	Bergabungnya Partai Perindo menjadi salah satu pendukung Capres ganjarpranowo akan semakin menambah kekuatan untuk menghadapi	bergabungnya partai perindo menjadi salah satu pendukung capres ganjarpranowo akan semakin menambah kekuatan untuk menghadapi
2	Kecerdasan dan kemampuan Airlangga Hartarto hrt dalam mengantisipasi situasi politik yang dinamis menjelang pemilu dan pilpres Airlangga akan menjadi penentu ritme dalam membangun koalisi	kecerdasan dan kemampuan airlangga hartarto hrt dalam mengantisipasi situasi politik yang dinamis menjelang pemilu dan pilpres airlangga akan menjadi penentu ritme dalam membangun koalisi

Proses Case Folding		
No	Tweet	Case Folding
	partai politik yang akan mengusung capres cawapres nantinya Terbaik	partai politik yang akan mengusung capres cawapres nantinya terbaik
3	In my prediction ini adalah lawan sepadan untuk Pak Prabowo Gerindra sebaiknya merangkul Demokrat agar bisa menang Pilpres PKS mungkin membuat koalisi baru dengan partai kecil untuk mengusung Anies sebagai Capres	in my prediction ini adalah lawan sepadan untuk pak prabowo gerindra sebaiknya merangkul demokrat agar bisa menang pilpres pks mungkin membuat koalisi baru dengan partai kecil untuk mengusung anies sebagai capres

Langkah selanjutnya adalah melakukan tokenisasi, yang merupakan proses memotong teks menjadi bagian-bagian token. Setelah tokenisasi, data diolah dan ditampilkan dalam tabel. Hasil tokenisasi dapat dilihat di Tabel 5.

Tabel 5 Data yang Dihasilkan dari Tokenisasi

Proses Tokenisasi		
No	Tweet	Tokenisasi
1	bergabungnya partai perindo menjadi salah satu pendukung capres ganjarpranowo akan semakin menambah kekuatan untuk menghadapi	['bergabungnya', 'partai', 'perindo', 'menjadi', 'salah', 'satu', 'pendukung', 'capres', 'ganjarpranowo', 'akan', 'semakin', 'menambah', 'kekuatan', 'untuk', 'menghadapi']
2	kecerdasan dan kemampuan airlangga hartarto hrt dalam mengantisipasi situasi politik yang dinamis menjelang pemilu dan pilpres airlangga akan menjadi penentu ritme dalam membangun koalisi partai politik yang akan mengusung capres cawapres nantinya terbaik	['kecerdasan', 'dan', 'kemampuan', 'airlangga', 'hartarto', 'hrt', 'dalam', 'mengantisipasi', 'situasi', 'politik', 'yang', 'dinamis', 'menjelang', 'pemilu', 'dan', 'pilpres', 'airlangga', 'akan', 'menjadi', 'penentu', 'ritme', 'dalam', 'membangun', 'koalisi', 'partai', 'politik', 'yang', 'akan', 'mengusung', 'capres', 'cawapres', 'nantinya', 'terbaik']

Proses Tokenisasi		
No	Tweet	Tokenisasi
		'politik', 'yang', 'akan', 'mengusung', 'capres', 'cawapres', 'nantinya', 'terbaik']
3	in my prediction ini adalah lawan sepadan untuk pak prabowo gerindra sebaiknya merangkul demokrat agar bisa menang pilpres pks mungkin membuat koalisi baru dengan partai kecil untuk mengusung anies sebagai capres	['in', 'my', 'prediction', 'ini', 'adalah', 'lawan', 'sepadan', 'untuk', 'pak', 'prabowo', 'gerindra', 'sebaiknya', 'merangkul', 'demokrat', 'agar', 'bisa', 'menang', 'pilpres', 'pks', 'mungkin', 'membuat', 'koalisi', 'baru', 'dengan', 'partai', 'kecil', 'untuk', 'mengusung', 'anies', 'sebagai', 'capres']

Di tahap selanjutnya, metode *remove stopwords* digunakan. Tujuan dari metode ini adalah untuk menghapus atau menghilangkan kata-kata *stopwords* dari teks yang sedang dianalisis, sehingga teks menjadi lebih sederhana dan perhatian dapat difokuskan pada kata-kata yang lebih informatif untuk dianalisis. Hal ini dapat meningkatkan kualitas analisis, mengurangi suara data, dan meningkatkan akurasi atau efisiensi proses berikutnya seperti analisis sentimen atau klasifikasi. Tabel 6 menunjukkan hasil penghilangan *stopwords*.

Tabel 6 Data Hasil Menghilangkan *Stopwords*

Proses Remove Stopwords		
No	Tweet	Hapus Stopwords
1	['bergabungnya', 'partai', 'perindo', 'menjadi', 'salah', 'satu', 'pendukung', 'capres', 'ganjarpranowo', 'akan', 'semakin', 'menambah', 'kekuatan', 'untuk', 'menghadapi']	['partai', 'perindo', 'pendukung', 'capres', 'ganjarpranowo', 'kekuatan']
2	['kecerdasan', 'dan', 'kemampuan', 'airlangga', 'hartarto', 'hrt', 'dalam',	['kecerdasan', 'kemampuan', 'airlangga', 'hartarto', 'hrt', 'mengantisipasi',

Proses Remove Stopwords		
No	Tweet	Hapus Stopwords
	'mengantisipasi', 'situasi', 'politik', 'yang', 'dinamis', 'menjelang', 'pemilu', 'dan', 'pilpres', 'airlangga', 'akan', 'menjadi', 'penentu', 'ritme', 'dalam', 'membangun', 'koalisi', 'partai', 'politik', 'yang', 'akan', 'mengusung', 'capres', 'cawapres', 'nantinya', 'terbaik']	'situasi', 'politik', 'dinamis', 'menjelang', 'pemilu', 'pilpres', 'airlangga', 'penentu', 'ritme', 'membangun', 'koalisi', 'partai', 'politik', 'mengusung', 'capres', 'cawapres', 'terbaik']
3	['in', 'my', 'prediction', 'ini', 'adalah', 'lawan', 'sepadan', 'untuk', 'pak', 'prabowo', 'gerindra', 'sebaiknya', 'merangkul', 'demokrat', 'agar', 'bisa', 'menang', 'pilpres', 'pks', 'mungkin', 'membuat', 'koalisi', 'baru', 'dengan', 'partai', 'kecil', 'untuk', 'mengusung', 'anies', 'sebagai', 'capres']	['in', 'my', 'prediction', 'lawan', 'sepadan', 'prabowo', 'gerindra', 'merangkul', 'demokrat', 'menang', 'pilpres', 'pks', 'koalisi', 'partai', 'mengusung', 'anies', 'capres']

3.2 Model Klasifikasi

Data dianalisis menggunakan 3 algoritma *Support Vector Machine* (SVM), *K-Nearest Neighbor* (K-NN), dan *Naive Bayes* (NB). Teknik penambangan data digunakan dalam proses ini. Model klasifikasi yang digunakan dalam penelitian ini membandingkan tiga metode tersebut. Desain model klasifikasi yang digunakan dalam aplikasi RapidMiner.

Confusion Matrix adalah metrik penilaian yang digunakan untuk mengukur skor akurasi. Ini adalah tabel data uji yang besar yang menunjukkan hasil prediksi yang kuat dan lemah dari model klasifikasi yang digunakan. Digunakan untuk mengetahui kinerja model klasifikasi yang dapat menghasilkan *recall*, *precision*, dan *accuracy* [12]. Tabel 7 menunjukkan hasil klasifikasi algoritma SVM, NB, dan KNN untuk kata kunci Capres dan Pilpres, dan Gambar 8 menunjukkan hasil klasifikasi algoritma NB. Dari ketiga gambar perbandingan, terlihat bahwa algoritma NB memiliki kinerja yang cukup baik dibandingkan dengan algoritma SVM dan KNN.

Tabel 7 Hasil tiap algoritma

Support Vector Machine			
Accuracy : 87.59%			
	True Negative	True Positive	Class Precision
Pred. Negative	1205	299	80.12%
Pred. Positive	0	906	100%
Class Recall	100%	75.19%	
Naïve Bayes			
Accuracy : 96.39%			
	True Negative	True Positive	Class Precision
Pred. Negative	1205	87	93.27%
Pred. Positive	0	1118	100%
Class Recall	100%	92.78%	
K-Nearest Neighbor			
Accuracy : 57.43%			
	True Negative	True Positive	Class Precision
Pred. Negative	1203	1024	54.02%
Pred. Positive	2	181	98.91%
Class Recall	99.83%	15.02%	

Hasil menunjukkan bahwa model klasifikasi berfungsi dengan baik. Setiap algoritma diklasifikasikan berdasarkan nilai *AUC*-nya, yang didefinisikan sebagai berikut: 0,50–0,60 = Gagal; 0,60–0,70 = Buruk; 0,70–0,80 = Cukup; 0,80–0,90 = Baik; dan 0,90–1,00 = Sangat Baik. Hasil klasifikasi algoritma ini disajikan dalam Tabel 8.

Tabel 8 Hasil dari perbandingan algoritma klasifikasi.

Perbandingan Algoritma Klasifikasi					
Algoritma	Keakuratan	<i>AUC</i>	<i>Precision</i>	<i>Recall</i>	Klasifikasi
SVM	87,59%	0,982	100%	75,19%	Sangat Baik
NB	96,39%	0,500	100%	92,78%	Gagal
KNN	57,43%	0,761	98,91%	15,02%	Cukup

Dengan nilai tingkat akurasi 96,39% dibandingkan dengan algoritma *Support Vector Machines* (SVM) 87,59% dan *K-Nearest Neighbor* (KNN) 57,43%, hasil yang diatas menunjukkan bahwa algoritma *Naive Bayes* (NB) memiliki perfoma yang baik.

4. KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa penerapan klasifikasi sentimen dalam analisis sentimen Capres dan Pilpres dapat diterapkan menggunakan beberapa metode algoritma seperti SVM, *Naive Bayes* dan K-NN. Tujuan dari penelitian ini untuk meningkatkan efisiensi pengukuran, berbeda dengan penelitian yang sudah dilakukan sebelumnya yang hanya menerapkan prosedur pendekatan SVM dalam mengklasifikasikan sentimen. Selanjutnya, penelitian ini melakukan perbandingan antara algoritma SVM, *Naive Bayes*, dan K-NN. Hasil pengujian menunjukkan bahwa algoritma *Naive Bayes* memiliki nilai *accuracy* yang signifikan sebesar 96,39% pada dataset Capres dan Pilpres, dengan tingkat *precision* sebesar 100% dan tingkat *recall* sebesar 92,78%. Ini menunjukkan bahwa algoritma ini memiliki performa yang sangat baik dibandingkan dengan algoritma SVM dan K-NN. Penelitian ini masih perlu dikembangkan seperti memperpanjang periode penarikan data, sehingga diharapkan bisa meningkatkan nilai akurasi, selain itu juga bisa dapat mencerminkan hasil analisis sentimen sesuai periode waktu.

DAFTAR PUSTAKA

- [1] D. D. Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (Dpr) Pada Twitter Menggunakan Metode Naive Bayes Classifier," vol. 10, no. 1, pp. 34–40, 2022.
- [2] A. Anjani, A. Chamid, and A. Murti, "Analisis Sentimen Kaum LGBT pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes," *JTINFO J. Tek. Inform.*, vol. 1, no. 2, pp. 1–8, 2022.
- [3] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," vol. 10, pp. 71–76, 2020.
- [4] D. R. Berliana and B. Santoso, "Elektabilitas Ridwan Kamil Dan Anies Baswedan Dalam Simulasi Pilpres 2024 Di Twitter (Analisis Jaringan Media Sosial Dan Analisis Sentimen Pengguna Twitter Terhadap #Ridwankamil Dan #Aniesbaswedan)," vol. 6, no. 2, pp. 150–162, 2022.
- [5] F. Fathonah and A. Herliana, "Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid - 19 Menggunakan Metode Naïve Bayes," vol. 7, no. November, pp. 155–164, 2021, doi: 10.34128/jsi.v7i2.331.
- [6] D. Ni, Z. Xiao, and M. K. Lim, "A systematic review of the research trends of machine learning in supply chain management," no. x, 2019, doi: 10.1007/s13042-019-01050-0.
- [7] J. C. Lyu, E. Le Han, and G. K. Luli, "Covid-19 vaccine-related discussion on twitter: Topic modeling and sentiment analysis," *J. Med. Internet Res.*, vol. 23, no. 6, 2021, doi: 10.2196/24435.
- [8] A. Ahmad and W. Gata, "Sentimen Analisis Masyarakat Indonesia di Twitter Terkait Metaverse dengan Algoritma Support Vector Machine," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 6, no. 4, pp. 548–555, 2022, doi: 10.35870/jtik.v6i4.569.
- [9] M. Raihan, F. Sya, U. Enri, and T. N. Padilah, "Analisis Sentimen Terhadap Bakal Calon Presiden 2024 dengan Algoritma Naïve Bayes," vol. 9, no. 2, pp. 265–273, 2024, doi: 10.30865/jurikom.v9i2.3989.
- [10] F. V. Sari and A. Wibowo, "Analisis Sentimen Pelanggan Toko Online Jd . Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi," vol. 10, no. 2, pp. 681–686, 2019.
- [11] D. Muhidin and A. Wibowo, "Perbandingan Kinerja Algoritma Support Vector Machine Dan K-Nearest Neighbor Terhadap Analisis Sentimen Kebijakan New Normal," vol. 5, no. 2, pp. 153–159, 2020.
- [12] A. P. Nardilasari, A. L. Hananto, S. S. Hilabi, T. Tukino, and B. Priyatna, "Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM Pada Media Sosial Twitter," *JOINTECS (Journal Inf. Technol. Comput. Sci.)*, vol. 8, no. 1, p. 11, 2023, doi:

- 10.31328/jointecs.v8i1.4265.
- [13] F. A. Rohmansyah, B. Bintoro, and I. Santoso, "Analisis Sentimen Terhadap Penerapan Sistem Ganjil Genap Menggunakan Metode K - Nearest Neighbor," vol. 7, no. 2, pp. 165–169, 2023.
 - [14] M. Rangga, A. Nasution, and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," vol. 6, no. 2, pp. 226–235, 2019.