

## **PERFORMANCE COMPARISON OF NAIVE BAYES, SUPPORT VECTOR MACHINE AND RANDOM FOREST ALGORITHMS FOR APPLE VISION PRO SENTIMENT ANALYSIS**

Rangga Rizky Pratama<sup>\*1</sup>, Ryan Randy Suryono<sup>2</sup>

<sup>1,2</sup>Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia  
Email: <sup>1</sup>[rangga\\_rizky\\_pratama@teknokrat.ac.id](mailto:rangga_rizky_pratama@teknokrat.ac.id), <sup>2</sup>[ryan@teknokrat.ac.id](mailto:ryan@teknokrat.ac.id)

(Article received: October 17, 2024; Revision: November 18, 2025; published: February 20, 2025)

### **Abstract**

*With the development of spatial computing devices, there arises a need to analyze consumer opinions about products such as the Apple Vision Pro (AVP), a technology that combines augmented reality (AR) and virtual reality (VR). This study aims to analyze consumer opinions on the Apple Vision Pro by utilizing data from the social media platform X. Three algorithms—Random Forest, Support Vector Machine (SVM), and Naïve Bayes—are used in text categorization to identify sentiment trends. Data was collected through a crawling process, resulting in 3,753 entries. After preprocessing and labeling, 2,609 clean data points were obtained, with 1,618 classified as negative and 991 as positive. In sentiment analysis, Random Forest delivered the best performance with an accuracy of 83%, followed by SVM at 80%, and Naïve Bayes at 75%. These results indicate that the Random Forest algorithm is more effective in sentiment categorization related to Apple Vision Pro. This study provides significant contributions to companies in understanding public perceptions and crafting more precise data-driven marketing strategies.*

**Keywords:** *Apple Vision Pro, Naive Bayes, Random Forest, Sentimen Analisis, Support Vector Machine*

## **PERBANDINGAN PERFORMA ALGORITMA NAIVE BAYES, SUPPORT VECTOR MACHINE DAN RANDOM FOREST UNTUK ANALISIS SENTIMEN APPLE VISION PRO**

### **Abstrak**

Dengan perkembangan perangkat komputasi spasial, muncul kebutuhan untuk menganalisis opini konsumen terhadap produk seperti Apple Vision Pro (AVP) yaitu teknologi yang menggabungkan augmented reality (AR) dan virtual reality (VR). Penelitian ini bertujuan untuk menganalisis opini konsumen terhadap Apple Vision Pro dengan memanfaatkan data dari media sosial X. Tiga algoritma, yaitu Random Forest, Support Vector Machine (SVM), dan Naïve Bayes, digunakan dalam pengkategorian teks guna mengidentifikasi tren sentimen. Data dikumpulkan melalui proses crawling, menghasilkan 3753 entri. Setelah melalui tahap preprocessing dan pelabelan, diperoleh 2609 data bersih dengan 1618 kelas negatif dan 991 kelas positif. Dalam analisis sentimen, Random Forest memberikan performa terbaik dengan akurasi 83%, disusul SVM dengan 80%, dan Naïve Bayes dengan 75%. Hasil ini menunjukkan bahwa algoritma Random Forest lebih efektif dalam mengklasifikasikan sentimen terkait Apple Vision Pro. Penelitian ini memberikan kontribusi penting bagi perusahaan dalam memahami persepsi publik dan menyusun strategi pemasaran berbasis data yang lebih tepat.

**Kata kunci:** Analisis Sentimen, Apple Vision Pro, Naive Bayes, Random Forest, Support Vector Machine,

### **1. PENDAHULUAN**

Perluasan media sosial dan platform online telah menghasilkan sejumlah besar data, termasuk opini dan perasaan masyarakat. Dalam penelitian ini, platform sosial yang digunakan adalah Twitter. Data yang dirilis oleh We Are Social menunjukkan bahwa, pada Oktober 2023, jumlah pengguna Twitter di seluruh dunia mencapai 666,2 juta, menempatkan platform tersebut di peringkat ke-12.

Twitter juga sering digunakan untuk melakukan analisis data tentang tren dan opini publik [1].

Sejak awal Juli 2023, akun media sosial Twitter telah berubah nama menjadi X. Kami memiliki akses ke data dan informasi yang akan bermanfaat untuk penelitian analisis sentimen [2]. *Extended Reality* (XR) adalah topik yang sedang diperdebatkan di media sosial seperti YouTube dan X [3]. *Extended Reality* (XR) adalah inovasi digital

yang memiliki potensi untuk mengubah dunia. Teknologi ini menawarkan pengalaman imersif dan interaktif yang melampaui batas-batas realitas fisik dengan memadukan elemen *Augmented Reality* (AR), *Virtual Reality* (VR) dan *Mixed Reality* (MR) [4]. Apple Vision Pro diperkenalkan pada tanggal 5 Juni 2023 adalah kemajuan besar dalam komputasi spasial yang menggabungkan konten digital dengan lingkungan fisik untuk membuat pendidikan lebih menarik dan interaktif [5].

Analisis sentimen adalah metode yang secara otomatis membaca, memahami, dan mengekstrak teks yang tidak terstruktur untuk mempelajari pendapat atau ide individu yang diekspresikan dalam frasa [6]. Metode dapat digunakan untuk mengklasifikasikan pendapat orang ke dalam perasaan positif atau negatif sesuai dengan seberapa besar emosi mereka [7]. Algoritma klasifikasi teks digunakan untuk mengkategorikan sentimen tersebut.

Klasifikasi teks adalah prosedur yang secara otomatis membagi dokumen teks ke dalam berbagai kategori berdasarkan isi yang terkandung di dalamnya [8]. Model *Naïve Bayes Classifier*, *Support Vector Machine* dan *Random Forest* merupakan klasifikasi teks untuk analisis sentimen pada penelitian ini. *Algoritma Naïve Bayes Classifier* adalah metodologi penggalian data yang berasal dari *teorema Bayes*. Konsepnya melibatkan pemanfaatan statistik dan probabilitas [9]. Vapnik pertama kali mengembangkan *Support Vector Machine (SVM)* di tahun 1992 dengan menggabungkan banyak metode pengenalan pola yang unggul [10]. Berbeda dengan metode *Classification and Regression Trees (CART)* yang juga menggunakan teknik pohon keputusan. *Random Forest* menggunakan metode *aggregating bootstrap (bagging)* dan seleksi fitur random (juga dikenal sebagai seleksi fitur random) [11].

Ketiga algoritma ini digunakan dalam penelitian sebelumnya perihal stigma terkait kesehatan mental dalam bentuk analisis sentimen. Setelah diproses dan diberi label, temuan tersebut menghasilkan 3.095 data, di mana 1.635 adalah data kelas negatif, 208 data kelas netral dan 633 data kelas positif. Akurasi 86,11% dihasilkan oleh model *Support Vector Machine*, *Random Forest* 82,55%, dan akurasi 78,19% oleh *Naive Bayes*. Maka, *Support Vector Machine* merupakan teknik terbaik untuk mengkategorikan materi yang membahas tentang kesehatan mental [12]. Penelitian tambahan yang menguji model algoritma SVM untuk analisis sentimen masyarakat terhadap penggunaan chatgpt. Hasil pengujian yang dilakukan menghasilkan performa yang cukup baik dengan nilai akurasi 87,6%. Sentimen yang lebih dominan pada penelitian ini yaitu sentimen netral dengan *precision* 96,53% dan *recall* 85,67%.

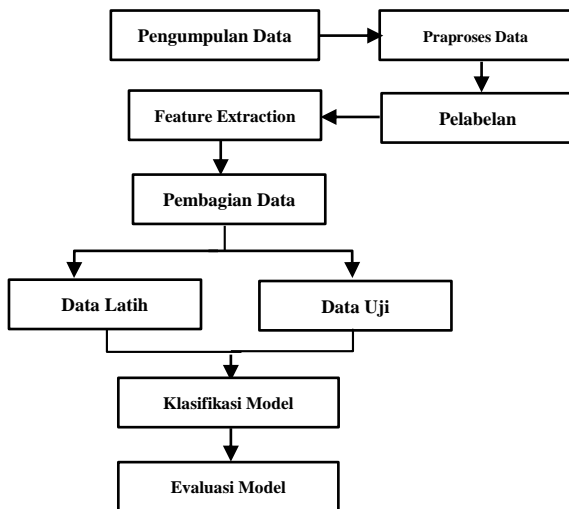
Perkembangan XR juga telah diuji sebelumnya dengan membandingkan Algoritma SVM dan *Random Forest* untuk analisis sentiment metaverse.

Hasil pengujian yang dilakukan menghasilkan akurasi 91% untuk *Random Forest* dan 90% untuk *Support Vector Machine*, dapat disimpulkan bahwa *Random Forest* lebih baik dari *Support Vector Machine* [13]. Penelitian tambahan tentang analisis sentimen metaverse menggunakan algoritma *Naïve Bayes* dan *Logistic Regression*. Hasil pengujian menghasilkan akurasi 90% untuk *Naïve Bayes* dan 91% untuk *Logistic Regression*, hasil tersebut belum dilakukan optimasi SMOTE. Setelah optimasi SMOTE mengalami peningkatan pada setiap algoritma. *Naïve Bayes* meningkat menjadi 91% dan *Logistic Regression* 95%. Dapat disimpulkan dari bahwa algoritma *Logistic Regression* lebih baik dari *Naïve Bayes* dalam topic penelitian ini [14].

Tujuan penelitian ini ialah membandingkan performa 3 algoritma yang berbeda, yaitu *Random Forest*, *Naïve Bayes* dan *Support Vector Machine (SVM)* dalam analisis sentimen terkait produk Apple Vision Pro berdasarkan data Twitter. Menemukan tren sentimen yang positif atau negatif pada item Apple Vision Pro adalah manfaat lain yang diantisipasi dari penelitian ini.

## 2. METODE PENELITIAN

Proses penelitian ini dibagi menjadi 7 tahapan, dimulai dengan data collection, kemudian preprocessing, labeling data untuk mendapatkan sentimen negatif dan positif, feature extraction menggunakan TF-IDF, selanjutnya dipisahkan menjadi set pelatihan dan pengujian untuk langkah berikutnya dalam *classification*, yang menggunakan algoritma *Random Forest*, *Support Vector Machine (SVM)*, dan *Naïve Bayes*. Penggunaan *Naïve Bayes*, *Support Vector Machine (SVM)*, dan *Random Forest* dalam analisis sentimen terhadap Apple Vision Pro bertujuan untuk mengeksplorasi keunggulan masing-masing algoritma dalam mengkategorikan opini konsumen. *Naïve Bayes* menawarkan kecepatan dan efisiensi dalam mengolah data besar, sementara SVM efektif dalam mengatasi data berdimensi tinggi dan pola non-linear. Di sisi lain, *Random Forest*, sebagai algoritma ensemble, meningkatkan akurasi dengan mengurangi risiko overfitting melalui penggabungan beberapa pohon keputusan. Dengan membandingkan ketiga metode ini, penelitian ini tidak hanya memberikan pemahaman yang lebih baik tentang performa masing-masing algoritma, tetapi juga berkontribusi secara praktis dalam membantu perusahaan merancang strategi pemasaran yang lebih tepat berdasarkan analisis sentimen. Langkah terakhir melibatkan evaluasi model dengan membandingkan kinerja ketiga algoritma dalam menginterpretasikan sentimen yang terkait dengan produk Apple Vision Pro menggunakan data Twitter. Hal ini dilakukan dengan menggunakan matriks untuk *accuracy*, *precision*, *recall* dan *F1-Score*. Gambar 1 merupakan Alur penelitian.



Gambar 1. Alur Penelitian

## 2.1. Pengumpulan data

Metode crawling adalah teknik pengumpulan data yang dilakukan secara otomatis dengan program komputer untuk mengambil data dari bermacam *source*. Pengumpulan data yang bertujuan untuk analisis, penelitian dan pengembangan aplikasi, dimungkinkan oleh metode ini [15]. Dalam penelitian ini pengumpulan data dilakukan mulai dari bulan mei sampai september 2024 melalui proses crawling pada platform media sosial twitter menggunakan bahasa pemrograman python dan *library harvest*. *Harvest* adalah *library* pada bahasa python yang digunakan untuk pengumpulan data di twitter dengan memanfaatkan *aut\_token* twitter [16]. Dataset berupa *tweet* pengguna tentang produk Apple Vision Pro dengan kata kunci “apple vision pro”.

## 2.2. Praproses Data

Praproses data adalah proses membenahi data yang mulanya tidak terstruktur menjadi terstruktur dalam format yang diperlukan [17]. Tujuan praproses ialah memastikan data mudah dibaca [18]. Metode preprocessing berikut digunakan dalam penelitian ini:

### 2.2.1. Data Cleaning

*Data cleaning* adalah tahap penghapusan tanda baca, termasuk tanda tanya, tanda seru, tanda kurung, dan titik. Selain itu, terdapat informasi dan simbol duplikat seperti URL, simbol *retweet*, dan nama pengguna. [19].

### 2.2.2. Casefolding

Dalam dokumen teks, metode case folding menghilangkan semua informasi yang mengandung karakter, kecuali huruf dan angka, dan mengubah huruf menjadi huruf kecil [20].

### 2.2.3. Tokenizing

Memecah teks atau data dalam bentuk kalimat menjadi kata dikenal sebagai tokenisasi. [21].

### 2.2.4. Filtering / Stopword removal

Dengan menghilangkan kata-kata yang sering muncul dan dianggap tidak perlu untuk menilai sentimen teks, penghapusan kata-kata tidak penting meningkatkan ketepatan dan penerapan analisis sentimen. Untuk melakukan ini, hanya istilah-istilah penting dengan signifikansi emotif atau evaluatif yang harus dipertahankan [22].

### 2.2.5. Stemming

Stemming adalah proses mengekstrak akar kata tinjauan dengan menghilangkan afiksnya. Awalan, akhiran, atau keduanya mungkin ada dalam afiks ini. Untuk memperlakukan kata-kata dengan akar yang sama sebagai satu kesatuan, proses ini berfungsi untuk mengubah berbagai bentuk kata menjadi bentuk yang paling mendasar [23].

## 2.3. Pelabelan

Pelabelan adalah tahap pemberian label pada data untuk dibagi menjadi klasifikasi positif dan negatif [24]. Penelitian ini menggunakan pustaka *textblob* untuk melabelkan data, yang menghasilkan sentiment pada kelas negatif sebanyak 1618 dan pada kelas positif sebanyak 991.

## 2.4. Future Extraction (TF-IDF)

Karena *machine learning* tidak dapat memahami kata dan karakter dalam dataset, *future extraction* dilakukan agar dapat memahami teks [25]. *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) digunakan untuk menghitung frekuensi kata dalam teks relatif terhadap jumlah total kata dalam dokumen [26].

- *Term Frequency* (TF) :

$$TF(t, d) = \frac{\text{Jumlah kemunculan term } t \text{ dalam dokumen } d}{\text{Total jumlah term dalam dokumen } d}$$

(1)

TF menghitung frekuensi kata dalam teks relative terhadap jumlah keseluruhan data.

- *Inverse Document Frequency* (IDF)

$$IDF(t, D) = \log\left(\frac{\text{Total jumlah dokumen } D}{\text{Jumlah dokumen dalam } D \text{ yang mengandung term } t}\right) + 1$$

(2)

IDF Menganalisis frekuensi kemunculan frasa dalam dokumen.

- TF-IDF

$$TF-IDF(t, d, D) = TF(t, d) \times IDF(t, D)$$

Mengalikan nilai TF dengan nilai IDF menghasilkan TF-IDF. TF-IDF menghasilkan skor dengan bobot yang ditingkatkan dengan mengintegrasikan data tentang distribusi istilah di seluruh koleksi dokumen (IDF) dan frekuensi istilah dalam dokumen (TF).

## 2.5. Pembagian Data

Penting untuk membagi data menjadi set uji dan set pelatihan pada langkah pembagian data. Data uji digunakan untuk mengevaluasi model guna memperoleh hasil terbaik, dan data pelatihan

digunakan untuk membuat model berdasarkan pandangan masing-masing kategori, yaitu pendapat positif dan negatif [27]. Dalam penelitian ini, model dikonfigurasi dengan membagi data menjadi 80% guna pelatihan dan 20% guna pengujian. Pembagian ini umumnya memberikan hasil yang lebih optimal karena mayoritas data dialokasikan untuk proses pelatihan model.

**2.6. Klasifikasi Model**

Model ini bekerja dengan mempelajari pola dari data latih, kemudian menggunakan pengetahuan tersebut untuk memprediksi label dari data baru. Keberhasilan model ini sangat bergantung pada kualitas data pelatihan dan algoritma yang digunakan, serta kemampuan model dalam generalisasi pada data baru.

**2.6.1. Naïve Bayes**

Dalam konteks klasifikasi, algoritma *Naive Bayes* menunjukkan performa yang unggul terkait perhitungan serta akurasi. Dengan menganggap bahwa setiap atribut objek bersifat independen, algoritma ini menggunakan pendekatan vektor informasi untuk menghitung kemungkinan bahwa kategori yang dipilih adalah benar [27]. Bentuk umum teorema bayes ditunjukkan pada Persamaan 3.

$$P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)} \tag{3}$$

Pada persamaan (1), X mewakili data yang kelasnya belum ditentukan, sedangkan Y menunjukkan kelas tertentu. P(Y|X) merupakan hipotesis yang dipengaruhi oleh kondisi X, P(X|Y) menunjukkan kemungkinan awal dari suatu kelas berdasarkan hipotesis yang ada, P(Y) adalah probabilitas hipotesis Y, dan P(X) menggambarkan probabilitas dari X [28].

**2.6.2. Support Vector Machine (SVM)**

*Hyperlane* merupakan garis vektor yang digunakan algoritma SVM berbasis diskriminatif untuk memisahkan data menjadi dua kelompok [29]. Persamaan 4 dan 5 menyajikan *hyperlane* ideal pada *kernel linear*.

$$(w \cdot xi + b) \leq 1, yi = -1 \tag{4}$$

$$(w \cdot xi + b) \geq 1, yi = 1 \tag{5}$$

Nilai *xi* merupakan Data ke *i*, nilai *w.xi* merupakan nilai bobot untuk kelas data ke *i*, *b* adalah nilai bias dan *yi* adalah kelas data ke *i* [25].

**2.6.3. Random Forest**

Dengan membuat properti secara acak untuk setiap node, pendekatan *Random Forest* dapat meningkatkan hasil [30]. Persamaan 6 dapat digunakan untuk melakukan pemilihan kelas *Random Forest* yang paling umum.

$$f(x) = Average(f_1(x), f_2(x), \dots, f_n(x)) \tag{6}$$

Keterangan :

- $f(x)$  : Hasil prediksi
- $f_{1-n}(x)$  : Hasil prediksi dari setiap pohon keputusan
- $(x)$  : Inputan data

**2.7. Evaluasi Model**

Salah satu cara untuk mengevaluasi kemampuan model pembelajaran mesin untuk meklasifikasikan data adalah dengan melakukan evaluasi model guna mengetahui kinerja proses yang menggunakan *confusion matrix*. Tabel berukuran dua kali dua dibuat oleh matriks ini untuk menunjukkan akurasi dan ketidaktepatan prediksi model. Tabel 1 menampilkan deskripsi setiap elemen *Confusion Matrix*.

Tabel 1. *Confusion Matrix*

Actual	Prediction	
	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

Keterangan :

- *True Positive (TP)* : Banyaknya data yang berhasil diidentifikasi dengan tepat sebagai positif oleh model.
- *False Positive (FP)* : Banyaknya data yang keliru diidentifikasi sebagai positif oleh model.
- *True Negative (TN)* : Banyaknya data yang berhasil diidentifikasi dengan tepat sebagai negatif oleh model.
- *False Negative (FN)* : Banyaknya data yang keliru diidentifikasi sebagai negatif oleh model.

Persamaan 7 dapat digunakan untuk menghitung akurasi sistem.

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{7}$$

Nilai presisi, *recall*, dan skor F1 dapat ditentukan dengan menggunakan rumus yang terdapat dalam persamaan 8, 9, dan 10. Rasio TP terhadap semua data positif yang diantisipasi dikenal sebagai nilai presisi, dan rasio TP terhadap semua data aktual dikenal sebagai nilai *recall*. Nilai presisi dan *recall* yang lebih tinggi ditunjukkan oleh nilai skor F1 yang lebih tinggi.

$$precision = \frac{TP}{TP+FP} \tag{8}$$

$$recall = \frac{TP}{TP+FN} \tag{9}$$

$$F1 - Score = 2 \times \frac{precision \times recall}{precision+recall} \tag{10}$$

Nilai yang dapat digunakan sebagai referensi untuk membandingkan model yang telah dibuat dapat diperoleh dari rumus yang disebutkan di atas.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Dataset

Dataset didapatkan pada platform X berdasarkan opini user X terhadap produk Apple Vision Pro. Pengumpulan data dilakukan dengan cara *crawling* dengan keyword “apple vision pro” sehingga mendapatkan dataset sebanyak 3753 data *tweets*. Kemudian hasil *crawling* disimpan dalam bentuk csv untuk dilakukan analisis sentimen.

#### 3.2. Tahap Preprocessing Data

Ada beberapa langkah yang terlibat dalam *preprocessing data*: *Cleaning*, *Casefolding*, *Tokenizing*, *Filtering/Stopword Removal*, *Stemming Data*.

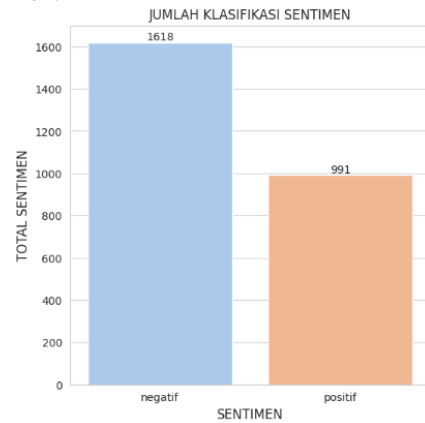
Tabel 2. Hasil *Preprocessing*

Tahapan	Hasil
Dataset	Saya menemukan bahwa @Meta #QuestPro dapat sangat mirip dengan Apple Vision Pro jika Anda melakukan pengaturan dengan benar. <a href="https://t.co/zuAnmYjO2N">https://t.co/zuAnmYjO2N</a>
<i>Cleaning</i>	Saya menemukan bahwa Meta QuestPro dapat sangat mirip dengan Apple Vision Pro jika Anda melakukan dengan benar
<i>Casefolding</i>	saya menemukan bahwa meta questpro dapat sangat mirip dengan apple vision pro jika anda melakukan dengan benar
<i>Tokenizing</i>	['saya', 'menemukan', 'bahwa', 'meta', 'questpro', 'dapat', 'sangat', 'mirip', 'dengan', 'apple', 'vision', 'pro', 'jika', 'anda', 'melakukan', 'dengan', 'benar']
<i>Filtering/Stopword Removal</i>	['menemukan', 'meta', 'questpro', 'apple', 'vision', 'pro']
<i>Stemming</i>	temu meta questpro apple vision pro

Seperti yang terlihat pada Tabel 2. Setelah dilakukan beberapa tahapan pada *preprocessing data*, dataset berkurang sebanyak 1144 data. Sehingga mendapatkan data bersih sebanyak 2609 data, dari data awal yaitu 3753. Data bersih inilah yang akan digunakan untuk dilakukan pengujian selanjutnya.

#### 3.3. Pelabelan

Pelabelan data dilakukan menggunakan Pustaka *textblob* yang menghasilkan kelas negatif dan positif. Dari 2609 data *tweets* terdapat 1618 data kelas negatif dan 991 kelas positif. Pada Gambar 2 merupakan visualisasi dari jumlah klasifikasi sentimen.



Gambar 2. Jumlah klasifikasi sentiment

#### 3.4. Visualisasi Wordcloud

*Wordcloud* adalah visualisasi dataset secara grafis yang membuatnya lebih mudah untuk menemukan kata yang paling sering muncul dalam dokumen. Pustaka *matplotlib* Python digunakan dalam proses pembuatan *wordcloud*. Gambar 3 menampilkan hasil visualisasi *wordcloud*.



Gambar 3. Visualisasi *Wordcloud*

#### 3.5 Tahap Pengujian

Setelah data diproses dalam beberapa tahapan, Untuk mengujinya, algoritma Random Forest, SVM, dan Naive Bayes diperbandingkan. Dua puluh persen data digunakan untuk pengujian, sedangkan delapan puluh persen sisanya digunakan untuk pelatihan. Selama fase pengujian, skor akurasi, presisi, *recall*, *F1-score*, dan *confusion matrix* akan dibandingkan.

##### 3.5.1 *Naïve Bayes*

Algoritma *Naïve Bayes* memiliki persentase akurasi keseluruhan (75%), persentase *recall* tertinggi ialah kelas negatif (97%), persentase *F1-Score* tertinggi yaitu kelas negatif (83%), dan persentase *precision* tertinggi yaitu kelas positif

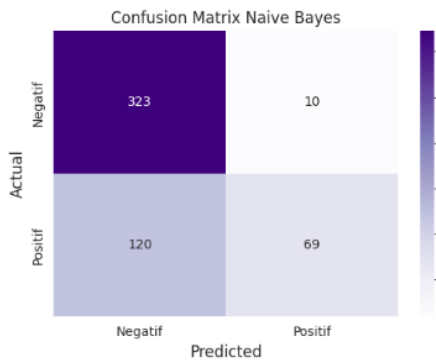
(87%). Hasil klasifikasi algoritma Naïve Bayes dapat diamati di Gambar 4.

	precision	recall	f1-score	support
negatif	0.73	0.97	0.83	333
positif	0.87	0.37	0.51	189
accuracy			0.75	522
macro avg	0.80	0.67	0.67	522
weighted avg	0.78	0.75	0.72	522

Confusion Matrix  
[[323 10]  
[120 69]]  
Accuracy Score :0.75

Gambar 4. Hasil klasifikasi Naïve Bayes

Hasil *confusion matrix* pada algoritma Naïve Bayes menunjukkan 323 data menunjukkan sebagai *True Negative*, 10 data *False Negative*, 69 *True Positif* dan 120 *False Positif*. Seperti yang terlihat pada Gambar 5.



Gambar 5. Hasil *Confusion Matrix* Algoritma Naïve Bayes

### 3.5.2. Support Vector Machine (SVM)

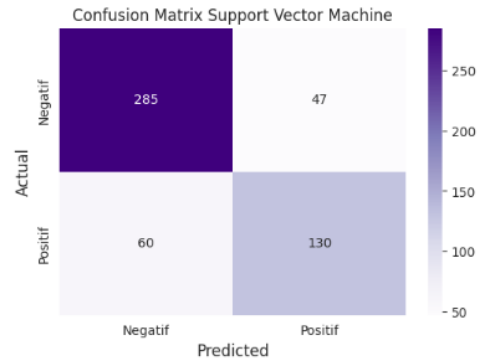
Tingkat *accuracy* keseluruhan metode SVM adalah 80%. Persentase *recall* tertinggi pada kelas negatif (86%), persentase F1-Score tertinggi pada kelas negatif (84%). Persentase *precision* tertinggi yaitu kelas negatif 83%). Gambar 6 menampilkan hasil kategorisasi metode SVM.

	precision	recall	f1-score	support
0	0.83	0.86	0.84	332
1	0.73	0.68	0.71	190
accuracy			0.80	522
macro avg	0.78	0.77	0.78	522
weighted avg	0.79	0.80	0.79	522

Confusion Matrix:  
[[285 47]  
[ 60 130]]  
Accuracy Score: 0.7950191570881227

Gambar 6. Hasil klasifikasi algoritma SVM

Hasil *confusion matrix* pada algoritma SVM menunjukkan 285 data menunjukkan sebagai *True Negative*, 47 data *False Negative*, 130 *True Positif* dan 60 *False Positif*. Gambar 7 menampilkan hasil *confusion matrix* algoritma Naïve Bayes.



Gambar 7. Hasil *Confusion Matrix* algoritma SVM

### 3.5.3. Random Forest

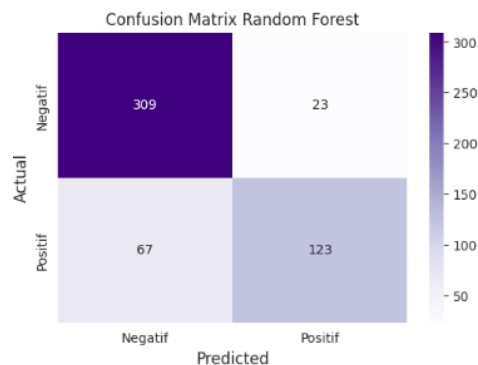
Persentase akurasi keseluruhan algoritma SVM adalah (83%), persentase *recall* tertinggi dikelas negatif (93%), persentase F1-Score tertinggi dikelas negatif (87%), dan persentase *precision* tertinggi di kelas positif (84%). Hasil klasifikasi algoritma SVM dapat dilihat pada Gambar 8.

	precision	recall	f1-score	support
0	0.82	0.93	0.87	332
1	0.84	0.65	0.73	190
accuracy			0.83	522
macro avg	0.83	0.79	0.80	522
weighted avg	0.83	0.83	0.82	522

Confusion Matrix:  
[[309 23]  
[ 67 123]]  
Accuracy Score: 0.8275862068965517

Gambar 8. Hasil klasifikasi algoritmas Random Forest

*Confusion matrix* yang dihasilkan pada model Random Forest menunjukkan 309 data menunjukkan sebagai *True Negative*, 23 data *False Negative*, 123 *True Positif* dan 67 *False Positif*. Gambar 9 menampilkan hasil *confusion matrix* pada *Random Forest*.

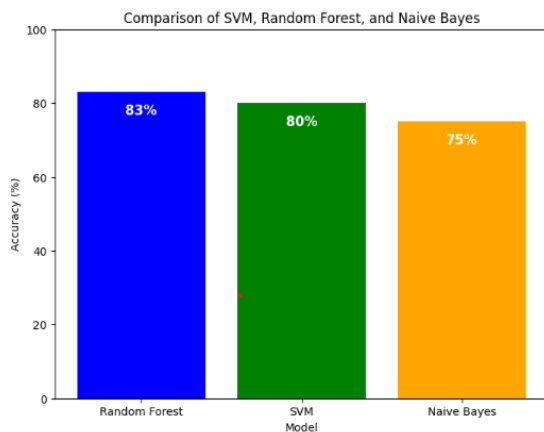


Gambar 9. Hasil *Confusion Matrix* algoritma Random Forest



### 3.6. Visualisasi Perbandingan Algoritma

Visualisasi ini menampilkan perbandingan akurasi algoritma Naïve Bayes, SVM dan Random Forest. Algoritma Naïve Bayes mendapatkan akurasi sebesar 75%, SVM mendapatkan akurasi 80%, sedangkan Random Forest mendapatkan akurasi sebesar 83%. Dari hasil perbandingan tersebut yang memiliki akurasi tertinggi yaitu algoritma Random Forest pada analisis sentimen produk Apple Vision Pro. Hasil visualisasi perbandingan algoritma pada dilihat pada Gambar .



Gambar 10. Visualisasi Perbandingan Algoritma

## 4. DISKUSI

Beberapa penelitian sebelumnya telah melakukan analisis sentimen metaverse menggunakan algoritma SVM dan Random Forest. Penelitian ini menunjukkan bahwa algoritma Random Forest memiliki akurasi lebih baik dibandingkan dengan SVM dengan akurasi sebesar 91% untuk Random Forest dan 90% untuk SVM [13].

Penelitian tambahan membahas tentang analisis sentimen metaverse menggunakan algoritma Naïve Bayes dan Logistic Regression. Hasil pengujian menghasilkan akurasi 90% untuk Naïve Bayes dan 91% untuk Logistic Regression, hasil tersebut belum dilakukan optimasi SMOTE. Setelah optimasi SMOTE mengalami peningkatan pada setiap algoritma. Naïve Bayes meningkat menjadi 91% dan Logistic Regression 95%. Dapat disimpulkan dari bahwa algoritma Logistic Regression lebih baik dari Naïve Bayes dalam topic penelitian ini [14].

Berdasarkan penelitian terdahulu menunjukkan bahwa hasil analisis sentimen terhadap teknologi XR menggunakan ketiga algoritma memiliki hasil yang berbeda-beda. Pada penelitian ini Algoritma *Random Forest* memiliki performa terbaik dengan nilai akurasi sebesar 83%, SVM 80% dan Naïve Bayes 75%. Nilai akurasi tersebut dapat dipengaruhi oleh beberapa faktor, seperti kualitas data, teknik proses data dan cara labeling data. Selain itu, penelitian ini berbeda dari penelitian sebelumnya dalam beberapa hal. Misalnya, penelitian

sebelumnya berfokus pada teknologi metaverse sedangkan pada penelitian ini berfokus pada teknologi *Extended Reality (XR)* yang dikembangkan oleh apple pada produk barunya yaitu *Apple Vision Pro*. Terdapat pada penelitian sebelumnya menggunakan perbandingan sebelum dan sesudah optimasi SMOTE untuk mendapatkan akurasi algoritma. Sedangkan penelitian ini hanya membandingkan algoritma tanpa melakukan optimasi SMOTE.

## 5. KESIMPULAN

Penelitian ini berhasil mengumpulkan 3.753 data melalui proses crawling, yang setelah praproses dan pelabelan menghasilkan 2.609 data bersih dengan 1.618 entri tergolong sebagai sentimen negatif dan 991 sebagai sentimen positif. Dalam analisis performa, algoritma Random Forest menunjukkan akurasi tertinggi sebesar 83%, diikuti oleh Support Vector Machine (SVM) dengan 80%, dan Naïve Bayes dengan 75%. Hal ini menunjukkan bahwa Random Forest paling unggul dalam mengkategorikan tweet terkait opini tentang produk Apple Vision Pro.

Dari temuan ini, dapat disimpulkan bahwa penggunaan algoritma yang tepat dalam analisis sentimen memainkan peran penting dalam menangkap opini publik, yang dapat dimanfaatkan oleh tim pengembang produk atau tim pemasaran untuk mendapatkan wawasan lebih dalam tentang persepsi konsumen. Secara praktis, hasil ini dapat digunakan untuk menyempurnakan strategi pemasaran Apple Vision Pro, khususnya dalam memahami dan merespon sentimen negatif yang lebih dominan.

Untuk penelitian selanjutnya, disarankan agar pendekatan ensemble learning diterapkan guna meningkatkan akurasi prediksi sentimen, terutama ketika dihadapkan pada distribusi data yang tidak seimbang. Selain itu, penggunaan teknik penyeimbangan data dapat membantu mengatasi bias kelas dan memberikan hasil yang lebih representatif. Eksplorasi terhadap dataset yang lebih besar juga akan memperkaya hasil penelitian.

Penelitian selanjutnya juga dapat mencoba menerapkan algoritma deep learning untuk mengeksplorasi performa yang lebih baik dalam sentiment analysis. Dengan menguji algoritma yang lebih canggih, diharapkan prediksi sentimen terhadap produk teknologi seperti Apple Vision Pro menjadi lebih akurat, yang pada akhirnya akan mendukung pengambilan keputusan strategis bagi pengembang dan pemasar dalam memetakan persepsi konsumen.

## DAFTAR PUSTAKA

- [1] N. Thakur, "MonkeyPox2022Tweets: A Large-Scale Twitter Dataset on the 2022 Monkeypox Outbreak, Findings from Analysis of Tweets, and Open Research

- Questions,” *Infect Dis Rep*, vol. 14, no. 6, pp. 855–883, Dec. 2022, doi: 10.3390/idr14060087.
- [2] J. López Díez, “Metaverso: Año Uno. La presentación en vídeo sobre Meta de Mark Zuckerberg (octubre 2021) en el contexto de los estudios previos y prospectivos sobre metaversos,” *Pensar la Publicidad. Revista Internacional de Investigaciones Publicitarias*, vol. 15, no. 2, pp. 299–303, Dec. 2021, doi: 10.5209/pepu.79224.
- [3] S. Stieglitz and L. Dang-Xuan, “Analisis Sentimen Twitter (X) Menggunakan Social Network Analysis Terhadap Kasus Pembunuhan Vin,” *Journal of Management Information Systems*, vol. 29, no. 4, pp. 217–248, Apr. 2024, doi: 10.2753/MIS0742-1222290408.
- [4] S. T. Yun, S. K. Olsen, K. C. Quigley, M. A. Cannady, and A. Hartry, “A Review of Augmented Reality for Informal Science Learning: Supporting Design of Intergenerational Group Learning,” *Visitor Studies*, vol. 26, no. 1, pp. 1–23, 2023, doi: 10.1080/10645578.2022.2075205.
- [5] E. Waisberg *et al.*, “Apple Vision Pro and the advancement of medical education with extended reality,” *Can Med Educ J*, Aug. 2023, doi: 10.36834/cmej.77634.
- [6] E. R. Lidinillah, T. Rohana, and A. R. Juwita, “Analisis sentimen twitter terhadap steam menggunakan algoritma logistic regression dan support vector machine,” *TEKNOSAINS: Jurnal Sains, Teknologi dan Informatika*, vol. 10, no. 2, pp. 154–164, Jul. 2023, doi: 10.37373/tekno.v10i2.440.
- [7] R. N. Ikhsani and F. F. Abdulloh, “Optimasi SVM dan Decision Tree Menggunakan SMOTE Untuk Mengklasifikasi Sentimen Masyarakat Mengenai Pinjaman Online,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, no. 4, p. 1667, Oct. 2023, doi: 10.30865/mib.v7i4.6809.
- [8] I. Athiyah Rahma and L. Hulliyatus Suadaa, “PENERAPAN TEXT AUGMENTATION UNTUK MENGATASI DATA YANG TIDAK SEIMBANG PADA KLASIFIKASI TEKS BERBAHASA INDONESIA STUDI KASUS: DETEKSI JUDUL CLICKBAIT DAN KOMENTAR HATE SPEECH PADA BERITA ONLINE,” *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 10, no. 6, pp. 1329–1340, 2023, doi: 10.25126/jtik.2023107325.
- [9] A. Felicia Watratan, A. B. Puspita, D. Moeis, S. Informasi, and S. Profesional Makassar, “Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia,” 2020. [Online]. Available: <http://journal.isas.or.id/index.php/JACOST>
- [10] A. Rahman Isnain, A. Indra Sakti, D. Alita, and N. Satya Marga, “SENTIMEN ANALISIS PUBLIK TERHADAP KEBIJAKAN LOCKDOWN PEMERINTAH JAKARTA MENGGUNAKAN ALGORITMA SVM,” *JDMSI*, vol. 2, no. 1, pp. 31–37, 2021, [Online]. Available: <https://t.co/NfhnmJtXw>
- [11] F. Azimah and K. Rizky Nova Wardani, “Sistem Pendeteksi Gejala Awal Covid-19 dengan Penggunaan Metode AI Project Cycle,” *Journal Locus Penelitian dan Pengabdian*, vol. 1, no. 6, pp. 405–418, Sep. 2022, doi: 10.36418/locus.v1i6.135.
- [12] P. Elisa and A. Rahman Isnain, “COMPARISON OF RANDOM FOREST, SUPPORT VECTOR MACHINE AND NAIVE BAYES ALGORITHMS TO ANALYZE SENTIMENT TOWARDS MENTAL HEALTH STIGMA,” *Jurnal Teknik Informatika (JUTIF)*, vol. 5, no. 1, pp. 321–329, 2024, doi: 10.52436/1.jutif.2024.5.1.1817.
- [13] P. Kumala Sari and R. Randy Suryono, “KOMPARASI ALGORITMA SUPPORT VECTOR MACHINE DAN RANDOM FOREST UNTUK ANALISIS SENTIMEN METAVERSE,” *Jurnal MNEMONIC*, vol. 7, no. 1, pp. 31–39, 2024.
- [14] B. Ramadhani and R. R. Suryono, “Komparasi Algoritma Naive Bayes dan Logistic Regression Untuk Analisis Sentimen Metaverse,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 2, p. 714, Apr. 2024, doi: 10.30865/mib.v8i2.7458.
- [15] S. Algifari Rismawan, Y. Syahidin, P. Piksi Ganesha Bandung, J. Gatot Subroto No, and K. Batununggal, “Implementasi Website Berita Online Menggunakan Metode Crawling Data Dengan Bahasa Pemrograman Python,” vol. 10, no. 3, pp. 167–178, 2023, [Online]. Available: <http://jurnal.mdp.ac.id>
- [16] B. A. Yuniarossy *et al.*, “ANALISIS SENTIMEN TERHADAP ISU FEMINISME DI TWITTER MENGGUNAKAN MODEL CONVOLUTIONAL NEURAL NETWORK (CNN),” vol. 5, no. 1, 2024, doi: 10.46306/lb.v5i1.
- [17] J. Lasama, A. Pradika, and A. Prasetiadi, “Prediksi Tsunami Pada Gempa Menggunakan Random Forest Classifier,” *Conference on Electrical Engineering, Telematics, Industrial Technology, and Creative Media*, pp. 32–39, 2019.



- [18] P. M. Nirmala Dharmapatni and N. L. P. Merawati, "Penerapan Algoritma Support Vector Machine Dalam Sentimen Analisis Terkait Kenaikan Tarif BPJS Kesehatan," *Jurnal Bumigora Information Technology (BITE)*, vol. 2, no. 2, pp. 105–112, Sep. 2020, doi: 10.30812/bite.v2i2.904.
- [19] U. Banten Jaya, J. Syeh Nawawi Albantani, and S. -Banten, "PERBANDINGAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE (SVM) DALAM KLASIFIKASI SMS SPAM BERBAHASA INDONESIA," *Jurnal Sains & Teknologi*, vol. 3, no. 2, pp. 178–194, 2019.
- [20] K. M. Hana, Adiwijaya, S. Al Faraby, and A. Bramantoro, "Multi-label Classification of Indonesian Hate Speech on Twitter Using Support Vector Machines," in *2020 International Conference on Data Science and Its Applications, ICoDSA 2020*, Institute of Electrical and Electronics Engineers Inc., Aug. 2020. doi: 10.1109/ICoDSA50139.2020.9212992.
- [21] A. Nofandi, N. Y. Setiawan, and D. W. Brata, "Analisis Sentimen Ulasan Pelanggan dengan Metode Support Vector Machine (SVM) untuk Peningkatan Kualitas Layanan pada Restoran Warung Wareg," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 1, pp. 458–466, 2023, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [22] M. I. Fikri, T. S. Sabrila, Y. Azhar, and U. M. Malang, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *SMATIKA Jurnal*, vol. 10, no. 02, pp. 71–76, 2020.
- [23] A. Ilham and W. Pramusinto, "3 rd Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) 30 Agustus 2023- Jakarta," *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, vol. 2, no. 2, pp. 539–547, 2023, Accessed: Nov. 18, 2024. [Online]. Available: <https://senafti.budiluhur.ac.id/index.php/senafti/index>
- [24] D. Apriliani, A. Susanto, M. Fikri Hidayattullah, and G. Wiro Sasmito, "Sentimen Analisis Pandangan Masyarakat Terhadap Vaksinasi Covid 19 Menggunakan K-Nearest Neighbors," *Jurnal Informatika: Jurnal pengembangan IT (JPIT)*, vol. 8, no. 1, pp. 34–37, 2023.
- [25] P. Elisa and A. Rahman Isnain, "COMPARISON OF RANDOM FOREST, SUPPORT VECTOR MACHINE AND NAIVE BAYES ALGORITHMS TO ANALYZE SENTIMENT TOWARDS MENTAL HEALTH STIGMA," *Jurnal Teknik Informatika (JUTIF)*, vol. 5, no. 1, pp. 321–329, 2024, doi: 10.52436/1.jutif.2024.5.1.1817.
- [26] Hermanto, A. Y. Kuntoro, T. Asra, E. B. Pratama, L. Effendi, and R. Ocanitra, "Gojek and Grab User Sentiment Analysis on Google Play Using Naive Bayes Algorithm and Support Vector Machine Based Smote Technique," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Nov. 2020. doi: 10.1088/1742-6596/1641/1/012102.
- [27] M. Fachriza and H. Artikel, "url : <http://studentjournal.umpo.ac.id/index.php/komputek> Analisis Sentimen Kalimat Depresi Pada Pengguna Twitter Dengan Naive Bayes, Support Vector Machine, Random Forest," *Komputek*, vol. 7, no. 2, pp. 49–58, 2023, [Online]. Available: <http://studentjournal.umpo.ac.id/index.php/komputek>
- [28] R. Tesalonika and E. Mailoa, "IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN ISU RESESI EKONOMI 2023 DI INDONESIA PADA PLATFORM TWITTER," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 1, pp. 34–40, Feb. 2024, doi: 10.29100/jupi.v9i1.4288.
- [29] M. Anwar Sadat, P. Pujiono, A. Pambudi, and S. Ibad, "COMPARISON OF ALGORITHM BETWEEN CLASSIFICATION & REGRESSION TREES AND SUPPORT VECTOR MACHINE IN DETERMINING STUDENT ACCEPTANCE IN STATE UNIVERSITIES," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 6, pp. 1589–1604, Jan. 2024, doi: 10.52436/1.jutif.2023.4.6.1565.
- [30] Suci Amaliah, M. Nusrang, and A. Aswi, "Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijjiwa Bantaeng," *VARIANSI: Journal of Statistics and Its application on Teaching and Research*, vol. 4, no. 3, pp. 121–127, Dec. 2022, doi: 10.35580/variensium31.