

---

## Perbandingan Nilai Akurasi Analisa Sentiment Pada Kata Kunci Pemilu 2024

I Wayan Suardi<sup>1</sup>, Irene R.H.T. Tangkawarow<sup>2</sup>, Efraim R. S. Moningkey<sup>3</sup>

Suardis289@gmail.com<sup>1</sup>, irene.tangkawarow@unima.ac.id<sup>2</sup>, fmoningkey@unima.ac.id<sup>3</sup>

<sup>1,2,3</sup> Program Studi Teknik Informatika, Universitas Negeri Manado

---

### Informasi Artikel

Diterima : 5 Mar 2025  
Direvisi : 13 Mar 2025  
Disetujui : 15 Apr 2025

---

### Kata Kunci

Analisis sentiment,  
Pemilu 2024, Naive  
Bayes Classifiers,  
Support Vector Machine,  
Media Sosial

---

### Abstrak

Demokrasi merupakan sistem pemerintahan yang memberi hak yang sama bagi setiap warga negara dalam pengambilan keputusan yang dapat mempengaruhi kehidupan mereka. Di Indonesia, sistem demokrasi diwujudkan melalui penyelenggaraan Pemilihan Umum (Pemilu) yang dilaksanakan secara periodik. Pemilu selalu ramai diperbincangkan di dunia nyata maupun dunia maya, khususnya di media sosial X. Mengingat jumlah tweet yang sangat besar, analisis manual terhadap opini publik menjadi tidak efisien. Oleh karena itu, diperlukan teknologi yang dapat menganalisis tweet secara otomatis dan mengkategorikannya ke dalam sentimen positif atau negatif. Penelitian ini akan melakukan perbandingan nilai akurasi analisa sentiment pada kata kunci data pemilu 2024 menggunakan metode naive bayes classifiers dan support vector machine. Data yang digunakan sebanyak 5651 tweet, dan mendapatkan nilai akurasi 64,59% pada metode naive bayes classifiers dan 76,14% pada metode support vector machine. Ini menunjukkan bahwa SVM dapat diandalkan dalam konteks analisis sentimen yang melibatkan data yang kompleks dan beragam.

---

### Keywords

Sentiment analysis, 2024  
General Election, Naive  
Bayes Classifiers, Support  
Vector Machine, Social Media

---

### Abstract

Democracy is a government system that gives equal rights to every citizen in making decisions that can affect their lives. In Indonesia, the democratic system is realized through the General Election (Pemilu) organization, which is held periodically. An election is always discussed in the real and virtual worlds, especially on X social media. Considering the huge number of tweets, manual analysis of public opinion is inefficient. Therefore, technology is needed to analyze and categorize tweets into positive or negative sentiments automatically. This research compared the accuracy value of sentiment analysis on 2024 election data keywords using the NAIVE BAYES CLASSIFIERS and SUPPORT VECTOR MACHINE methods. The data used was 5651 tweets and obtained an accuracy value of 64.59% in the naive bayes classifiers method and 76.14% in the support vector machine method. It shows that SVM is reliable in the context of sentiment analysis involving complex and diverse data.

## A. Pendahuluan

Dalam sebuah demokrasi, setiap warga negara memiliki suara yang setara dalam keputusan yang memiliki kekuatan untuk mengubah hidup mereka. Demokrasi memungkinkan orang untuk secara langsung atau melalui wakil mereka berpartisipasi dalam penyusunan, perumusan, dan pembuatan undang-undang [1].

Untuk menyatukan tujuan masyarakat, pemilihan umum dilakukan sebagai salah satu bentuk demokrasi. Publik memiliki kesempatan untuk memilih wakil politik melalui pemilihan. Pasal 22 E, ayat (1) Undang-Undang Dasar 1945 [2], yang menyatakan bahwa "Pemilihan umum dilaksanakan secara langsung, umum, bebas, rahasia, jujur, dan adil setiap lima tahun," juga mengatur pelaksanaan pemilihan umum di Indonesia untuk menjamin bahwa pemilihan tersebut mengikuti aturan dan prinsip pemilihan. Oleh karena itu, pengawasan diperlukan untuk setiap tahap pemilihan yang dilaksanakan. Di Indonesia, Badan Pengawas Pemilu (Bawaslu) bertanggung jawab untuk mengawasi berbagai tahapan proses penyelenggaraan pemilu. Bawaslu memerlukan dukungan masyarakat dalam bentuk pengawasan partisipatif untuk memenuhi tanggung jawabnya dalam mengawasi pemilihan umum. Peraturan Badan Pengawas Pemilu Nomor 2 Tahun 2023 [3] mengatur pengawasan partisipatif, yang menyatakan bahwa fungsi Panwaslu Kecamatan, Bawaslu Kabupaten/Kota, Bawaslu Provinsi, dan Bawaslu adalah untuk meningkatkan keterlibatan masyarakat dalam pengawasan pemilu.

Antusiasme masyarakat terhadap pemilihan umum 2024 tidak terbatas pada dunia nyata; hal ini juga hadir di dunia Maya, seperti di platform media sosial seperti X. Pada periode saat ini, media sosial dan internet telah berkembang pesat, menyediakan berbagai informasi dan sudut pandang dari orang lain. X, yang dulunya dikenal sebagai Twitter sebelum mengalami perubahan nama, sekarang menjadi situs jejaring sosial terpopuler. X adalah salah satu platform media sosial yang paling banyak digunakan untuk mengumpulkan informasi. Informasi ini bisa bersifat faktual atau subjektif, seperti opini positif atau negatif yang diungkapkan dalam bentuk tweet di X. Karena hal ini, peneliti dapat mengumpulkan informasi di X karena pengguna dapat menggunakannya untuk memahami opini publik mengenai pemilihan umum 2024, mengidentifikasi baik opini positif maupun negatif dalam sebuah tweet. Namun, seperti yang dapat dilihat dari jumlah pengguna yang besar. Jadi, dibutuhkan lebih banyak waktu dan kesabaran. Karena analisis manusia masih belum terlalu efisien, maka diperlukan alat yang dapat menganalisis tweet dan secara otomatis mengklasifikasikan sejumlah besar tweet sebagai positif atau negatif.

Penerapan machine learning dalam analisis sentimen juga dapat mempercepat proses pemantauan opini publik yang selama ini memerlukan banyak tenaga manusia dan waktu yang panjang. Dengan menggunakan algoritma seperti Naive Bayes dan Support Vector Machine (SVM), analisis terhadap jutaan tweet yang tersebar di media sosial dapat dilakukan secara otomatis dalam waktu yang relatif singkat. Ini akan memungkinkan penyelenggara pemilu, partai politik, dan lembaga pengawas untuk lebih cepat menanggapi perubahan opini publik dan potensi kerawanan yang dapat mempengaruhi jalannya pemilu.

perkembangan analisis sentimen juga dapat dimanfaatkan untuk memperbaiki kualitas debat politik dan meningkatkan partisipasi masyarakat dalam proses demokrasi. Dengan memahami opini publik melalui analisis sentimen,

para pemangku kepentingan dapat menciptakan kampanye yang lebih efektif dan mengatasi masalah yang penting bagi masyarakat umum.

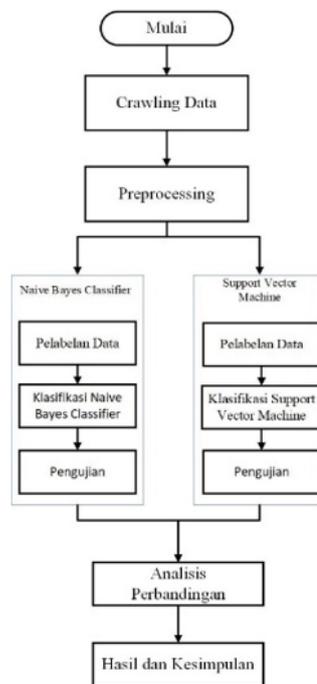
Pemilihan algoritma Naive Bayes Classifiers dan Support Vector Machine (SVM) dalam penelitian ini didasarkan pada keunggulan dan karakteristik unik dari masing-masing metode dalam menangani tugas klasifikasi teks. Naive Bayes Classifiers, yang didasarkan pada teorema Bayes, terkenal karena kesederhanaannya dan kemampuannya dalam memberikan hasil yang baik meskipun dengan data latihan yang relatif sedikit [4] Algoritma ini sangat efisien dan cepat dalam memproses data, sehingga cocok untuk analisis sentimen dalam jumlah besar seperti tweet. Di sisi lain, SVM dikenal karena kemampuannya dalam menghasilkan akurasi tinggi, SVM bekerja dengan menggunakan hyperplane ideal yang meminimalkan margin antara kelas-kelas yang berbeda, terutama dalam kasus ketika data tidak linear, yang membuatnya sangat efektif dalam menangani data dengan dimensi tinggi seperti teks [5].

Dengan demikian, pada penelitian ini akan dilakukan Perbandingan Nilai Akurasi Analisa Sentiment Pada kata kunci Pemilu 2024 Menggunakan Metode Naive Bayes Classifiers Dan Support Vector Machine untuk mengetahui metode manakah yang memiliki tingkat akurasi yang lebih baik untuk mengklasifikasi sentiment positif dan negatif pada kata kunci data pemilu 2024.

## **B. Metode Penelitian**

Metode yang digunakan pada penelitian ini adalah metode Naive Bayes dan Support Vector Machine (SVM), di mana kedua metode ini diterapkan untuk menganalisis sentimen berdasarkan data yang diperoleh dari media sosial. Naive Bayes digunakan karena kemampuannya yang sederhana namun efektif dalam memprediksi sentimen berdasarkan probabilitas, sementara SVM dipilih karena kekuatannya dalam mengklasifikasikan data dengan margin yang optimal. kedua metode ini diharapkan dapat memberikan hasil analisis yang lebih akurat dan mendalam terhadap sentimen publik terkait pemilu 2024, Pengambilan sampel data dilakukan secara acak pada tanggal 5 Maret 2024, menggunakan kata kunci "pemilu 2024," yang berhasil mengumpulkan sebanyak 5.651 tweet sebagai data penelitian.

Tahapan penelitian ini dilakukan secara sistematis untuk menganalisis sentimen publik terkait pemilu 2024, dengan langkah-langkah utama yang meliputi pengumpulan data melalui crawling, preprocessing data, pelabelan data, pengklasifikasian menggunakan metode Naive Bayes dan Support Vector Machine (SVM), pengujian model, hingga analisis perbandingan hasil kedua metode. Berikut adalah penjelasan untuk masing-masing tahapan:



**Gambar 1.** Tahapan Penelitian

### Crawling Data

Crawling adalah proses pengumpulan data dari web, baik yang berukuran besar maupun kecil, Data ini diambil berdasarkan kata kunci tertentu yang telah ditetapkan, kemudian disimpan di penyimpanan local [6]. Proses pengambilan data menggunakan pustaka Tweepy. Penggunaan pustaka ini memerlukan akses ke OAuth 1 yang dapat diperoleh melalui situs web Pengembang Twitter. OAuth 1 memberikan izin untuk mengakses API yang disediakan oleh Twitter. Library ini memanfaatkan API Twitter berupa fitur search, yang memungkinkan pencarian tweet berdasarkan kata kunci tertentu. Pengambilan sampel data menggunakan kata kunci "pemilu 2024" berhasil mengumpulkan sebanyak 5.651 tweet yang dilakukan pada tanggal 5 Maret 2024, yang kemudian digunakan sebagai data utama dalam penelitian ini.

### Preprocessing

preprocessing pada tahapan ini Data yang terkumpul sebagian besar terdiri dari materi tidak terstruktur dari setiap tweet, sebagian besar dalam bahasa non-baku. Tujuan dari langkah ini adalah untuk menghilangkan karakter yang tidak relevan dan menurunkan kualitas model; selain itu, hal ini dapat meningkatkan kualitas data. dengan tujuan supaya data digunakan dapat digunakan pada tahap selanjutnya. Adapapun tahapan yang dilakukan pada penelitian ini adalah :

Pembersihan data, Proses pembersihan data ini bertujuan untuk menghilangkan karakteristik yang menurunkan kualitas data, seperti tautan, nama pengguna, handle X (Twitter), dan tagar. Program akan mencari karakter yang disebutkan sebelumnya, dan kemudian karakter tersebut akan dihilangkan dari data Twitter [7].

Tokenizing adalah Proses mengubah dokumen menjadi kumpulan kata Tokenisasi dapat dilakukan dengan menghilangkan tanda baca dan memisahkannya per spasi. Selain itu, taktik ini menonaktifkan beberapa karakter, seperti tanda baca, dan mengubah semua token menjadi huruf kecil. [8].

Proses normalisasi berkaitan dengan model data relasional dan digunakan untuk mengatur data dengan sensitivitas dan korelasi tinggi atau rendah. Hasil dari proses normalisasi adalah data (tabel-tabel) dalam bentuk normal [9].

Remove Stopwords, tahapan ini menggunakan kata-kata yang tidak deskriptif maupun krusial. Kata ini harus dihindari karena tidak mengandung atau memberikan data. Stopword adalah kata-kata yang memiliki frekuensi kemunculan yang tinggi [10].

Stemming merupakan Proses menghilangkan akhiran dan imbuhan untuk setiap token [11]. Proses ini menggunakan pustaka Pysastrawi, yang didasarkan pada algoritma Nazief dan Adriani, untuk mengubah imbuhan sebuah kata menjadi kata dasar. Algoritma ini mencakup fitur bahasa yang kompleks, ambiguitas, overstemming, understemming, bentuk jamak, serapan, akronim, dan hal-hal lainnya. Sastrawi memiliki 40 aturan pemenggalan yang digunakan untuk menjalankan proses stemming.

### **Pelabelan Data**

Dalam proses pembuatan dataset untuk sistem klasifikasi, diperlukan mekanisme untuk memastikan bahwa dataset yang telah dibuat memiliki label yang akurat. Proses pelabelan dataset cenderung mudah jika jumlah data relatif kecil, Namun, ini akan memakan waktu yang sangat lama, bahkan mungkin tidak memungkinkan untuk dilakukan secara manual, jika dataset berukuran sangat besar. Penelitian ini menggunakan kamus InSet Lexicon untuk melabeli dataset secara otomatis, khususnya pada setiap tweet. Kamus InSet Lexicon merupakan kumpulan kata yang bernada positif dan negatif. Dalam proses pelabelan, setiap kata dalam kalimat akan dibandingkan dengan entri dalam kamus berdasarkan polaritasnya. Kalimat yang memiliki polaritas lebih dari 0 akan diberi label positif, sedangkan kalimat dengan polaritas kurang dari 0 akan diberi label negatif. Sementara itu, jika polaritasnya sama dengan 0, maka kalimat tersebut diberi label netral.

### **Naïve Bayes Classifier**

Pengklasifikasi Naïve Bayes (NBC) digunakan untuk menentukan probabilitas atau kemungkinan hasil tertentu berdasarkan data dari eksperimen sebelumnya atau untuk memungkinkan pembuatan pengujian untuk suatu sistem [12]. Dalam studi ini, metode NBC digunakan untuk mengelompokkan bahasa sesuai dengan label yang akan digunakan sehingga setiap label akan memiliki bobot dan bahasa yang digunakan akan diidentifikasi sesuai dengan label.

### **Support Vector Machine (SVM)**

Support Vector Machine (SVM) adalah algoritma klasifikasi yang dilakukan dengan mengidentifikasi sebuah hyperplane. Hyperplane tersebut akan diposisikan secara akurat di kedua kelas, sehingga akan memiliki jarak yang paling akurat ke data-data yang ada di kedua kelas tersebut. SVM digunakan untuk mengumpulkan

data yang telah dipahami atau dibuat berdasarkan klasifikasi sehingga dapat menilai akurasi data dalam sebuah sistem [13].

**Confusion Matrix**

Penelitian ini akan menggunakan Confusion Matrix, atau dikenal juga sebagai Matriks Kebingungan, yang merupakan salah satu teknik yang dapat digunakan untuk menentukan beberapa model yang baik untuk klasifikasi sebuah dataset dalam machine learning atau pembelajaran mesin. Matriks kebingungan biasanya digunakan dalam konteks masalah klasifikasi biner (dua kelas). Matriks kebingungan berisi prediksi dan hasil aktual, dan masing-masing memiliki kombinasi yang unik. [14]. Dengan menggunakan Confusion Matrix memungkinkan untuk mendapatkan hasil perhitungan yang lebih baik tentang seberapa baik model klasifikasi.

**Perbandingan nilai akurasi**

Setelah dilakukan pengujian dengan Confusion Matrix dari metode Naïve Bayes dan Support Vector Machine maka hasil yang di dapat selanjutnya akan di lakukan perbandingan nilai akurasi untuk mengetahui metode manakah yang paling akurat untuk melakukan sentiment analisis pada kata kunci pemilu 2024.

**C. Hasil dan Pembahasan**

Penelitian ini menggunakan tweet dari media sosial X (Twitter) dengan memanfaatkan library bahasa pemrograman python yang berisikan API untuk dapat mengakses X (Twitter). Proses pengambilan Data dilakukan tanggal 5 maret 2024, tweet yang diambil adalah tweet yang berbahasa Indonesia dengan kata kunci “PEMILU 2024” dengan tweet yang berhasil dikumpulkan berjumlah 5651 tweet yang kemudian disimpan dengan format .csv.

**Preprocessing**

Setelah data set ditetapkan, tahapan selanjutnya adalah preprocessing. Pada tahapan ini, data hasil dari proses crawling akan diproses hingga menjadi data yang bersih dan siap digunakan untuk analisis lebih lanjut. Berikut tahapan dalam melakukan preprocessing:

Untuk mendapatkan hasil optimal, maka data yang digunakan harus memiliki kualitas yang baik. Untuk itu, diperlukan proses pembersihan pada dataset awal untuk memperbaiki atau menghapus data yang tidak akurat, tidak lengkap, tidak konsisten, atau tidak relevan. data tweet sering kali mengandung teks yang tidak relevan seperti spam, tautan, atau mention yang tidak diperlukan. Oleh karena itu, langkah awal dalam pembersihan adalah menghapus elemen-elemen tersebut untuk mempertahankan teks yang bermakna. Selain itu, tweet sering kali mengandung kesalahan penulisan, bahasa informal, serta penggunaan singkatan atau emoji yang dapat mempengaruhi hasil analisis. Pada Tabel 1 Menunjukkan hasil cleaning data.

**Tabel 1.** Hasil Cleaning data

Text	Hasil pembersihan data
@Yayanjay_	@KawalPemilu_org brojikalau dari awal sudah ada
@mazzini_gsp	@ainunnajib kesalahan fatalmaka sulit

Bro,,jikalau dari awal sudah ada mengembalikan kepercayaan  
kesalahan fatal,maka sulit rakyat saat ini  
mengembalikan kepercayaan rakyat  
saat ini .

Tokenizing merupakan proses di mana teks yang telah dibersihkan dipecah menjadi unit-unit kecil yang disebut token. Dalam konteks analisis teks, token biasanya berupa kata, frasa, atau bahkan karakter individu, tergantung pada kebutuhan analisis. Proses ini bertujuan untuk mengubah teks yang berupa kalimat atau paragraf menjadi daftar token yang lebih mudah dianalisis secara komputasional. Tabel 2 menunjukkan hasil dari proses tokenizing.

**Tabel 2.** Hasil Tokenizing

Text	Hasil Tokenizing
brojikalau dari awal sudah ada kesalahan fatalmaka sulit mengembalikan kepercayaan rakyat saat ini	brojikalau, dari, awal, sudah, ada, kesalahan, fatalmaka, sulit, mengembalikan, kepercayaan, rakyat, saat, ini,

Normalisasi adalah proses yang bertujuan untuk menyamakan format dan struktur teks agar lebih konsisten dan seragam. Pada tahap ini, token-token yang telah dihasilkan dari proses tokenizing mengalami beberapa penyesuaian, seperti konversi huruf kapital menjadi huruf kecil, penghilangan tanda baca yang tidak relevan, serta penggantian singkatan dengan kata yang lengkap. Tabel 3 menunjukkan hasil dari tahapan normalisasi.

**Tabel 3.** Hasil Normalisasi

Text	Hasil Normalisasi
brojikalau, dari, awal, sudah, ada, kesalahan, fatalmaka, sulit, mengembalikan, kepercayaan, rakyat, saat, ini,	brojikalau, dari, awal, sudah, ada, kesalahan, fatalmaka, sulit, mengembalikan, kepercayaan, rakyat, saat, ini,

Stopword merupakan kata-kata umum yang sering muncul dalam teks namun tidak memiliki makna signifikan dalam konteks analisis, seperti 'dan', 'yang', 'di', 'ke', dan sebagainya. Penghapusan stopwords bertujuan untuk mengurangi jumlah kata yang tidak penting, sehingga fokus analisis dapat lebih diarahkan pada kata-kata yang memiliki makna kontekstual lebih kuat.

**Tabel 4.** Hasil Penghapusan Stopword

Text	Hasil Stopword
brojikalau, dari, awal, sudah, ada, kesalahan, fatalmaka, sulit, mengembalikan, kepercayaan, rakyat, saat, ini,	kesalahan, fatalmaka, sulit, mengembalikan, kepercayaan, rakyat

Pada Tabel 4 menunjukkan hasil dari proses penghapusan stopwords, kata kata seperti "dari", "awal", "sudah", "saat", dan "ini" dihapus karena tidak memiliki makna yang signifikan. Hasilnya berupa dataset yang siap untuk di gunakan dalam tahapan analisis selanjutnya.

Pada proses stemming mengubah kata-kata yang telah di-tokenisasi dan dinormalisasi menjadi bentuk dasarnya (root form). Tujuan utama dari stemming adalah untuk mengurangi berbagai bentuk kata menjadi bentuk dasar yang sama. Tabel 5 merupakan hasil dari dataset yang sudah melalui proses stemming.

**Tabel 5.** Hasil Stemming

Text	Hasil Stemming
kesalahan, fatalmaka, sulit, salah, fatalmaka, sulit, kembali, mengembalikan, kepercayaan, rakyat	percaya, rakyat,

Stemming data merupakan tahapan terakhir pada tahapan Preprocessing Data, Setelah tahap text preprocessing yang dilakukan pada 5.651 tweet, kemudian didapatkan sebanyak 4.591 tweet bersih yang sudah sesuai dengan kebutuhan proses text mining.

### Pelabelan

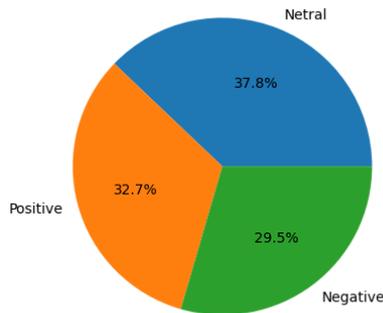
Dataset yang sudah melalui tahapan preprocessing kemudian dilanjutkan ke proses pelabelan. Pada tahap ini, setiap tweet dalam dataset diberi label berdasarkan sentimen yang dikandungnya, apakah positif, negatif, atau netral. Pada penelitian ini saya menggunakan kamus Inset Lexicon, Inset Lexicon merupakan sebuah kamus sentimen yang dirancang khusus untuk analisis teks dalam bahasa Indonesia. Cara kerjanya melibatkan pencocokan setiap kata dalam teks dengan daftar kata yang ada di kamus. Sebagian kamus kata Inset Lexicon dari kata keseluruhan yang berjumlah 3585 kata. Ketika sebuah kata dari teks ditemukan dalam Inset Lexicon, kata tersebut akan diberikan skor sentimen sesuai dengan labelnya. Skor ini kemudian diakumulasikan untuk seluruh teks, sehingga menghasilkan skor total yang menunjukkan apakah teks tersebut cenderung positif, negatif, atau netral. Gambar 2 merupakan hasil dari tahap pelabelan data yang dilakukan dengan Inset Lexicon.

	tweet	positive_counts	negative_counts	label
0	salah fatalmaka sulit kembali percaya rakyat p...	3	3	Netral
1	abud kavling surga sakjose buta data real inde...	1	1	Netral
2	terima kasih senang baca komentar paham data	2	0	Positive
3	bahas politik mabuk politik harga masuk suka n...	0	1	Negative
4	hancur negara inidg dinasti oligarki culas kkn...	2	3	Negative
5	kumpul manusia rakus bagi cerita meja makan me...	0	0	Netral
6	tokoh bangsa mantan wapres bijak buka kompor l...	2	0	Positive
7	percaya kpu rekap beres bukti video kacau hitu...	2	2	Netral
8	milih program sogo ancam apatis lalu records g...	3	1	Positive
9	ayo buka kawal pemilu berani engga bilang dung...	2	3	Negative

**Gambar 2.** Hasil Dari Tahapan Pelabelan

Setelah dilakukan pelabelan menggunakan kamus Inset Lexicon pada dataset yang berjumlah 4.591 tweet. Dari total dataset, sebanyak 1.737 tweet diberi label Netral, 1.500 tweet diberi label Positif, Sementara itu, 1.353 tweet diberi label Negatif. Jika divisualisasikan dalam bentuk pie chart, hasilnya seperti yang ditunjukkan pada Gambar 3 di mana kategori Netral mendominasi dengan

persentase 37,8%, diikuti oleh kategori Positif sebesar 32,7%, dan kategori Negatif sebesar 29,5%.



**Gambar 3.** Pie Chart Hasil Pelabelan

Berdasarkan hasil perhitungan pelabelan menggunakan kamus inset lexicon dapat disimpulkan bahwa respon positif dan netral berada pada posisi lebih dari 50% respon di media sosial. Hal ini menunjukkan respon yang baik terhadap pemilu 2024.

**Analisis metode Naïve Bayes Classifier**

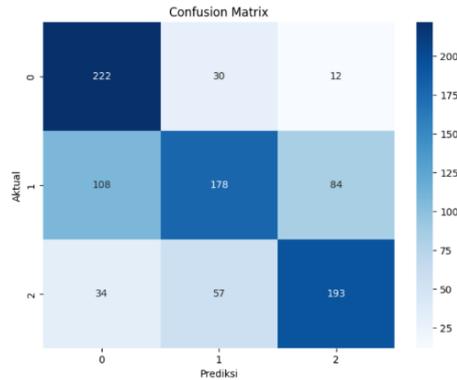
Untuk menentukan sentimen pada tweet, peneliti menggunakan algoritma Naive Bayes yang diimplementasikan menggunakan bahasa Python. Proses ini menghasilkan laporan klasifikasi (classification report), yang mencakup metrik seperti recall, presisi (precision), dan akurasi (accuracy). Sehingga didapatkan perhitungan nilai akurasi menggunakan metode Naïve Bayes yaitu:

**Tabel 6.** Hasil Akurasi Naïve Bayes

accuracy_score	0.6459694989106753			
	precision	recall	f1-score	support
Negatif	0.61	0.84	0.71	264
Netral	0.67	0.48	0.56	370
Positif	0.67	0.68	0.67	284

hasil evaluasi model Naive Bayes menunjukkan akurasi keseluruhan sebesar 64.6%, yang berarti sekitar 64.6% dari prediksi model terhadap data uji adalah benar. Pada kelas Negatif, model menunjukkan performa yang cukup baik dengan presisi sebesar 0.61 dan recall sebesar 0.84, menghasilkan F1-Score sebesar 0.71. Hal ini mengindikasikan bahwa model lebih sering berhasil mengidentifikasi instance kelas Negatif dengan benar. Kelas Positif juga menunjukkan performa yang cukup memuaskan dengan presisi dan recall masing-masing sebesar 0.67 dan 0.68, menghasilkan F1-Score sebesar 0.67, yang berarti model dapat mengenali instance kelas Positif dengan cukup baik.

Kinerja evaluasi sangat penting untuk menentukan hasil model dengan akurat, Berdasarkan hasil analisis menggunakan metode Naïve Bayes, selanjutnya akan dilakukan evaluasi dalam bentuk Confusion Matrix yang dapat dilihat pada Gambar 4.



**Gambar 4.** Confusion Matrix Analisis naïve bayes

Hasil evaluasi menunjukkan bahwa model Naive Bayes mengklasifikasikan tiga kelas: kelas 0 (Negatif), kelas 1 (Netral), dan kelas 2 (Positif). Untuk kelas Negatif, model berhasil memprediksi 222 instance dengan benar, namun salah memprediksi 30 instance sebagai kelas Netral dan 12 instance sebagai kelas Positif. Pada kelas Netral, model memprediksi dengan benar 178 instance, tetapi salah mengklasifikasikan 108 instance sebagai kelas Negatif dan 84 instance sebagai kelas Positif. Di kelas Positif, model memprediksi 193 instance dengan benar, tetapi salah memprediksi 34 instance sebagai kelas Negatif dan 57 instance sebagai kelas Netral. Secara keseluruhan, akurasi model adalah 64.6%, yang menunjukkan bahwa 64.6% dari prediksi model adalah benar. Perhitungan akurasi dihitung dengan Persamaan 1:

$$\begin{aligned}
 &= \frac{\text{True positif} + \text{True netral} + \text{True Negatif}}{\text{Total data matrix}} \times 100\% \quad 1 \\
 &= \frac{222 + 178 + 193}{918} \times 100\% \\
 &= 64,6\%
 \end{aligned}$$

**Analisis metode Support Vector Machine**

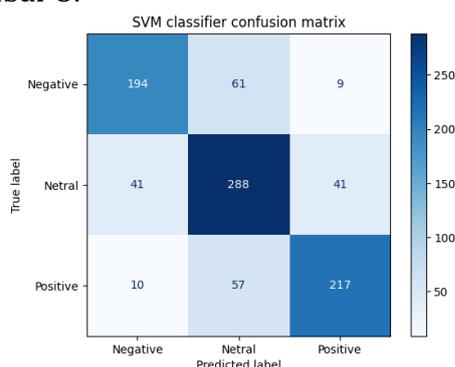
Menentukan sentimen pada tweet, peneliti juga menggunakan algoritma Support Vector Machine (SVM) yang diimplementasikan menggunakan bahasa Python, SVM adalah metode klasifikasi yang dikenal efektif dalam menangani masalah klasifikasi dengan data yang memiliki fitur yang tinggi dan sering kali memberikan hasil yang lebih baik. Model SVM dilatih menggunakan data pelatihan dengan memilih kernel yang sesuai, seperti kernel linier atau Radial Basis Function (RBF), tergantung pada kebutuhan spesifik data dan tujuan analisis. Setelah pelatihan, model SVM diuji pada data pengujian. Sehingga didapatkan perhitungan nilai akurasi menggunakan metode Support vector machine yaitu:

**Tabel 7.** Hasil Akurasi Analisis Support Vector Machine

accuracy_score	0.761437908496732			
	precision	recall	f1-score	support
Negatif	0.79	0.84	0.71	264
Netral	0.71	0.48	0.56	370
Positif	0.81	0.68	0.67	284

Evaluasi model SVM (Support Vector Machine) untuk klasifikasi sentimen menunjukkan kinerja yang cukup baik dengan akurasi keseluruhan sebesar 76,14%. Model ini mampu mengklasifikasikan data dengan baik di tiga kelas utama: negatif, netral, dan positif. Pada kelas negatif, model memiliki precision sebesar 79%, yang berarti 79% dari semua prediksi negatif benar-benar negatif, serta recall sebesar 73%, yang menunjukkan model berhasil mendeteksi 73% dari semua data yang sebenarnya negatif, menghasilkan F1-score sebesar 0,76. Untuk kelas netral, precision sebesar 71% dan recall sebesar 78% menunjukkan bahwa model lebih baik dalam mengidentifikasi data yang sebenarnya netral dibandingkan dengan memprediksi data netral, menghasilkan F1-score sebesar 0,74. Sementara itu, kelas positif memiliki precision sebesar 81% dan recall sebesar 76%, dengan F1-score tertinggi di antara ketiga kelas, yaitu 0,79, yang menunjukkan model sangat baik dalam mengidentifikasi sentimen positif. Hal ini mengindikasikan bahwa model SVM yang digunakan memiliki performa yang seimbang dan akurat dalam mengklasifikasikan sentimen dengan mempertimbangkan distribusi data di setiap kelas.

Berdasarkan hasil analisis menggunakan metode SVM (Support Vector Machine), selanjutnya akan dilakukan evaluasi dalam bentuk Confusion Matrix. Yang bisa di lihat pada Gambar 5.



**Gambar 5.** Confusion Matrix Analisis Support Vector Machine

Pada Gambar 5 menunjukkan Confusion Matrix yang mencakup tiga kelas, yaitu Negatif, Netral, dan Positif, dengan baris yang mewakili label sebenarnya dan kolom yang mewakili label prediksi. Pada kelas Negatif, 194 instance diklasifikasikan dengan benar, sementara 61 instance salah diklasifikasikan sebagai Netral, dan 9 instance salah diklasifikasikan sebagai Positif. Kelas Netral memiliki 288 instance yang diklasifikasikan dengan benar, dengan 41 instance salah diklasifikasikan sebagai Negatif, dan 41 instance lainnya salah diklasifikasikan sebagai Positif. Untuk kelas Positif, 217 instance diklasifikasikan dengan benar, sementara 57 instance salah diklasifikasikan sebagai Netral, dan 10 instance salah diklasifikasikan sebagai Negatif. Matriks ini menunjukkan bahwa model bekerja cukup baik dalam mengklasifikasikan sebagian besar instance dengan benar, terutama pada kelas Netral yang memiliki jumlah klasifikasi benar tertinggi. Dalam konteks Confusion Matrix, jumlah prediksi benar adalah jumlah dari semua nilai True Positif (TP) dan True Negatif (TN) untuk setiap kelas dengan Perhitungan akurasi manual dari perhitungan matrix pada persamaan 2:

$$\begin{aligned}
 &= \frac{\text{True positif} + \text{True netral} + \text{True Negatif}}{\text{Total data matrix}} \times 100\% \quad 2 \\
 &= \frac{217 + 288 + 194}{918} \times 100\% \\
 &= 76,14\%
 \end{aligned}$$

Model klasifikasi SVM yang diuji menunjukkan performa akurasi keseluruhan sebesar 76,14%, dengan precision tertinggi pada kelas positif (0,81) dan recall tertinggi pada kelas netral (0,78). F1-score juga menunjukkan nilai yang baik, dengan yang tertinggi pada kelas positif (0,79). Hasil dari matriks memperlihatkan bahwa model berhasil mengklasifikasikan sebagian besar data dengan benar, terutama pada kelas positif dan netral.

### Analisis Perbandingan

Setelah didapatkan hasil klasifikasi dari kedua metode, selanjutnya akan dilakukan perbandingan akurasi antara Naïve Bayes Classifier dan Support Vector Machine (SVM). Proses perbandingan ini mencakup evaluasi dan analisis metrik kinerja utama, yaitu akurasi. Dengan membandingkan nilai-nilai tersebut, kita akan mengevaluasi kemampuan masing-masing metode dalam melakukan klasifikasi sentimen secara akurat.

**Tabel 8.** Tabel Perbandingan Akurasi

Naïve bayes	SVM
64,6%	76,14%

Dari tabel 8, metode SVM menunjukkan performa paling unggul dengan akurasi sebesar 76,14%, yang mengindikasikan bahwa metode ini mampu memberikan hasil yang sangat baik dan lebih presisi dalam analisis sentimen. Sebaliknya, metode Naïve Bayes hanya mencapai akurasi sebesar 64,6%, yang menunjukkan performa yang kurang memadai jika dibandingkan dengan SVM. Perbedaan akurasi ini menunjukkan bahwa SVM lebih efektif dalam menangani kompleksitas data. Oleh karena itu, untuk analisis sentimen pada kata kunci Pemilu 2024, SVM dapat diandalkan sebagai metode yang lebih unggul dalam menghasilkan hasil prediksi yang akurat dan konsisten.

### D. Simpulan

Berdasarkan hasil pengujian dan perbandingan yang dilakukan dalam penelitian ini, kedua metode yang dipakai, yaitu Naïve Bayes Classifier dan Support Vector Machine (SVM), sama-sama bagus dalam melakukan analisis sentimen pada kata kunci Pemilu 2024. Namun, hasil menunjukkan bahwa SVM memiliki keunggulan yang lebih jelas dibandingkan dengan Naïve Bayes. SVM mencatatkan akurasi yang lebih tinggi, yaitu 76,14%, dibandingkan dengan Naïve Bayes yang mencapai 64,59%. Selain itu, SVM juga unggul dalam hal precision, recall, dan F1-score untuk setiap kelas sentimen, menunjukkan kemampuannya dalam memberikan hasil klasifikasi yang lebih akurat dan konsisten. Naïve Bayes, meskipun memiliki performa yang baik, terutama dalam mengidentifikasi sentimen Negatif dengan

recall yang tinggi, menunjukkan keterbatasan dalam hal akurasi keseluruhan dan keseimbangan antara precision dan recall pada kelas-kelas lain. Di sisi lain, SVM tidak hanya memiliki akurasi keseluruhan yang lebih tinggi, tetapi juga menunjukkan kinerja yang lebih baik dalam mendeteksi kelas-kelas sentimen lainnya, yaitu Netral dan Positif, dengan nilai precision, recall, dan F1-score yang lebih baik. Secara keseluruhan, meskipun kedua metode efektif dalam analisis sentimen kata kunci Pemilu 2024, SVM terbukti lebih superior dalam memberikan akurasi dan keseimbangan klasifikasi yang lebih baik. Oleh karena itu, untuk aplikasi yang membutuhkan klasifikasi sentimen yang akurat dan andal, SVM adalah pilihan yang lebih baik dibandingkan Naïve Bayes. Penelitian ini menunjukkan bahwa SVM dapat diandalkan dalam konteks analisis sentimen yang melibatkan data yang kompleks dan beragam.

### E. Referensi

- [1] L. A. Andika, P. A. N. Azizah, and R. Respatiwan, "Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier," *Indones. J. Appl. Stat.*, vol. 2, no. 1, p. 34, 2019, doi: 10.13057/ijas.v2i1.29998.
- [2] A. H. Sholahuddin, A. Iftitah, and U. D. Mahmudah, "Pelaksanaan Pasal 280 Undang-Undang Republik Indonesia Nomor 7 Tahun 2017 tentang Pemilihan Umum," *J. Supremasi*, vol. 9, no. 2, pp. 17–27, 2019, doi: 10.35457/supremasi.v9i2.793.
- [3] Peraturan Badan Pengawas Pemilihan Umum Republik Indonesia No 2, "tentang Pengawasan Partisipatif," pp. 1–15, 2023.
- [4] I. F. Rahman, A. N. Hasanah, and N. Heryana, "Analisis Sentimen Ulasan Pengguna Aplikasi Samsat Digital Nasional (Signal) Dengan Menggunakan Metode Naïve Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, pp. 963–969, 2024, doi: 10.23960/jitet.v12i2.4073.
- [5] Septi Putri, Yohanes Agung Apriyanto, and Andri Wijaya, "Analisis Sentimen Ulasan Aplikasi Deepl Pada Google Play Dengan Metode Support Vector Machine (Svm)," *J. Sist. Inf.*, vol. 4, no. 2, pp. 59–66, 2023, doi: 10.32546/jusin.v4i2.2368.
- [6] C. Jovanica, D. D. Rahmintangrum, H. A. Nuradni, and A. Salsabila, "Analisis Pengaruh Aktor Pada Tagar #Roketchina Di Media Sosial Twitter Menggunakan Social Network Analysis (Sna)," *J. Ilm. Komun. Makna*, vol. 10, no. 1, p. 43, 2022, doi: 10.30659/jikm.v10i1.15644.
- [7] R. I. Borman and M. Wati, "Penerapan Data Mining Dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandar Lampung Dengan Algoritma Naïve Bayes," *J. Ilm. Fak. Ilmu Komput.*, vol. 09, no. 01, pp. 25–34, 2020.
- [8] H. Y. Firzania *et al.*, "Analisis Sentimen Berbasis Aspek Terhadap Richeese Factory pada Platform Twitter Menggunakan Algoritme Naïve Bayes dan Menggunakan Marketing Mix 4P," vol. 11, no. 4, pp. 3792–3807, 2024.
- [9] D. F. M. Dina, "Normalisasi Database Rancangan Sistem Penyewaan Buku Berbayar," *Comput. Insight J. Comput. Sci.*, vol. 4, no. 1, pp. 56–61, 2024, doi: 10.30651/comp\_insight.v4i1.15814.
- [10] A. Santosa, I. Purnamasari, and Mayasari Rini, "Pengaruh Stopword Removal dan Stemming Terhadap Performa Klasifikasi Teks Komentar Kebijakan New

- Normal Menggunakan Algoritma LSTM," *J. Sains Komput. Inform.*, vol. 6, pp. 81–93, 2022.
- [11] NURUL AFIFAH, Dony Permana, Dodi Vionanda, and Dina Fitria, "Sentiment Analysis of Electric Cars Using Naive Bayes Classifier Method," *UNP J. Stat. Data Sci.*, vol. 1, no. 4, pp. 289–296, 2023, doi: 10.24036/ujsds/vol1-iss4/68.
- [12] Y. S. Sari, "Penerapan Metode Naïve Bayes Untuk Mengetahui Kualitas Air Di Jakarta," *J. Ilm. FIFO*, vol. 13, no. 2, p. 222, 2021, doi: 10.22441/fifo.2021.v13i2.010.
- [13] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)," *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjee.v5i1.16830.
- [14] N. Windy Mardiyah, N. Rahaningsih, and I. Ali, "Penerapan Data Mining Menggunakan Algoritma K-Nearest Neighbor Pada Prediksi Pemberian Kredit Di Sektor Finansial," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 1491–1499, 2024, doi: 10.36040/jati.v8i2.9010.