

Perbandingan Beberapa Algoritma *Machine Learning* Dalam Analisis Sentimen Terkait Pemilihan Presiden RI 2024

Nopan^{1*}, Evangs Mailoa²

Teknik Informatika, Universitas Kristen Satya Wacana, Salatiga, Indonesia
 *e-mail *Corresponding Author*: 672020079@student.uksw.edu

Abstract

The 2024 Presidential Election has become a hot topic among the Indonesian community. People often share their opinions and criticisms openly on social media, such as Twitter or X, leading to debates among the public in supporting a presidential candidate. These debates often result in various public comments that are positive, neutral, or negative. Sentiment analysis is used as a method to analyze public opinions and comments using machine learning algorithms. The purpose of this research is to compare the accuracy levels of the Support Vector Machine, Naive Bayes, and Decision Tree algorithms. This research uses a dataset from Twitter, which will go through stages of data merging, text preprocessing, translation, labeling, and algorithm classification. The results of this research show that the Support Vector Machine algorithm has a higher accuracy rate than the other two algorithms, with an accuracy rate of 81.49%.

Keywords: *Election 2024; Sentiment Analysis; Support Vector Machine; Naive Bayes; Decision Tree.*

Abstrak

Pemilihan Presiden 2024 menjadi topik yang hangat dikalangan masyarakat indonesia, masyarakat sering membagikan pendapat dan kritik secara terbuka dimedia sosial *twitter* atau X, yang menyebabkan subjek perdebatan bagi masyarakat dalam mendukung salah satu kandidat presiden, dampak dari perdebatan ini sering memunculkan berbagai komentar masyarakat yang bersifat positif, netral maupun negatif. Analisis sentimen digunakan sebagai metode untuk menganalisis tentang pendapat dan komentar masyarakat dengan penggunaan algoritma *machine learning*. Tujuan dari penelitian ini dilakukan untuk membandingkan tingkat akurasi dari algoritma *Support Vector Machine*, *Naive Bayes* dan *Decision Tree*, penelitian ini menggunakan *dataset* dari *twitter* yang akan diproses melalui tahapan penggabungan data, *preprocessing text*, *translate*, pelabelan dan klasifikasi algoritma. Hasil dari penelitian ini menunjukkan bahwa algoritma *Support Vector Machine* memiliki tingkat akurasi yang lebih tinggi dari kedua algoritma lainnya dengan nilai akurasi 81.49%.

Kata kunci: *Pilpres 2024; Analisis Sentimen; Support Vector Machine; Naive Bayes; Decision Tree.*

1. Pendahuluan

Pemilu adalah sebuah sistem demokrasi yang ada di setiap negara. Indonesia merupakan salah satunya, yang dimana setiap warga negara bebas untuk memilih pemimpin yang relevan dan baik untuk menentukan nasib negara dan masyarakat. Di era yang *modern* media sosial merupakan salah satu yang menjadi tempat dimana orang-orang mendapatkan informasi dari berbagai aspek, salah satunya adalah Pemilu Presiden 2024 yang akan mendatang. Sudah ada beberapa tokoh ahli politik yang akan mencalonkan diri sebagai Presiden dan Wakil Presiden. Media sosial *twitter* memainkan peran yang semakin *signifikan* dalam membentuk opini dan menyuarakan pendapat masyarakat secara bebas dan terbuka. Berbagai instansi pemerintah menjadikan *twitter* sebagai tempat bagi pemerintah dan masyarakat melakukan komunikasi[1]. Dan juga bisa digunakan untuk mengetahui sejauh mana pandangan dan pendapat masyarakat terkait isu politik pemilihan presiden 2024 dengan memanfaatkan penggunaan metode analisis sentimen.

Twitter atau sekarang disebut *X* merupakan salah satu dari berbagai macam *platform* media sosial yang biasanya digunakan untuk berbagi opini kepada masyarakat, terdapat berbagai fitur seperti mengambil gambar, membagikan, serta mengomentari postingan dengan konteks tertentu[2]. Dalam menjalankan tugas demokratis masyarakat semakin aktif menggunakan *platform* ini untuk berdiskusi, memberikan dukungan dan kritik terhadap kandidat, dan memberikan pandangan pribadi tentang proses pemilihan Presiden. Disisi lain media sosial diharapkan bisa menjadikan tempat menyampaikan pandangan masyarakat secara konstruktif, tapi hal itu tidak selalu sejalan dengan apa yang diharapkan, masyarakat sering memberikan suatu kritik yang beragam baik bersifat positif, netral dan negatif. Dipemilu Presiden 2024 yang mendatang di negara ini akan menjadi momen penting dalam dinamika politik yang demokratis, dan hasil dari pemilihan ini yang nantinya akan mempengaruhi arah masa depan negara[3].

Analisis sentimen merupakan metode dalam melihat dan memahami perasaan dan pandangan masyarakat umum menanggapi isu politik dalam kategori positif, negatif dan netral dimedia sosial menggunakan *machine learning*[4]. Dalam melakukan penelitian ini akan digunakan metode algoritma *Machine Learning Support Vector Machine*(SVM), *Naive Bayes*, dan *Decision Tree*, dengan menggunakan *dataset* dari *twitter* tentang pemilihan presiden 2024. Pelabelan sentimen data dilakukan terlebih dulu dengan menggunakan *library* NLTK(*Natural Language Toolkit*) yang kemudian akan diklasifikasi menggunakan metode algoritma *machine learning* untuk melakukan perbandingan.

Support Vector Machine merupakan algoritma klasifikasi dan regresi yang digunakan dalam pemisahan data *linear* dan *non linear*. *Naive Bayes* adalah algoritma klasifikasi data yang memanfaatkan teorema *Bayes* berdasarkan probabilitas yang digunakan dalam model fitur independen[5]. Dan *Decision Tree* adalah algoritma yang membagi data ke dalam struktur pohon keputusan berdasarkan fitur-fitur yang ada didalam *dataset*[6].

Rumusan masalah dari penelitian ini adalah bagaimana melakukan perbandingan analisis sentimen menggunakan algoritma SVM, *Naive Bayes* dan *Decision Tree*, dan mana yang memiliki tingkat akurasi terbaik dari ketiga algoritma tersebut, dan diharapkan dari penelitian ini dapat memberikan panduan tentang memilih penggunaan algoritma yang efektif dalam melakukan analisis sentimen dan klasifikasi *text* lainnya.

2. Tinjauan Pustaka

Penelitian analisis sentimen dilakukan Mohd. Amiruddin Saddam, Erno Kurniawan D dan Indra tentang “Analisis Sentimen Fenomena PHK Massal Menggunakan *Naive Bayes* dan *Support Vector Machine*” menggunakan data sebanyak 441 *tweet*, hasil dari penelitian tersebut menghasilkan jumlah sentimen positif 108 dan negatif 333, nilai akurasi dari algoritma yang digunakan menghasilkan *Support Vector Machine* dengan nilai 84% dan *Naive Bayes* 74,1%[7].

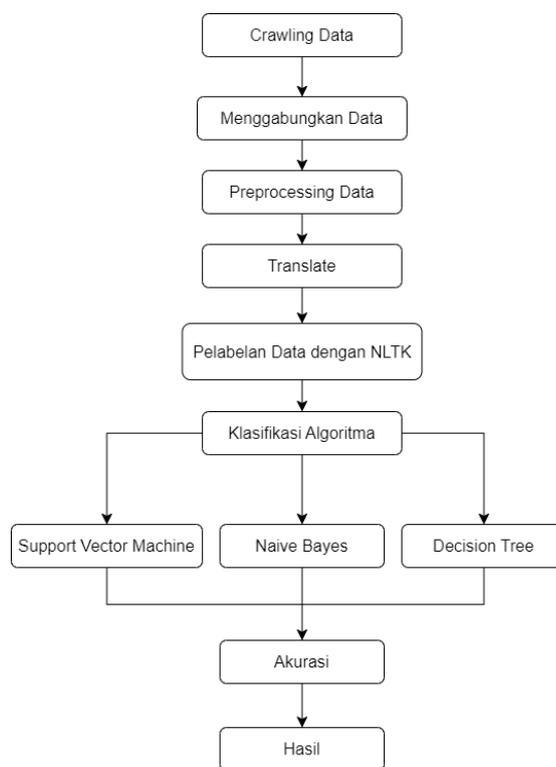
Penelitian yang dilakukan Chandra Ayunda Apta Soemedhy, dkk berjudul “Analisis Komparasi Algoritma *Machine Learning* Untuk *Sentiment Analysis* (Studi Kasus: Komentar *YouTube* “Kekerasan Seksual”) dengan menggunakan algoritma *machine learning Support Vector Machine*, *Naive Bayes* dan *Random Forest*. Dari penelitian yang dilakukan menghasilkan algoritma *Random Forest* dengan akurasi tertinggi dengan nilai 78% dan untuk algoritma SVM dan *Naive Bayes* memiliki tingkat akurasi yang sama dengan nilai 77%[8].

Penelitian yang dilakukan Muhammad Syarifuddin dengan judul “Analisis Sentimen Opini Publik Terhadap Efek PSBB Pada *Twitter* Dengan Algoritma *Decision Tree-KNN-Naive Bayes*” menggunakan data dari *twitter* dengan *search* “Efek PSBB” sebanyak 170 opini. Hasil dari penelitian tersebut menggunakan *tools Rapidminer* menunjukan *Decision Tree* sebagai algoritma dengan nilai tertinggi untuk akurasi sebesar 83,3%, KNN 80,80% dan *Naive Bayes* 80,03%[9].

Pada penelitian terdahulu yang telah dilakukan, menunjukkan penggunaan metode algoritma bervariasi dan menunjukkan hasil yang berbeda. Penelitian ini akan dilakukan dengan perbedaan pada data yang digunakan dengan perubahan dari data yang berbahasa indonesia kedalam bahasa inggris.

3. Metode Penelitian

Metode penelitian adalah sebuah teknik pendekatan yang digunakan dalam melakukan penelitian dengan arti sebagai pendukung untuk membantu mencapai tujuan dari penelitian. Metode penelitian memiliki beberapa langkah yang harus sesuai dengan sistematis penelitian yang akan dilakukan[10].



Gambar 1. Tahap Metode Penelitian

Gambar 1 merupakan kerangka metode penelitian yang akan digunakan untuk melakukan analisis sentimen, dari setiap langkah yang akan dilakukan memiliki pengertian dan tujuan sendiri.

Crawling merupakan teknik yang digunakan dalam pengumpulan data dari berbagai jenis sumber baik berupa website atau media sosial. Didalam penelitian ini menggunakan data dari berbagai *tweet*(postingan) yang ada di aplikasi media sosial *twitter* atau X dengan objek kajian tentang pemilihan presiden 2024, yang kemudian akan dilakukan proses penggabungan data dan *preprocessing text*. Proses pengumpulan data dimulai pada tanggal 22 November 2023 sampai 16 Maret 2024 karena jumlah batasan pengambilan data dari *twitter* maka proses pengambilan data harus dilakukan setiap saat dengan judul *search* "Pemilu Presiden 2024".

Proses penggabungan data digunakan sebagai salah satu metode yang digunakan pada tahap penelitian, berdasarkan dari proses *crawling* data yang telah dilakukan dalam beberapa waktu yang lama untuk mendapatkan data, sehingga hasil dari *crawling* data menghasilkan sejumlah beberapa file yaitu 100 file dengan format *csv*. Jumlah hasil file yang didapatkan digabungkan menjadi satu file menghasilkan jumlah 38.567 baris data.

Preprocessing merupakan tahap dalam proses membersihkan data yang akan digunakan, proses ini bertujuan untuk mengubah data bentuk mentah menjadi data yang bersih sehingga data yang telah siap bisa digunakan pada proses klasifikasi berdasarkan algoritma yang sudah ditentukan. *Dataset* tentang pemilihan presiden 2024 yang digunakan akan dilakukan proses pembersihan data dengan melewati proses tahap *preprocessing* yang akan digunakan berupa normalisasi kata, *cleaning*, *case folding*, *stopword*, *tokenize*, dan *stemming*. Normalisasi kata adalah proses perubahan pada teks data dari awal data yang tidak baku menjadi baku, *cleaning* bertujuan menghapus atau menghilangkan beberapa karakter pada teks seperti emoji, simbol, *mention*, *hashtag*, *retweet*, *url*, data yang duplikat dan karakter lainnya[11], *case folding* bertujuan untuk mengubah data teks dari huruf kapital menjadi data dengan huruf kecil(*lower*)[12], *tokenize* mengubah dari teks data dari kalimat menjadi tiap-tiap kata, *stopword* proses membuang kata yang tidak bermakna dan tidak dibutuhkan, *stemming* proses mengubah kata dari teks yang asalnya memiliki makna kata imbuhan diubah menjadi kata dasar[13].

Translate data melakukan perubahan pada data teks yang awalnya berbahasa Indonesia akan diubah ke dalam bahasa Inggris dengan menggunakan *google spreadsheet*.

Pelabelan data akan menggunakan NLTK (*Natural Language Toolkit*) yang merupakan *library* yang digunakan di dalam bahasa pemrograman *python*, dengan menentukan sentimen positif, netral dan negatif pada data. Sentimen pada data ditentukan berdasarkan *polarity scores* dengan nilai lebih besar dari nol maka masuk ke dalam kategori positif, lebih kecil dari nol kategori negatif dan sama dengan nol kategori netral.

Hasil dari tahap pelabelan data menggunakan NLTK ini akan menjadi informasi sejauh mana pendapat dan komentar masyarakat terkait pemilihan presiden 2024, target utama dalam penelitian yang dilakukan untuk menghasilkan akurasi mana yang terbaik dari ketiga algoritma yang akan dilakukan perbandingan analisis sentimen dengan penggunaan *dataset* pemilihan presiden 2024.

Data yang telah dilakukan pelabelan dalam menentukan sentimen menggunakan NLTK akan dilakukan pembagian data dengan perbandingan 80:20, 80% data latih dan 20% data uji. Fitur yang diterapkan adalah vektor TF-IDF dan kemudian akan dilakukan klasifikasi menggunakan algoritma. Penggunaan rumus persamaan dalam algoritma *Machine Learning Support Vector Machine* persamaan (1), *Naive Bayes* persamaan (2), dan *Decision Tree* persamaan (3)(4) sebagai berikut [14][15][16]:

$$D_{ij} = y_i y_j (K(\vec{x}_i, \vec{x}_j) + \lambda^2) \quad (1)$$

$$P(c|d) \propto P(c) \prod_{k=1}^n P(t_k|c) \quad (2)$$

$$Entropy(S) = \sum_{i=0}^n -p_i * \log^2 p_i \quad (3)$$

$$Gian(S, A) = Entropy(S) = \sum_i^n \frac{|s_i|}{|S|} * Entropy(S_i) \quad (4)$$

4. Hasil dan Pembahasan

Data yang digunakan merupakan *tweet* (postingan) tentang pemilihan presiden 2024 didapat dari hasil proses *Crawling* data dari *twitter* menggunakan *twitter (auth_token)* dengan *search* "Pemilu Presiden 2024" data yang didapatkan sebanyak 38.567 (*tweet*), data yang telah ada akan dilakukan proses *preprocessing* atau pembersihan data agar data bisa digunakan dalam proses pelabelan sentimen dan klasifikasi menggunakan algoritma. Tabel 1 merupakan tampilan data yang masih dalam keadaan mentah yang belum dilakukan pembersihan.

Tabel 1. Sampling Data Hasil *Crawling*

<i>full_text</i>
Forum Silaturahmi Purnawiraaan (FSP) TNI/Polri dan Keluarga Besar ABRI (KBA) menyatakan dukungan terhadap Kepala Staf Kepresidenan (KSP) @Dr_Moeldoko untuk maju sebagai salah satu kandidat pada Pemilihan Presiden (Pilpres) 2024. #MoeldokoPenjagaNKRI https://t.co/fSa1sXbQl0
@abu_waras @PDemokrat @aniesbaswedan @PKSejahtera @NasDem Presiden Indonesia th 2024 Anies dan wakil presiden ahy Cocok memimpin Indonesia Anak muda yg cerdas dan berpengalaman di berbagai daerah Anies pengalaman gubernur jkt yg sukses memimpin dan ahy dari TNI yg punya prestasi komando pasukan pengaman nkri ny Anies&ahy presiden Ri

Pembersihan data ini dilakukan untuk membantu mempermudah dalam proses pelabelan dan klasifikasi algoritma *machine learning*.

Tabel 2. Sampling Hasil Dari Pembersihan Data

Normalisasi Kata
Forum Silaturahmi Purnawiraan (FSP) TNI/Polri dan Keluarga Besar ABRI (KBA) menyatakan dukungan terhadap Kepala Staf Kepresidenan (KSP) @Dr_Moeldoko untuk maju sebagai salah satu kandidat pada Pemilihan Presiden (Pilpres) 2024. #MoeldokoPenjagaNKRI https://t.co/fSa1sXbQI0
Cleaning
Forum Silaturahmi Purnawiraan FSP TNI/Polri dan Keluarga Besar ABRI KBA menyatakan dukungan terhadap Kepala Staf Kepresidenan KSP untuk maju sebagai salah satu kandidat pada Pemilihan Presiden Pilpres
Case Folding
forum silaturahmi purnawiraan fsp tnipolri dan keluarga besar abri kba menyatakan dukungan terhadap kepala staf kepresidenan ksp untuk maju sebagai salah satu kandidat pada pemilihan presiden pilpres
Tokenize
['forum', 'silaturahmi', 'purnawiraan', 'fsp', 'tnipolri', 'dan', 'keluarga', 'besar', 'abri', 'kba', 'menyatakan', 'dukungan', 'terhadap', 'kepala', 'staf', 'kepresidenan', 'ksp', 'untuk', 'maju', 'sebagai', 'salah', 'satu', 'kandidat', 'pada', 'pemilihan', 'presiden', 'pilpres']
Stopword
['forum', 'silaturahmi', 'purnawiraan', 'fsp', 'tnipolri', 'keluarga', 'abri', 'kba', 'dukungan', 'kepala', 'staf', 'kepresidenan', 'ksp', 'maju', 'salah', 'kandidat', 'pemilihan', 'presiden', 'pilpres']
Stemming
forum silaturahmi purnawiraan fsp tnipolri keluarga abri kba dukung kepala staf presiden ksp maju salah kandidat pilih presiden pilpres

Tabel 2 adalah sampling hasil dari tahap *preprocessing* data yang telah dilakukan normalisasi kata dilakukan untuk mengubah teks tidak baku ke kata baku, *cleaning* menghilangkan *mention*, *hashtag*, *url*, emoji, *retweet* dan karakter lainnya, *case folding* mengubah teks ke bentuk *lower*, *tokenize* mengubah bentuk kalimat menjadi tiap kata, *stopword* membuang kata yang tidak memiliki makna dan *stemming* melakukan perubahan pada teks yang memiliki kata imbuhan menjadi kata dasar. Dalam hasil tersebut terlihat beberapa karakter dan kata yang tidak dibutuhkan sudah dihapus. Data yang awal 38.567 *tweet* dilakukan pembersihan untuk menghapus data yang duplikat menghasilkan sisa 7721 *tweet*, data inilah yang digunakan dalam analisis sentimen.

Tabel 3. Sampling Hasil Translate

Translate
friendly forum for retired fsp tnipolri abri kba family supports ksp president's chief of staff to run for wrong candidate to vote for president in presidential election
president of indonesia anies vice president ahy is fit to lead indonesia, intelligent young people, natural area, anies alam, governor of jakarta, successful in leading ahy tni, ahy's achievements in the republic of indonesia's pasu kam aniesamp ahy president ri

Tabel 3 menunjukkan hasil dari *translate* data, proses *translate* data ini menggunakan *google spreadsheet* dengan menggunakan rumus (*GoogleTranslate(cell yang ditranslate,"bahasa digunakan","bahasa tujuan")*).

Hasil *translate* data akan dilakukan proses *labeling* data dengan menggunakan *library* NLTK (*Natural Language Toolkit*). NLTK adalah salah satu *library*(perpustakaan) yang ada di bahasa pemrograman *python*, *library* NLTK menyediakan kumpulan algoritma yang digunakan untuk melakukan tugas dalam pemrosesan bahasa alami. Labeling yang dilakukan akan menentukan sentimen yang memiliki makna positif, netral dan negatif berdasarkan *polarity scores*.

Tabel 4. Sampling Sentimen Positif

<i>Translate</i>	<i>Sentiment</i>
friendly forum for retired fsp tnipolri abri kba family supports ksp president's chief of staff to run for wrong candidate to vote for president in presidential election	Positif
president of indonesia anies vice president ahy is fit to lead indonesia, intelligent young people, natural area, anies alam, governor of jakarta, successful in leading ahy tni, ahy's achievements in the republic of indonesia's pasu kam aniesamp ahy president ri	Positif

Tabel 4 adalah sampel *dataset* yang dikategorikan oleh *library* NLTK ke dalam label sentimen positif. NLTK mengkategorikan kalimat pertama dalam kategori positif karena kata "*friendly forum*" dan "*support*" memiliki bobot kata positif, meskipun di dalam kalimat tersebut ada kata "*wrong*" yang memiliki bobot negatif, tapi kalimat tersebut memiliki kata bobot positif lebih besar dari pada bobot negatif. Begitupun di kalimat kedua bobot positif lebih besar seperti kata "*fit to lead indonesia*" dan "*successful in leading*".

Tabel 5. Sampling Sentimen Netral

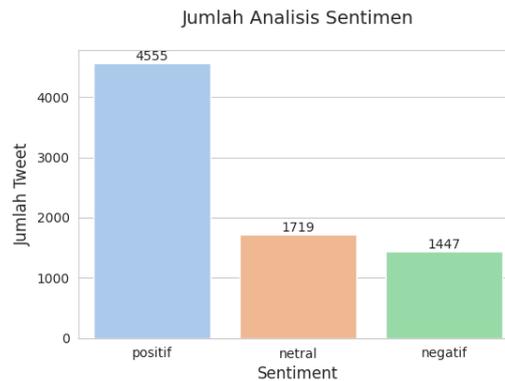
<i>Translate</i>	<i>Sentiment</i>
yess president for	Netral
following the nasdem democrat pks officially nominating anies as presidential candidate	Netral

Tabel 5 adalah *dataset* yang berisi kalimat netral yang dikategorikan oleh *library* NLTK, kalimat pertama dikategorikan netral karena dalam kalimat tersebut singkat dan tidak memiliki kalimat yang mengandung ungkapan secara positif maupun negatif, sedangkan untuk kata "*Yes*" sendiri cenderung netral karena bisa digunakan dalam berbagai konteks. Kalimat yang kedua memiliki arti sama tidak ada ungkapan positif maupun negatif, dalam kalimat tersebut menggambarkan pencalonan anies sebagai kandidat oleh partai-partai.

Tabel 6. Sampling Sentimen Negatif

<i>Translate</i>	<i>Sentiment</i>
dpr pdip pp pkb nasdem pks rejects approval of milu revision completed by legislators regarding urgency	Negatif
president joko widodo jokowi publishes arranging leave for ministers, governors, regents, mayors, and milu regulates orders for ministers' numbers, mayors, and milu resigns	Negatif

Tabel 6 adalah data yang memiliki sentimen negatif, NLTK mengkategorikan kalimat pertama ke dalam negatif karena kata "*rejects*" yang menunjukkan penolakan tentang revisi undang-undang. Sedangkan untuk kalimat kedua, NLTK mengkategorikan kalimat tersebut karena kata "*resigns*" yang menunjukkan pengunduran diri.



Gambar 2. Hasil *Labeling* Data

Gambar 2 merupakan hasil dari keseluruhan data yang telah di *labeling* dalam kategori sentimen menggunakan *library* NLTK(*Natural Language Toolkit*), hasil dari *labeling* tersebut menampilkan jumlah 4555 positif, 1719 netral dan 1447 negatif.

Setelah *labeling* sentimen, data hasil dari sentimen NLTK akan dibagi menjadi dua subset data, berupa data latih dan uji dengan perbandingan 80% data latih 20% data uji, jumlah data latih 6176 dan data uji 1545. Transformasi data teks akan menggunakan *vector* TF-IDF dan kemudian melakukan pelatihan model algoritma *Support Vector Machine*, *Naive Bayes*, dan *Decision Tree*, model yang sudah dilatih akan dilakukan prediksi untuk menentukan hasil *accuracy*, *precision*, *recall* dan *f1-score*. Adapun untuk rumus perhitungan sebagai berikut[17][18].

$$Accuracy = \frac{TP(Kelas\ 1) + TP(Kelas\ 2) + \dots + TP(Kelas\ n)}{Total(Kelas\ 1) + Total(Kelas\ 2) + \dots + Total(Kelas\ n)} \quad (5)$$

$$Precision = \frac{TP(Kelas\ i)}{Prediksi(Kelas\ i)} \quad (6)$$

$$Recall = \frac{TP(Kelas\ i)}{Total(Kelas\ i)} \quad (7)$$

$$F1 - Score = 2 \left(\frac{Recall \times Precision}{Recall + Precision} \right) \quad (8)$$

Persamaan (5) adalah perhitungan dalam mencari hasil nilai *accuracy* seluruh data, persamaan (6), (7) dan (8) adalah perhitungan dalam mencari nilai *precision*, *recall* dan *f1-score* untuk sentimen negatif, netral dan positif.

	precision	recall	f1-score	support
negatif	0.83	0.66	0.74	290
netral	0.76	0.72	0.74	368
positif	0.83	0.90	0.87	887
accuracy			0.81	1545
macro avg	0.81	0.76	0.78	1545
weighted avg	0.81	0.81	0.81	1545
Accuracy:	81.49%			
Precision:	81.41%			
Recall:	81.49%			
F1-score:	81.16%			

Gambar 3. Hasil Klasifikasi *Support Vector Machine*

Gambar 3 menunjukkan hasil dari klasifikasi sentimen menggunakan algoritma *Support Vector Machine* dengan ekstraksi fitur TF-IDF, dengan Data uji yang digunakan berjumlah 1545 menghasilkan hasil *accuracy* 81.49%, *precision* 81.41%, *recall* 81.49% dan *f1-score* 81.16%.

	precision	recall	f1-score	support
negatif	0.95	0.12	0.22	290
netral	0.93	0.10	0.19	368
positif	0.60	1.00	0.75	887
accuracy			0.62	1545
macro avg	0.83	0.41	0.39	1545
weighted avg	0.74	0.62	0.52	1545
Accuracy: 61.94%				
Precision: 74.44%				
Recall: 61.94%				
F1-score: 51.64%				

Gambar 4. Hasil Klasifikasi *Naive Bayes*

Gambar 4 adalah hasil klasifikasi menggunakan algoritma *Naive Bayes* dengan menggunakan perbandingan 80:20 dengan fitur TF-IDF menunjukkan bahwa hasil *accuracy* 61.94%, *precision* 74.44%, *recall* 61.94% dan *f1-score* 51.64%.

	precision	recall	f1-score	support
negatif	0.58	0.47	0.52	290
netral	0.56	0.63	0.59	368
positif	0.77	0.78	0.77	887
accuracy			0.68	1545
macro avg	0.64	0.63	0.63	1545
weighted avg	0.68	0.68	0.68	1545
Accuracy: 68.41%				
Precision: 68.38%				
Recall: 68.41%				
F1-score: 68.25%				

Gambar 5. Hasil Klasifikasi *Decision Tree*

Gambar 5 menunjukkan hasil dari klasifikasi menggunakan *Decision Tree* menggunakan fitur dan perbandingan yang sama dengan kedua algoritma sebelumnya, dari hasil tersebut *Decision Tree* memiliki tingkat *accuracy* 68.41%, *precision* 68.38%, *recall* 68.41% dan *f1-score* 68.25%.

Tabel 7. Sampling Data Perbandingan Hasil Prediksi

<i>Translate</i>	<i>Sentiment</i>	<i>SVM</i>	<i>Naive Bayes</i>	<i>Decision Tree</i>
president jokowi, people are fighting against each other, clashing with each other ahead of the presidential election	Negatif	Positif	Positif	Netral
vice presidential candidate prabowo subiantogibran rakabuming raka milu rot prabowo subianto wins	Positif	Positif	Positif	Positif

Tabel 7 adalah data dengan hasil prediksi menggunakan algoritma *Machine Learning* yang menampilkan ada perbedaan hasil dari prediksi dari setiap algoritma.

Tabel 8. Hasil Perbandingan Algoritma

Algoritma	Accuracy (%)
<i>Support Vector Machine</i>	81.49%
<i>Naive Bayes</i>	61.94%
<i>Decision Tree</i>	68.41%

Tabel 8 adalah hasil dari perbandingan ketiga algoritma *machine learning* dalam melakukan analisis sentimen menggunakan *dataset* tentang Pemilu Presiden 2024. Berdasarkan tabel 8, penelitian yang dilakukan menghasilkan akurasi tertinggi adalah algoritma *Support Vector Machine* dengan nilai 81.49% yang dikategorikan sangat baik dalam melakukan analisis sentimen, dengan hasil dari perbandingan ini disimpulkan bahwa algoritma SVM pada penelitian ini mendukung hasil dari penelitian terdahulu[7], sedangkan algoritma *Decision Tree* pada penelitian ini menunjukkan nilai akurasi 68.41% dan *Naive Bayes* 61.94% menunjukkan bahwa akurasi *Decision Tree* lebih tinggi dibandingkan *Naive Bayes* yang mendukung pada penelitian[9] meskipun dengan perbedaan hasil nilai akurasi dengan penelitian terdahulu berbeda.

5. Kesimpulan

Penelitian ini menghasilkan analisis sentimen berupa *Support Vector Machine* sebagai algoritma dengan tingkat *accuracy* tertinggi dengan nilai 81.49%, kemudian diikuti algoritma *Decision Tree* dengan 68.41% dan yang terakhir *Naive Bayes* dengan 61.94%, untuk hasil sentimen pada penelitian ini menunjukkan sebanyak 4555 positif, 1719 netral dan 1447 negatif. Saran untuk penelitian selanjutnya; 1) melakukan perbandingan antara algoritma *machine learning* dan *deep learning*. 2) menggunakan *dataset* yang lebih banyak dan melakukan *preprocessing* dengan lebih akurat. 3) menggunakan opsi lain dalam melakukan *translate* data.

DAFTAR PUSTAKA

- [1] N. A. Kurniawati, A. Karim, and Ghufron, "RUANG PUBLIK VIRTUAL TWITTER PADA AKUN @ Pemkotbpn SEBAGAI ALAT KONTROL PUBLIK," *eJournal Ilmu Komun.*, vol. 6, no. 3, pp. 333–347, 2018.
- [2] R. Vindua and A. U. Zailani, "Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python," *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 2, pp. 479–487, 2023, doi: 10.30865/jurikom.v10i2.5945.
- [3] O. Manullang and C. Prianto, "Analisis Sentimen dalam Memprediksi Hasil Pemilu Presiden dan Wakil Presiden : Systematic Literature Review," *J. Inform. dan Teknol. Komput.*, vol. 04, no. 02, pp. 104–113, 2023.
- [4] H. Utami, "Analisis Sentimen dari Aplikasi Shopee Indonesia Menggunakan Metode Recurrent Neural Network," *Indones. J. Appl. Stat.*, vol. 5, no. 1, pp. 31–38, 2022, doi: 10.13057/ijas.v5i1.56825.
- [5] R. R. Sani, Y. A. Pratiwi, S. Winarno, E. D. Udayanti, and F. AlZami, "Analisis Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Hoax pada Berita Online Indonesia," *J. Masy. Inform.*, vol. 13, no. 2, pp. 85–98, 2022, doi: 10.14710/jmasif.13.2.47983.
- [6] R. A. Putri and N. S. Fatonah, "Perbandingan Metode Klasifikasi serta Analisis Faktor Akademis Pola Kelulusan Mahasiswa di Perguruan Tinggi," *J. Inform. J. Pengemb. IT*, vol. 7, no. 2, pp. 109–117, 2022, doi: 10.30591/jpit.v7i2.3082.
- [7] M. A. Saddam, E. Kurniawan D, and Indra, "Analisis Sentimen Fenomena PHK Massal Menggunakan Naive Bayes dan Support Vector Machine," *J. Inform. J. Pengemb. IT*, vol. 8, no. 3, pp. 226–233, 2023, doi: 10.30591/jpit.v8i3.4884.
- [8] C. A. A. Soemedhy, N. Trivetisia, N. A. Winanti, D. P. Martyaningsih, T. W. Utami, and S. Sudianto, "Analisis Komparasi Algoritma Machine Learning untuk Sentiment Analysis (Studi Kasus: Komentar YouTube 'Kekerasan Seksual')," *J. Inform. J. Pengemb. IT*, vol. 7, no. 2, pp. 80–84, 2022, doi: 10.30591/jpit.v7i2.3547.
- [9] M. Syarifuddin, "Analisis Sentimen Opini Publik Terhadap Efek Psbb Pada Twitter Dengan Algoritma Decision Tree-Knn-Naive Bayes," *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 87–94,

- 2020, doi: 10.33480/inti.v15i1.1433.
- [10] M. Nashrullah, O. Maharani, A. Rohman, E. F. Fahyuni, Nurdyansyah, and R. S. Untari, *Metodologi Penelitian Pendidikan (Prosedur Penelitian, Subyek Penelitian, Dan Pengembangan Teknik Pengumpulan Data)*. Sidoarjo: UMSIDA Press, 2023. doi: 10.21070/2023/978-623-464-071-7.
- [11] R. Ramlan, N. Satyahadewi, and W. Andani, "Analisis Sentimen Pengguna Twitter Menggunakan Support Vector Machine Pada Kasus Kenaikan Harga BBM," *Jambura J. Math.*, vol. 5, no. 2, pp. 431–445, 2023, doi: 10.34312/jjom.v5i2.20860.
- [12] S. Khairunnisa, Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, pp. 406–414, 2021, doi: 10.30865/mib.v5i2.2835.
- [13] I. S. Thalib, S. K. Gusti, F. Yanto, and M. Affandes, "Klasifikasi Sentimen Tragedi Kanjuruhan Pada Twitter Menggunakan Algoritma Naïve Bayes," *J. Sist. Komput. dan Inform.*, vol. 4, no. 3, pp. 467–473, 2023, doi: 10.30865/json.v4i3.5852.
- [14] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *J. Media Inform. Budidarma*, vol. 4, no. 3, pp. 650–658, 2020, doi: 10.30865/mib.v4i3.2181.
- [15] A. Sabrani, I. G. W. W. W., and F. Bimantoro, "Metode Multinomial Naïve Bayes Untuk Klasifikasi Artikel Online Tentang Gempa Di Indonesia," *J. Teknol. Informasi, Komputer, dan Apl. (JTika)*, vol. 2, no. 1, pp. 89–100, 2020, doi: 10.29303/jtika.v2i1.87.
- [16] C. Cahyaningtyas, Y. Nataliani, and I. R. Widiyari, "Analisis Sentimen Pada Rating Aplikasi Shopee Menggunakan Metode Decision Tree Berbasis SMOTE," *Aiti*, vol. 18, no. 2, pp. 173–184, 2021, doi: 10.24246/aiti.v18i2.173-184.
- [17] A. Fitriadin and A. S. Purnomo, "Analisis Sentimen Masyarakat Terhadap Pandemi Covid-19 Pada Sosial Media Menggunakan Naïve Bayes Clasifier," *INFORMAL Informatics J.*, vol. 8, no. 1, pp. 51–57, 2023, doi: 10.19184/isj.v8i1.33937.
- [18] M. M. Khoirudin, Wiranto, and Winarno, "News Opinion Mining around Universitas Sebelas Maret Using Naive Bayes Algorithm," *ITSMART J. Ilm. Teknol. dan Inf.*, vol. 7, no. 1, pp. 44–50, 2018.