

## Optimasi Teknik Voting pada Sentimen Analisis Pemilihan Presiden 2024 Menggunakan *Machine Learning*

Kharisma Rahayu<sup>1</sup>, M.Khairul Anam<sup>2</sup>, Lusiana Efrizoni<sup>3</sup>, Nurjayadi<sup>4</sup>, Triyani Arita Fitri<sup>5</sup>

rahayukharisma716@gmail.com, khairulanam@sar.ac.id, lusiana@stmik-amikriau.ac.id,

nurjayadi@sar.ac.id, triyani@sar.ac.id

<sup>1,3,4,5</sup> Universitas Sains dan Teknologi Indonesia

<sup>2</sup> Universitas Samudra

---

### Informasi Artikel

Diterima : 11 Jun 2024

Direview : 20 Jun 2024

Disetujui : 25 Jul 2024

---

### Kata Kunci

Pemilihan presiden, *Machine Learning*, *Syntethic Minority Over-Sampling* (SMOTE), *Ensamble*, *Voting*

---

### Abstrak

Pemilihan presiden adalah peristiwa penting dalam sistem demokrasi Negara Kesatuan Republik Indonesia atau NKRI yang diadakan setiap lima tahun. Pro dan kontra pilpres banyak terjadi, khususnya di media sosial Twitter atau X. X adalah salah satu platform media dimana orang meninggalkan komentar positif, netral, bahkan negatif. Oleh karena itu penelitian ini bertujuan untuk membangun model analisis sentimen untuk melakukan klasifikasi sentimen pilpres 2024. Penelitian ini menggunakan algoritma *Support Vector machine*, *Naïve Bayes* dan *Decision Tree* dalam melakukan klasifikasi dengan penambahan metode *Syntethic Minority Over-Sampling* dan *Ensemble Voting*. Hasil pengujian menunjukkan bahwa sentimen masyarakat terhadap pilpres mendominasi sentimen negatif sebesar 5008 positif 3582 dan netral 1411 sentimen. Kemudian hasil pemrosesan data algoritma SVM, NB dan DT ditambah SMOTE dan optimasi *ensemble voting*, memberikan akurasi 92,8%, presisi 93%, *recall* 93% dan *F1-Score* 93%. Penelitian ini dapat memberikan kontribusi signifikan dengan mengklasifikasikan sentimen masyarakat terhadap data pilpres 2024.

---

### Keywords

*Presidential election*, *Machine Learning*, *Syntethic Minority Over-Sampling* (SMOTE), *Ensamble*, *Voting*

---

### Abstract

*The presidential election is an important event in the democratic system of the Republic of Indonesia or NKRI held every five years. There are many pros and cons of the presidential election, especially on social media Twitter or X. X is one of the media platforms where people leave positive, neutral, and even negative comments. Therefore, this research aims to build a sentiment analysis model to classify the sentiment of the 2024 presidential election. This research uses the Support Vector machine, Naïve Bayes and Decision Tree algorithms in performing classification with the addition of the Syntethic Minority Over-Sampling and Ensemble Voting methods. The test results show that public sentiment towards the presidential election dominates negative sentiment of 5008 positive 3582 and neutral 1411 sentiments. Then the results of data processing SVM, NB and DT algorithms plus SMOTE and ensemble voting optimization, provide 92.8% accuracy, 93% precision, 93% recall and 93% F1-Score. This research can make a significant contribution by classifying public sentiment towards the 2024 presidential election data.*

## A. Pendahuluan

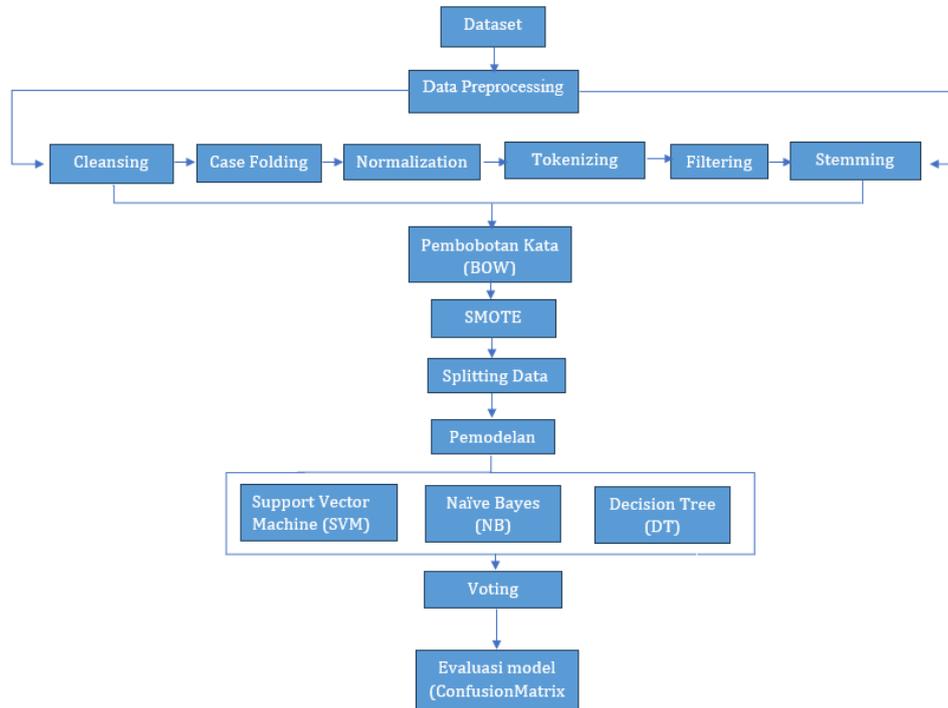
Demokrasi adalah jenis pemerintahan di mana setiap warga memiliki hak yang sama untuk membuat keputusan yang dapat mengubah kehidupan mereka [1]. Sistem demokrasi ini ditandai dengan mengadakan pemilihan umum serentak pada tahun 2024 untuk memilih presiden serta wakilnya atau pemilihan presiden (Pilpres), dan pemilihan legislatif (Pileg). Banyak respon masyarakat Indonesia terhadap padangan politik menuai opini pro dan kontra yang dapat dilihat pada media sosial [16]. Komentar pada media sosial dapat bersifat positif, negatif ataupun netral. Salah satu media sosial yang paling populer di masyarakat adalah media sosial twitter atau X. Di X, komentar tentang pemilihan presiden 2024 sangat memanas, sehingga banyak pola komunikasi netizen maupun opini masyarakat terhadap pemilihan presiden 2024, untuk melihat pola komunikasi tersebut dilakukanlah analisa sentimen.

Analisis sentimen adalah mengelompokkan polaritas dalam teks yang telah dilabeli untuk mengekstrak informasi tentang sikap seseorang terhadap masalah. Metode ini digunakan untuk mengetahui, melihat, dan membuat kesimpulan apakah teks tersebut bersifat positif, negatif, atau netral. Dengan menggunakan analisis sentimen, kita dapat mengetahui pendapat masyarakat tentang masalah yang terjadi [3]. Analisis sentimen semakin populer dan banyak dibahas dalam berbagai jurnal penelitian, salah satunya adalah penelitian yang dilakukan oleh [4]. Penelitian ini membahas analisis sentimen calon presiden 2024, dengan tujuan mengetahui kecenderungan komentar masyarakat terhadap calon presiden, seberapa banyak tanggapan positif dan negatif menggunakan algoritma *Support Vector Machine* (SVM), *Naïve Bayes*, dan *K-Nearest Neighbors* (KNN). Hasil penelitian algoritma SVM memberikan kinerja superior akurasi total sebesar 0.88. Selanjutnya penelitian yang dilakukan oleh [5] membahas opini masyarakat mengenai efek PSBB menggunakan algoritma *Decision Tree*, *K-Nearest Neighbors* (K-NN), dan *Naïve Bayes*. Hasil tertinggi dari penelitian ini adalah *Decision Tree* dengan nilai *accuracy* 83,3%. Kemudian pada penelitian [6] membahas analisis sentimen pembelajaran daring pada twitter di masa pandemi COVID-19 menggunakan metode *naïve bayes* dan didapatkan hasil bahwa algoritma *naïve bayes* memiliki akurasi yang sangat baik sebesar 97,15%. Dari penelitian sebelumnya mengenai analisis sentimen algoritma SVM, NB, dan DT memiliki akurasi tertinggi, sehingga algoritma tersebut sangat tepat karena nilai hasil yang bagus, namun terkadang algoritma *based*, terkadang masih mendapatkan akurasi yang cukup rendah, untuk itu perlunya peningkatan menggunakan metode lain seperti metode *ensemble learning* yaitu voting.

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, maka dilakukanlah analisis sentimen yang bertujuan untuk melihat pola komunikasi di masyarakat terhadap pemilihan presiden 2024. Penelitian ini menambahkan metode *ensemble* yaitu voting guna melihat seberapa optimal meningkatkan kinerja model. hasil penelitian ini diharapkan dapat membantu masyarakat mengetahui tanggapan masyarakat terhadap pilpres 2024. Penelitian ini diharapkan dapat memberikan kontribusi yang relevan dalam bidang pemodelan sentimen.

## B. Metode Penelitian

Metodologi penelitian dijadikan panduan untuk menentukan tahapan penelitian. Tahap-tahap analisis sentimen terhadap pemilihan presiden 2024 seperti disajikan pada gambar 1.



**Gambar 1.** Alur Penelitian

### Pengambilan Data

Pengumpulan datatset yang digunakan pada penelitian ini adalah data *public* yang diambil menggunakan aplikasi *Drone Empirit Academy* (DEA). Data yang diambil menggunakan kata kunci “Pilpres2024”. Dataset terdiri dari 5 atribut (i.e., *Datetime*, *Tweet\_Id*, *Tweet*, *Username*, dan *Sentimen*) dan 10.001 *tweet*.

### Data Preprocessing

*Preprocessing* adalah salah satu proses data mentah. Tahap ini menghapus data yang bernilai null atau kosong [7]. Data *preprocessing* bertujuan agar data yang diproses lebih terstruktur dan proses pemodelan menjadi lebih mudah[8]. Dalam penelitian ini dilakukan 6 tahapan *Preprocessing* yaitu *data cleaning*, *case folding*, *text normalization*, *tokenizing*, *filtering* dan *stemming*.

1. *Data Cleaning*, untuk membersihkan tweet dari simbol, angka, URL, hashtag. Contohnya menghilangkan tanda koma (,)
2. *Case folding* adalah proses dalam pemrosesan teks di mana semua huruf dalam teks dikonversi menjadi format huruf kecil (*lowercase*) atau huruf besar (*uppercase*) untuk meminimalkan variasi yang disebabkan oleh perbedaan huruf besar dan kecil.

3. *Text Normalisasi* bertujuan agar pembaca dapat memahami arti kata yang dimaksudkan secara jelas, begitu juga dengan sistem ketika melakukan analisis klasifikasi.
4. *Tokenizing* adalah proses memecah kalimat per-kata/token [9]
5. *Filtering* adalah Lanjutan dari tahapan tokenizing adalah tahapan filtering yang digunakan untuk mengambil kata-kata yang penting dari hasil token tadi. Kata umum yang biasanya muncul dan tidak memiliki makna disebut dengan *stopword*.
6. *Stemming* merupakan proses dalam mentransformasi kata menjadi kata dasar dengan menghilangkan imbuhan, sehingga bisa meminimalkan dimensi kata yang berbeda bentuk namun dengan makna yang sama [9].

### **Pembobotan Kata**

Pembobotan kata adalah suatu mekanisme untuk memberikan skor kepada kata-kata tertentu dalam dokumen teks. Sedangkan menurut [10] adalah proses mengubah kata menjadi vektor kata (representasi numerik). Pada penelitian ini menggunakan teknik pembobotan kata *Bag Of Words* (BoW). BoW yaitu menghitung frekuensi kemunculan kata pada seluruh dokumen. Pada tahap ini metode BoW hanya menghitung berapa kali kata muncul di seluruh dokumen yang diproses.

1. Langkah pertama adalah membuat sekumpulan kata atau kosa kata dari seluruh dokumen yang akan diproses.
2. Menghitung frekuensi kata yang muncul dari setiap dokumen.
3. Masukkan hasil perhitungan frekuensi kata ke dalam vektor.
4. Vektor tersebut menunjukkan dokumen BOW.

### **Penyeimbangan Data (SMOTE)**

Penelitian ini menggunakan teknik penyeimbangan data yaitu (*Synthetic Minority Oversampling Technique*) SMOTE. SMOTE bekerja dengan menyeimbangkan kelas minoritas dengan membuat sintesis data kelas minoritas. Di penelitian ini metode *oversampling* lebih cocok digunakan dibandingkan *undersampling* karena dataset pemilihan presiden keseluruhannya hanya berjumlah 10.001. *Oversampling* menyimpan seluruh informasi dari data mentah sementara *undersampling* mengharuskan penghapusan sebagian informasi dari data mentah. Informasi tersebut memiliki peran penting dalam pemodelan klasifikasi walaupun berasal dari kelas mayoritas. Jadi secara umum, untuk dataset yang berjumlah sedikit *oversampling* lebih cocok dibandingkan *undersampling*.

### **Splitting Data**

*Splitting data* adalah memisahkan dataset menjadi data uji dan data latih. Penelitian ini menggunakan *splitting* data 70:30, Data latih digunakan untuk melatih algoritma *machine learning* agar algoritma dapat belajar dari data yang tersedia. Data uji digunakan untuk menguji algoritma *machine learning* tersebut yang menghasilkan tingkat akurasi atau skor dari algoritma yang digunakan.

### **Pemodelan**

Tahap berikutnya adalah pemodelan *algoritma machine* menggunakan menggunakan *Support Vector Machine* (SVM), *Naive Bayes* (NB), dan *Decision Tree*

(DT) untuk membangun klasifikasi penelitian ini menggunakan SVM *Linear*, NB *Complement*, dan DT *C4.5*. Ketiga algoritma tersebut juga nantinya akan ditambahkan SMOTE dan metode *ansamble learning* yaitu soft voting. Sehingga nantinya akan diketahui algoritma yang paling optimal untuk mendapatkan akurasi terbaik dari beberapa uji coba yang dilakukan.

### **Support Vector Machine**

*Support Vector Machine* (SVM) adalah seperangkat metode pembelajaran terbimbing (*supervised learning*) yang menganalisis data dan mengenali pola, digunakan untuk klasifikasi dan analisis regresi [11] Prinsip utama SVM adalah mencari fungsi *hyperplane*, atau garis pemisah, yang dapat memisahkan kedua kelas secara maksimal.

Pada Penelitian ini menggunakan kernel SVM *linear* dimana fungsi kernel yang digunakan dalam SVM pada penelitian ini dengan kernel *linier*. Fungsi kernel ini didefinisikan sebagai berikut

$$\text{Kernel Linear: } x^t x \quad (1)$$

$x$ , yang merupakan produk dot dari dua vektor fitur.

### **Naive Bayes**

Berdasarkan teorema *Bayes*, *Naive Bayes Classifier* menggunakan metode probabilitas dan statistik untuk memprediksi peluang di masa depan berdasarkan pengalaman masa lalu. Salah satu ciri utama dari *Naive Bayes Classifier* ini adalah keyakinan yang sangat kuat (naif) bahwa setiap kondisi atau peristiwa sendiri. Setiap kelas keputusan, *Naive Bayes* menghitung probabilitas dengan asumsi bahwa kelas keputusan adalah benar, mengingat vektor informasi objek, menurut [12]. Algoritma ini menganggap atribut objek independen.

Pada penelitian ini menggunakan NB *Multinomial Multinomial Naive Bayes* adalah salah satu varian algoritma *Naive Bayes* yang biasanya digunakan untuk klasifikasi data dan teks yang diwakili sebagai distribusi frekuensi kata-kata. Algoritma ini menganggap distribusi multinomial sebagai fitur yang digunakan untuk klasifikasi.

Rumus umum untuk *naive Bayes* adalah:

$$P(c|d) = \frac{P(d|c) - P(c)}{P(d)} \quad (2)$$

### **Decision Tree**

Salah satu metode klasifikasi yang menggunakan representasi struktur pohon (*tree*) adalah *Decision tree*. Dalam metode ini, setiap node menunjukkan atribut, cabang menunjukkan nilai atribut, dan daun menunjukkan kelas. Noda yang paling atas dari pohon ini disebut *root*. Hasil dari model ini mudah dipahami dan dibangun dengan cepat. Terdapat tiga jenis node pada *decision tree*: *root node*, *internal node*, dan *leaf node* [13].

Pada penelitian ini menggunakan *decision tree* C4.5. DT C4.5 adalah algoritma pembuatan pohon keputusan yang digunakan untuk klasifikasi. Dikembangkan oleh Ross Quinlan, algoritma ini berfokus pada pembuatan pohon yang optimal dengan menggunakan metode informasi dan pengukuran entropi (C4.5). C4.5 menggunakan

rumus pengukuran entropi dan informasi untuk menentukan fitur mana yang memberikan pemisahan kelas yang paling baik. Pengukuran entropi dan informasi melibatkan perhitungan probabilitas dan distribusi kelas pada setiap simpul pohon.

### Voting

Voting adalah metode *ensambel* yang menggabungkan performa beberapa model untuk membuat keputusan akhir yang lebih akurat dibandingkan dengan prediksi dari satu model tunggal. Salah satu metode kelompok, *voting classifier* memungkinkan penggabungan berbagai model untuk menghasilkan klasifikasi yang lebih baik [14]. Voting atau Pemungutan suara diklasifikasikan menjadi "*Hard voting* (keras)" dan "*Soft Voting* (lunak)" menurut pengklasifikasi kelompok.

### Evaluasi Model

Tahap berikutnya ialah evaluasi model algoritma dimana menguraikan perbandingan kinerja berbagai model yang telah diuji dalam penelitian. Disini dapat melihat algoritma mana yg terbaik dan dapat melihat peningkatan algoritma pada data pilpres 2024 yang digunakan. penelitian ini menggunakan *confusion matrix* sebagai evaluasi. *confusion matrix* yang akan mengetahui tingkat akurasi. Semakin tinggi nilai akurasi, semakin baik pula model yang dihasilkan. "Evaluasi dengan *Confusion matrix* menghasilkan nilai akurasi, presisi, *recall* dan *f1-score*" [15].

#### 1. Akurasi.

Akurasi adalah jumlah rasio klasifikasi yang benar dari data. Rumus akurasi adalah sebagai berikut:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

#### 2. Recall.

Recall Sebuah matriks evaluasi yang digunakan untuk mengevaluasi seberapa efektif model klasifikasi menganalisis kasus positif yang sebenarnya. Recall rumusnya adalah sebagai berikut:

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

#### 3. Presisi.

Presisi adalah matriks evaluasi yang digunakan untuk mengukur keberhasilan model klasifikasi dalam mengidentifikasi instance positif dengan akurat dari semua instance yang bernilai positif. Presisi rumusnya adalah sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

#### 4. F1-Score.

F1-Score adalah ukuran untuk evaluasi pembelajaran mesin yang menggabungkan skor perolehan dan presisi. Rumusnya adalah sebagai berikut:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (6)$$

### C. Hasil dan Pembahasan

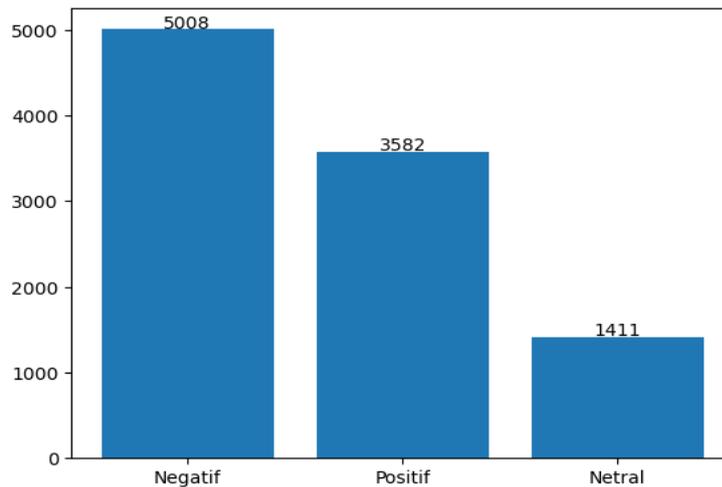
#### Pengambilan Data

Pada penelitian ini data diambil menggunakan aplikasi *Drone Empirit Academy* (DEA). Data yang diambil menggunakan kata kunci "Pilpres2024". Data yang diambil berjumlah 10.001 *tweet* dengan 5 variabel yaitu *Datetime*, *Tweet\_id*, *Tweet*, *Username* dan Sentimen. Data diambil dalam bentuk data frame dan disimpan dengan format excel sebagai 'pemilu.excel', disajikan dalam tabel dibawah ini.

**Tabel 1** Hasil pengambilan dataset

	Datetime	Tweet_id	Tweet	User name	Sentimen
0	2023-03-30 23:58:53+00:00	1.641591e+ 18	Kami menolak, capres pengusung khalifah. Presid....	presiden pribumi	Positif
1	2023-03-30 23:40:00+00:00	1.641586e+ 18	Diapresiasi Kemenkes RI !! Tuntaskan Problem B..	harian ganjar	Negatif
2	2023-03-30 23:28:56+00:00	1.641583e+ 18	Indonesia menolak khalifah. Presiden Republik....	presiden pribumi	Positif
3	2023-03-30 23:00:01+00:00	1.641576e+ 18	Kejar Target Open Defaction Fee (ODF) di 2023...	harian ganjar	Negatif
4	2023-03-30 23:00:00+00:00	1.641576e+ 18	"itu kan keputusan politik luar negri kita ya..	catatan ganjar	Negatif
...	...	...	...	...	...
9996	2023-03-30 10:10:53+00:00	1.622176e+ 18	Apa yang menjadi keluhan rakyat saat ini, mak..	ilham fadli	Negatif

Kolom *Datetime*, *Tweet\_id*, *Username* dihapus karna yang diambil hanya data pada kolom *Tweet* dan *Sentimen*. Data diatas memiliki 3 label sentimen yaitu negatif, positif dan netral.



**Gambar 2** Bar Chart Label Negatif, Netral dan Positif

Gambar 2 menunjukkan bahwa Negatif (5008), Netral (1411) dan Positif (3582). Terdapat ketidakseimbangan pada label data yang digunakan sehingga akan dilakukan metode *Synthetic Minority Over-Sampling Technique* (SMOTE) untuk menanggulangnya. Namun sebelum itu data akan dilakukan tahap *pre-processing*.

**Pre-Processing**

Tujuan dari proses *pre-processing* data adalah untuk meningkatkan kualitas data. Teknik yang digunakan dalam proses *pre-processing* termasuk menghapus data yang hilang atau duplikat, mengisi nilai yang hilang, menormalisasi data, dan mengubah format data ke format yang lebih standar. Berikut lanjutan dari data *pre-processing* yang dilakukan.

**Tabel 2** Tahap *Preprocessing*

Tahap	Sebelum	Sesudah
<i>Data Cleaning</i>	Kami menolak capres pengusung khalifah. Presiden Republik Indonesia adalah orang Indonesia asli #presidenpribumi#pribumi #Pilpres2024	Kami menolak capres pengusung khalifah. Presiden Republik Indonesia adalah orang Indonesia asli
<i>Case Folding</i>	Kami menolak capres pengusung khalifah. Presiden Republik Indonesia adalah orang Indonesia asli	kami menolak capres pengusung khalifah. presiden republik indonesia adalah orang indonesia asli
<i>Text Normalisasi</i>	kami menolak capres pengusung khalifah. presiden republik indonesia adalah orang indonesia asli	kami menolak capres pengusung khalifah presiden republik indonesia adalah orang indonesia asli

<i>Tokenizing</i>	kami menolak capres pengusung khalifah. presiden republik indonesia adalah orang indonesia asli	[kami, menolak, capres, pengusung, khalifah, presiden, republik, indonesia, adalah, orang, indonesia, asli]
<i>Filtering</i>	[kami, menolak, capres, pengusung, khalifah, presiden, republik, indonesia, adalah, orang, indonesia, asli]	menolak, capres, pengusung, khalifah, presiden, republik, indonesia, adalah, orang, indonesia, asli
<i>Stemming</i>	menolak, capres, pengusung, khalifah, presiden, republik, indonesia, adalah, orang, indonesia, asli	tolak capres usung khalifah presiden, republik indonesia asli

**Pembobotan Kata**

Tujuan pembobotan kata ini adalah untuk mengubah kata dalam data menjadi angka sehingga algoritma dapat memprosesnya. Studi ini menggunakan pembobotan kata BoW dengan *CountVectorizer* di library sklearn. Setiap kata yang berulang ditulis sekali saat menggunakan korpus yang ada.

	aa	aaaaaaaa	aaaaaaaaah	aaahhhh	aahh	aamiin	abad	abai	abdi	abdullah	...	yupi	yusri	zainul	zakat	zaman	ziva	zodiak	zonk	zulhas	zulkifli	
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
9996	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
9997	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
9998	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
9999	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
10000	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0

**Gambar 3.** Hasil perhitungan *Bag of Word*

Keterangan:

Menghitung frekuensi setiap kemunculan kata pada korpus tersebut pada data diatas, Jika kata tersebut muncul maka diberi nilai satu sebaliknya jika tidak muncul maka diberi nilai 0.

**Penyeimbangan Data**

Proses selanjutnya, variabel data X dan y didefinisikan. Pada tahap ini, variabel y diubah menjadi kategorikal dengan bantuan library imbalanced-learn dan variabel x yang sebelumnya satu dimensi diubah menjadi dua dimensi. Karena yang digunakan terdapat ketidakseimbangan, dilakukan teknik oversampling menggunakan metode *Synthetic Minority Over-Sampling Technique* (SMOTE) untuk menyeimbangkan label pada data.

Jumlah sampel sebelum oversampling: 10001  
 Jumlah sampel setelah oversampling: 15024

**Gambar 4** Hasil *Synthetic Minority Over-Sampling Technique* (SMOTE)

Dapat dilihat pada gambar 4 bahwa setelah penggunaan metode *Synthetic Minority Over-Sampling Technique* (SMOTE), data yang berjumlah sebanyak 10001 menjadi sebanyak 15024 data. Sehingga untuk proses selanjutnya data yang digunakan adalah sebanyak 15024 data.

### **Splitting Data**

Tahap selanjutnya adalah pembagian data menjadi data latih dan data uji dengan menggunakan *splitting* data. *Splitting* data yang digunakan adalah 70% data latih dan 30% data uji dan didapatkan sebanyak 10516 data latih dan 4508 data uji.

### **Pemodelan**

Algoritma yang digunakan dalam penelitian ini *Support Vector Machine* (SVM), *Naive Bayes* (NB), *Decision Tree* (DT) dan optimasi *ensemble learning soft voting*. *Confusion matrix* digunakan untuk mengevaluasi kinerja model yang dibuat oleh setiap algoritma klasifikasi.

Berikut tabel perbandingan algoritma yang digunakan pada penelitian ini.

**Tabel 3.** Perbandingan *Algoritma* SVM, NB dan DT

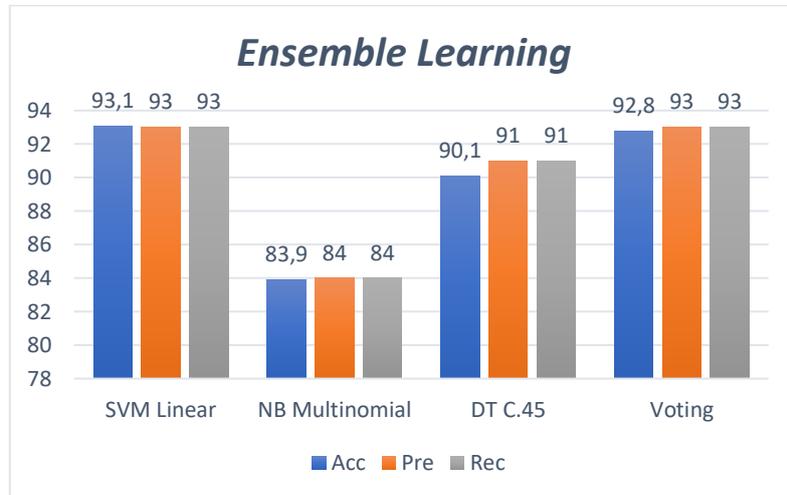
Algoritma	Hasil Evaluasi		
	Acc (%)	Pre (%)	Rec (%)
SVM	93,1%	90%	90%
NB	83,9%	84%	84%
DT	90,1%	91%	91%

Berdasarkan Tabel 3, proses *learning* yang dilakukan pada data sentimen pemilihan presiden 2024 menggunakan algoritma SVM, NB, dan DT menunjukkan bahwa penerapan SVM menghasilkan nilai akurasi terbaik sebesar 93,1% atau 93%. Dalam penelitian ini, nilai akurasi 93% dianggap sangat baik. Selain itu, hasil dari model yang menggunakan tambahan *ensemble learning soft voting* dapat dilihat dalam Tabel 4 berikut, yang menunjukkan peningkatan performa model.

**Tabel 4** Perbandingan *Algoritma* SVM, NB dan DT + Soft Voting

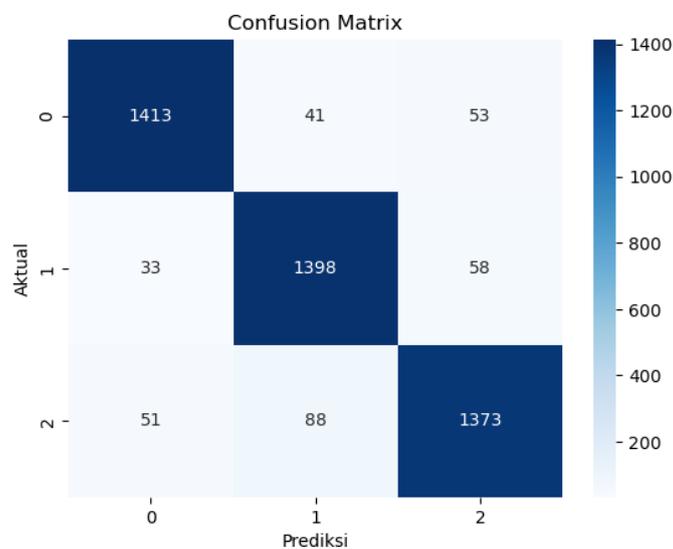
Algoritma	Hasil Evaluasi		
	Acc (%)	Pre (%)	Rec (%)
SVM	93,1%	90%	90%
NB	83,9%	84%	84%
DT	90,1%	91%	91%
Voting	92,8%	93%	93%

Berdasarkan Tabel 4, penelitian ini menunjukkan bahwa kombinasi algoritma yang dioptimalkan dengan soft voting mencapai nilai akurasi yang tinggi. Hasil ensemble learning menggunakan tiga algoritma tunggal SVM, NB, dan DT menghasilkan akurasi sebesar 92,8% atau 93%. Visualisasi keseluruhan algoritma yang digunakan dapat dilihat pada Gambar 5, yang memberikan gambaran lebih jelas tentang performa dari kombinasi algoritma tersebut.



**Gambar 5.** Visualisasi Perbandingan SVM, NB, DT + Voting

Evaluasi akurasi mengukur seberapa baik model cocok dengan data secara keseluruhan. Hasil menunjukkan bahwa model *ensemble soft voting* memiliki nilai akurasi yang baik. Meskipun penerapan voting menurun sedikit terhadap SVM dengan selisih 1%, kombinasi algoritma lainnya mengalami peningkatan akurasi. Ini menunjukkan bahwa model mampu mengklasifikasikan data dengan benar secara keseluruhan. Setelah menguji masing-masing algoritma, hasil tersebut disajikan dalam *confusion matrix* yang dapat dilihat pada Gambar 6.



**Gambar 6** Confusion matrix ensemble learning voting

Gambar diatas menunjukkan *confusion matrix* dari *ansamble voting* algoritma SVM, NB, dan DT. Dari hasil *confusion matrix* diatas, dapat dilakukan perhitungan akurasi, presisi, *recall*, dan *F1 Score*.

Berdasarkan hasil perhitungan tersebut dapat dibuktikan pada laporan klasifikasi yang disajikan pada Gambar 7.

	precision	recall	f1-score	support
Negatif	0.94	0.94	0.94	1507
Netral	0.92	0.94	0.93	1489
Positif	0.93	0.91	0.92	1512
accuracy			0.93	4508
macro avg	0.93	0.93	0.93	4508
weighted avg	0.93	0.93	0.93	4508

**Gambar 7** Laporan Klasifikasi

Nilai akurasi voting lebih rendah daripada SVM dikarenakan pengaruh NB yang lebih rendah, (NB) memiliki akurasi yang cukup rendah (83.9%), dan kontribusi model yang kurang dapat mempengaruhi kinerja keseluruhan kelompok akan menurunkan keakuratan voting.

#### D. Simpulan

Hasil analisis sentimen masyarakat terhadap pemilihan presiden mendominasi sentimen negatif sebesar 5008 diikuti dengan sentimen positif 3582 dan netral 1411 sentimen. Kemudian untuk hasil pengujian menunjukkan bahwa algoritma *machine learning* SVM, NB dan DT dengan penambahan SMOTE dan Voting sangat berpotensi untuk menyelesaikan masalah klasifikasi pada dataset yang digunakan pada penelitian ini. Akurasi yang sangat baik diberikan oleh masing-masing model, Berdasarkan perhitungan tersebut dapat disimpulkan nilai akurasi tertinggi pada algoritma tunggal adalah pada SVM sebesar 93,1 % diikuti dengan DT 90,1% dan NB 83,9%, kemudian pada penerapan metode voting didapatkan hasil 92,8%.

Kesimpulan bahwa penggunaan model *ensemble* dalam klasifikasi sentimen masyarakat tidak selalu berhasil meningkatkan performa algoritma. Hal ini disebabkan oleh pengaruh salah satu algoritma yang memiliki tingkat akurasi yang rendah, yang dapat menurunkan kinerja keseluruhan model *ensemble*. Penelitian ini memberikan wawasan penting mengenai penggunaan algoritma *machine learning* dalam klasifikasi data, dan menyarankan agar peneliti selanjutnya menerapkan teknik optimasi lain yang dapat dikombinasikan dengan metode voting pada algoritma yang ingin digunakan dalam penelitian. Selain itu, peneliti disarankan untuk membandingkan berbagai pendekatan ensemble, seperti Voting dengan *Stacking Ensemble*, untuk menemukan kinerja terbaik dari hasil kolaborasi model.

#### 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," *Indonesian Journal Of Applied Statistics*, Vol. 2, No. 1, P. 34, Jul. 2019, Doi: 10.13057/Ijas.V2i1.29998.
- [2] M. I. Ghazali, W. H. Sugiharto, And A. Fajar Iskandar, "Klik: Kajian Ilmiah Informatika Dan Komputer Analisis Sentimen Pinjaman Online Di Media Sosial Twitter Menggunakan Metode Naive Bayes," *Media Online*), Vol. 3, No. 6, Pp. 1340–1348, 2023, Doi: 10.30865/Klik.V3i6.936.
- [3] A. Ikhsan, M. F. A. Kusuma, A. C. M. Wibowo, And N. A. Rakhmawati, "Pengaruh Akun Bot Pada Sentiment Masyarakat Terhadap Pinjaman Online Di Twitter," *Sistemasi: Jurnal Sistem Informasi* , Vol. 11, No. 1, Pp. 137–147, 2021, [Online]. Available: [Http://Sistemasi.Ftik.Unisi.Ac.Id](http://Sistemasi.Ftik.Unisi.Ac.Id)
- [4] H. Hilmi Zain, R. Maulana Awangga, And W. Isti Rahayu, "Perbandingan Model Svm, Knn Dan Naive Bayes Untuk Analisis Sentiment Pada Data Twitter: Studi Kasus Calon Presiden 2024," *Jimps: Jurnal Ilmiah Mahasiswa Pendidikan Sejarah*, Vol. 8, No. 3, Pp. 2083–2093, 2023, Doi: 10.24815/Jimps.V8i3.25342.
- [5] M. Syarifuddin, "Analisis Sentimen Opini Publik Terhadap Efek Psbb Pada Twitter Dengan Algoritma Decision Tree, Knn, Dan Naive Bayes," *Inti Nusa Mandiri*, Vol. 15, No. 1, Pp. 87–94, Aug. 2020, Doi: 10.33480/Inti.V15i1.1433.
- [6] S. Samsir, A. Ambiyar, U. Verawardina, F. Edi, And R. Watrionthos, "Analisis Sentimen Pembelajaran Daring Pada Twitter Di Masa Pandemi Covid-19 Menggunakan Metode Naive Bayes," *Jurnal Media Informatika Budidarma*, Vol. 5, No. 1, P. 157, Jan. 2021, Doi: 10.30865/Mib.V5i1.2580.
- [7] D. Septhya *Et Al.*, "Implementasi Algoritma Decision Tree Dan Support Vector Machine Untuk Klasifikasi Penyakit Kanker Paru," *Malcom: Indonesian Journal Of Machine Learning And Computer Science*, Vol. 3, No. 1, Pp. 15–19, May 2023, Doi: 10.57152/Malcom.V3i1.591.
- [8] K. Rahayu, V. Fitria, D. Septhya, R. Rahmadden, And L. Efrizoni, "Klasifikasi Teks Untuk Mendeteksi Depresi Dan Kecemasan Pada Pengguna Twitter Berbasis Machine Learning," *Malcom: Indonesian Journal Of Machine Learning And Computer Science*, Vol. 3, No. 2, Pp. 108–114, Sep. 2023, Doi: 10.57152/Malcom.V3i2.780.
- [9] A. Kurniawan And S. Adinugroho, "Analisis Sentimen Opini Film Menggunakan Metode Naive Bayes Dan Lexicon Based Features," 2019. [Online]. Available: [Http://J-Ptiik.Ub.Ac.Id](http://J-Ptiik.Ub.Ac.Id)
- [10] W. Yulita, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naive Bayes Classifier," *Jurnal Data Mining Dan Sistem Informasi*, Vol. 2, No. 2, P. 1, Aug. 2021, Doi: 10.33365/Jdmsi.V2i2.1344.
- [11] N. Wayan And S. Saraswati, "Naive Bayes Classifier Dan Support Vector Machines Untuk Sentiment Analysis," 2020.
- [12] D. L. Olson, *Advanced Data Mining Techniques*. 2008.
- [13] M. Zarlis, R. Widia Sembiring, S. Tunas Bangsa Pematangsiantar, And J. A. Jend Sudirman Blok No, "Analisa Terhadap Perbandingan Algoritma Decision Tree Dengan Algoritma Random Tree Untuk Pre-Processing Data," *Jurnal Sains Komputer & Informatika (J-Sakti)*, No. 1, 2017, [Online]. Available: [Http://Tunasbangsa.Ac.Id/Ejurnal/Index.Php/Jsakti](http://Tunasbangsa.Ac.Id/Ejurnal/Index.Php/Jsakti)

- [14] K. Khadijah And R. Kusumaningrum, "Ensemble Classifier Untuk Klasifikasi Kanker Payudara," *It Journal Research And Development*, Vol. 4, No. 1, Pp. 61–71, Aug. 2019, Doi: 10.25299/Itjrd.2019.Vol4(1).3540.
- [15] P. Mayadewi And E. Rosely, "Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining," 2015.
- [16] Xia, H., Liu, J., & Zhang, Z. J. (2020). Identifying Fintech risk through machine learning: analyzing the Q&A text of an online loan investment platform. *Annals of Operations Research*, 1-21.