

## **SENTIMENT ANALYSIS OF POST-COVID ONLINE EDUCATION AMONG GEN Z WITH VARIOUS CLASSIFICATION METHODS**

**Dai Rahman Bakti<sup>1</sup>, Ryan Randy Suryono<sup>\*2</sup>**

<sup>1,2</sup>Information Systems Studies, Faculty of Engineering and Computer Science, Universitas Teknokrat Indonesia,  
Indonesia

Email: <sup>1</sup>[dai\\_rahman\\_bakti@teknokrat.ac.id](mailto:dai_rahman_bakti@teknokrat.ac.id), <sup>2</sup>[ryan@teknokrat.ac.id](mailto:ryan@teknokrat.ac.id)

(Article received: October 12, 2024; Revision: November 11, 2024; published: February 20, 2025)

### **Abstract**

*The COVID-19 pandemic has significantly changed the education sector, shifting from traditional learning to online learning. Generation Z's perception of online education is influenced by their experience as "Digital Natives" who have been familiar with technology since childhood. However, this sudden transition brings new challenges, such as screen fatigue, lack of social interaction, and difficulty in maintaining learning motivation. Sentiment analysis is an important tool to evaluate their experiences and views on online learning. This study aims to investigate Generation Z's views on online education after the pandemic, utilizing various classification methods. Data was collected from Twitter through scraping technique with specific keywords, resulting in a total of 4,986 data obtained using the Tweet Harvest library in Python programming language. The dataset then went through a preprocessing stage, including data cleaning, case folding, tokenizing, stopword removal, and stemming. Before applying Random Forest, SVM, and Naïve Bayes methods, the data is divided into two parts, namely, 3988 training data and 998 testing data with a ratio of 80:20. The accuracy results show that Naïve Bayes achieved 95.49% on training data and 76.05% on testing data, SVM recorded 94.77% accuracy on training data and 87.33% on testing data, and Random Forest obtained 99.97% accuracy on training data and 92.21% on testing data. This research provides important insights into Generation Z's perceptions of post-COVID-19 online education and learning platforms to improve the effectiveness of online learning and identify student challenges in the digital era.*

**Keywords:** Naïve Bayes, Online Education, Random Forest, Sentiment Analysis, SVM.

## **ANALISIS SENTIMEN PENDIDIKAN ONLINE PASCA COVID DI KALANGAN GEN Z DENGAN BERBAGAI METODE KLASIFIKASI**

### **Abstrak**

Pandemi COVID-19 telah mengubah sektor pendidikan secara signifikan, beralih dari pembelajaran tradisional ke pembelajaran daring. Persepsi Generasi Z terhadap pendidikan daring dipengaruhi oleh pengalaman mereka sebagai "Digital Natives" yang telah terbiasa dengan teknologi sejak kecil. Namun, transisi mendadak ini membawa tantangan baru, seperti kelelahan layar, kurangnya interaksi sosial, dan kesulitan dalam menjaga motivasi belajar. Analisis sentimen menjadi alat penting untuk mengevaluasi pengalaman dan pandangan mereka terhadap pembelajaran daring. Penelitian ini bertujuan untuk menyelidiki pandangan Generasi Z mengenai pendidikan online setelah pandemi, dengan memanfaatkan berbagai metode klasifikasi. Data dikumpulkan dari Twitter melalui teknik scraping dengan kata kunci spesifik, menghasilkan total 4.986 data yang diperoleh dengan menggunakan library Tweet Harvest dalam bahasa pemrograman Python. Dataset tersebut kemudian melalui tahap preprocessing, termasuk pembersihan data, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Sebelum menerapkan metode *Random Forest*, *SVM*, dan *Naïve Bayes*, data dibagi menjadi dua bagian yaitu, data training 3988 data dan data testing 998 data dengan rasio 80:20. Hasil akurasi menunjukkan bahwa Naïve Bayes mencapai 95,49% pada data training dan 76,05% pada data testing, SVM mencatat akurasi 94,77% pada data training dan 87,33% pada data testing, serta Random Forest memperoleh akurasi 99,97% pada data training dan 92,21% pada data testing. Penelitian ini memberikan wawasan penting mengenai persepsi Generasi Z terhadap pendidikan online pasca COVID-19 dan platform pembelajaran untuk meningkatkan efektivitas pembelajaran online serta mengidentifikasi tantangan siswa di era digital.

**Kata kunci:** Naïve Bayes, Pendidikan Online, Random Forest, Sentimen Analist, SVM.

## 1. PENDAHULUAN

Pandemi COVID-19 sejak awal 2020 membawa perubahan besar di berbagai sektor, termasuk Pendidikan [1]. Dampak utamanya adalah peralihan mendadak dari pembelajaran konvensional ke pembelajaran daring, menjadi salah satu perubahan paling menonjol selama pandemi [2]. Pendidikan online, yang sebelumnya merupakan opsi tambahan, kini menjadi bentuk utama dalam proses belajar mengajar [3]. Fenomena ini menuntut evaluasi mendalam mengenai dampaknya, terutama terhadap kelompok usia tertentu yang paling terdampak oleh perubahan tersebut. Generasi Z merupakan kelompok yang menarik untuk diteliti, karena tumbuh di tengah perkembangan teknologi digital dan mengalami transformasi signifikan dalam cara mereka mengakses Pendidikan [4]. Persepsi Generasi Z terhadap pendidikan daring selama pandemi COVID-19 dipengaruhi oleh berbagai faktor yang berakar pada pengalaman mereka dengan teknologi digital. Generasi ini dikenal sebagai “digital natives” yang berarti mereka sudah terbiasa dengan teknologi sejak usia dini. Namun, perpindahan tiba-tiba ke pembelajaran daring selama pandemi membawa tantangan baru, seperti kelelahan layar, keterbatasan interaksi sosial, dan kesulitan dalam menjaga motivasi belajar di lingkungan yang terisolasi. [5]. Analisis sentimen menjadi alat yang sangat berguna dalam menilai pengalaman dan persepsi individu terhadap fenomena ini [6]. Dengan mengumpulkan dan menganalisis data dari beragam sumber, termasuk media sosial, forum pendidikan, dan survei online melalui aplikasi X, peneliti dapat mendapatkan pemahaman mendalam tentang persepsi Generasi Z terhadap pendidikan online setelah COVID-19 [7]–[9].

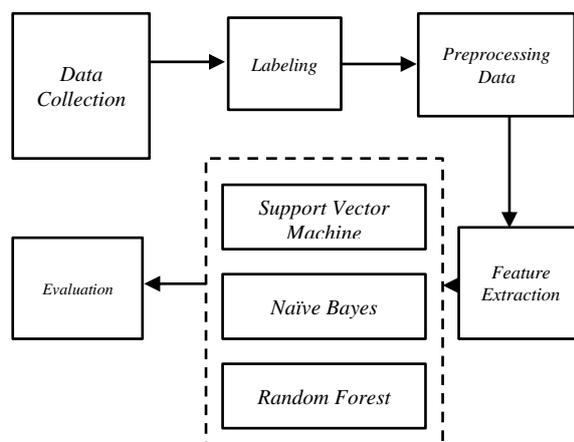
Peneliti terdahulu menunjukkan bahwa pandemi COVID-19 berdampak besar pada pendidikan di Indonesia, sehingga pemerintah menerapkan pembelajaran daring. Data diklasifikasikan dengan metode TF-IDF dan algoritma *Naïve Bayes* serta SVM. Hasilnya menunjukkan bahwa SVM menghasilkan akurasi tertinggi sebesar 84%, di mana sentimen netral mendominasi dengan persentase 43,28%, diikuti oleh sentimen negatif sebesar 32,91% dan sentimen positif sebesar 23,82% [10]. Selain itu, penelitian yang membahas tentang *Twitter* populer di kalangan anak muda Indonesia, sering digunakan untuk berbagi opini. Pada Maret 2020, COVID-19 masuk ke Indonesia, memengaruhi sektor bisnis dan pendidikan. Pemerintah menerapkan pembelajaran daring, yang memicu pro dan kontra. Penelitian ini menganalisis sentimen *tweet* terkait pembelajaran daring menggunakan metode *Naïve Bayes*. Hasil menunjukkan rata-rata akurasi 0.59, precision 0.61, recall 0.60, dan F-measure 0.58 dari 10 percobaan [7].

Analisis sentimen membahas tentang COVID-19, pertama kali muncul di Wuhan, Tiongkok, menyebabkan pandemi global dan mendorong

pembelajaran daring. Penelitian ini menganalisis sentimen mahasiswa di Twitter menggunakan algoritma *Naïve Bayes* dan SVM. *Naïve Bayes* mencapai akurasi 81.20% sedangkan SVM mencapai akurasi 85% [11]. Penelitian ini menganalisis sentimen publik terhadap metaverse menggunakan data media sosial X dan membandingkan algoritma *Random Forest* dan *Support Vector Machine* (SVM). Hasil menunjukkan akurasi 91% untuk *Random Forest* dan 90% untuk SVM, dengan *Random Forest* terbukti lebih baik. Penerapan SMOTE meningkatkan deteksi sentimen positif, meskipun ada trade-off antara recall dan precision [12].

Berdasarkan penelitian sebelumnya, hingga saat ini, perbandingan efektivitas algoritma *Naïve Bayes*, SVM, dan *Random Forest* dalam konteks pendidikan daring di kalangan Generasi Z masih jarang diteliti. Penelitian-penelitian yang ada umumnya hanya fokus pada satu atau dua algoritma tanpa mengeksplorasi potensi keunggulan masing-masing dalam analisis sentimen terkait pengalaman belajar daring [13]–[15]. Penelitian ini berfokus pada analisis sentimen terkait pendidikan online di kalangan Generasi Z pascapandemi COVID-19 dengan menggunakan berbagai metode klasifikasi. Selain itu, penelitian ini menyajikan perbandingan akurasi antar metode yang diterapkan. Diharapkan, hasil penelitian ini dapat memberikan pemahaman mendalam mengenai tantangan, keuntungan, dan potensi perbaikan dalam sistem pendidikan daring. Selain itu, penelitian ini diharapkan dapat memberikan wawasan yang berharga bagi pendidik, pembuat kebijakan, dan pengembang platform pendidikan dalam meningkatkan kualitas pengalaman belajar daring bagi Generasi Z.

## 2. METODE PENELITIAN



Gambar 1. Tahapan Penelitian.

Penelitian ini melibatkan beberapa langkah, yaitu pengumpulan data dari berbagai sumber mengenai stunting, pemrosesan data untuk membersihkan dan menyiapkan informasi, pemberian label pada entitas di dalam dataset,

pembobotan kata untuk menentukan relevansi tiap kata, serta evaluasi model. Evaluasi dilakukan dengan menguji metode *Naïve Bayes*, *Support Vector Machine* (SVM), dan *Random Forest* untuk menganalisis sentimen terkait pendidikan online.

### 2.1. Data Collection

Proses pengumpulan data melibatkan pengambilan informasi dari sumber-sumber yang berkaitan dengan tujuan penelitian atau analisis yang sedang dijalankan [16]. Dalam penelitian ini, data dari Twitter diperoleh dengan melakukan ekstraksi informasi terkait pengguna dan tweet menggunakan kata kunci tertentu, melalui pemanfaatan *Application Programming Interface* (API). Proses pengumpulan dataset dilakukan dengan menerapkan teknik *scraping* untuk mengumpulkan data tweet yang berkaitan dengan pembelajaran online dari platform X. Parameter yang digunakan dalam pencarian mencakup kata kunci “daring” atau “belajar online”. Pengumpulan data ini dilakukan menggunakan *library Tweet Harvest* yang tersedia dalam bahasa pemrograman *Python* yang memungkinkan peneliti untuk melakukan pencarian dan mengunduh tweet berdasarkan kata kunci yang ditentukan, dataset yang terkumpul sebanyak 4299 data. Setelah data berhasil diunduh, file CSV yang dihasilkan dibaca ke dalam *DataFrame* menggunakan pustaka *pandas*. Dari keseluruhan data yang diperoleh, hanya kolom yang relevan diambil, yaitu *created\_at*, *username*, dan *full\_text*, untuk fokus pada informasi penting yang berkaitan dengan waktu pengiriman tweet, pengirim, dan isi tweet itu sendiri. Data yang telah dipilih kemudian disimpan kembali dalam format CSV untuk analisis lebih lanjut.

### 2.2. Labeling

Pelabelan data dalam *text mining* adalah langkah penting di mana setiap elemen teks diberi kategori atau label sesuai dengan ciri-ciri tertentu. Tujuan utamanya adalah untuk mempersiapkan data agar siap untuk analisis yang lebih mendalam. [17]. Dalam penelitian ini, proses pelabelan data dilakukan secara otomatis dengan memanfaatkan pustaka *TextBlob*, dan hasilnya akan dikelompokkan ke dalam tiga kategori: positif, negatif, dan netral. Kategori tersebut diberikan dengan keterangan bahwa nilai 0 mewakili kategori negatif, nilai 1 untuk kategori positif, dan nilai 2 untuk kategori netral.

### 2.3. Preprocessing Data

Proses *preprocessing* data adalah tahap penting dalam menyiapkan data untuk analisis. Tahapan ini mencakup pembersihan data, penyesuaian format huruf, tokenisasi, penghilangan kata umum, dan *stemming*. Melalui langkah-langkah *preprocessing* ini, data menjadi lebih siap untuk dianalisis, yang pada gilirannya meningkatkan akurasi dan relevansi hasil yang diperoleh.

## 1. Cleaning Data

Tabel 1 *Cleaning* bertujuan untuk menghilangkan karakter atau simbol seperti tautan URL, nama pengguna atau *mention* (*@username*), *hashtag* (*#*), *retweet*, dan *emotikon* dari teks [3].

Tabel 1. *Cleaning Data*

<i>Tweet</i>	<i>Cleaning Data</i>
At the end of the day aku bersyukur banget bisa ketemu banyak orang di tahun ini. Mau online kek mau offline. Dari mereka semua aku belajar banyak hal. Thanks all	the end the day aku bersyukur banget bisa ketemu banyak orang tahun ini Mau online kek mau offline Dari mereka semua aku belajar banyak hal Thanks all
Pembelajaran Digital: Materi pelajaran dapat diakses dari mana saja dan kapan saja Menggunakan media interaktif seperti video simulasi kuis online dan forum. Pembelajaran dapat disesuaikan dengan kebutuhan dan kecepatan belajar individu.#nbsoft #elearning #pagii https://t.co/BZ2wNK5cp5	Pembelajaran Digital Materi pelajaran dapat diakses dari mana saja dan kapan saja Menggunakan media interaktif seperti video simulasi kuis online dan forum Pembelajaran dapat disesuaikan dengan kebutuhan dan kecepatan belajar individu #nbsoft #elearning pagii

## 2. Case Folding

Tabel 2 *Case folding* adalah teknik normalisasi teks yang mengubah semua karakter menjadi huruf kecil, sehingga mengurangi variasi penulisan dan meningkatkan akurasi dalam pengolahan teks [18].

Tabel 2. *Case Folding*

<i>Tweet</i>	<i>Case Folding</i>
the end the day aku bersyukur banget bisa ketemu banyak orang tahun ini Mau online kek mau offline Dari mereka semua aku belajar banyak hal Thanks all	the end the day aku bersyukur banget bisa ketemu banyak orang tahun ini mau online kek mau offline dari mereka semua aku belajar banyak hal thanks all
Pembelajaran Digital Materi pelajaran dapat diakses dari mana saja dan kapan saja Menggunakan media interaktif seperti video simulasi kuis online dan forum Pembelajaran dapat disesuaikan dengan kebutuhan dan kecepatan belajar individu nbsoft elearning pagii	pembelajaran digital materi pelajaran dapat diakses dari mana saja dan kapan saja menggunakan media interaktif seperti video simulasi kuis online dan forum pembelajaran dapat disesuaikan dengan kebutuhan dan kecepatan belajar individu nbsoft elearning pagii

## 3. Tokenizing

Tabel 3 merupakan proses *tokenizing* dilakukan untuk menghasilkan token-token dari teks, di mana token merupakan unit terkecil yang memiliki makna, seperti kata atau frasa [3].

Tabel 3. *Tokenizing*

<i>Tweet</i>	<i>Tokenizing</i>
the end the day aku bersyukur banget bisa ketemu banyak orang tahun ini mau online kek mau offline dari mereka semua aku belajar banyak hal thanks all	['', 'the', 'end', 'the', 'day', 'aku', 'bersyukur', 'banget', 'bisa', 'ketemu', 'banyak', 'orang', 'tahun', 'ini', 'mau', 'online', 'kek', 'mau', 'offline', 'dari', 'mereka', 'semua', 'aku', 'belajar', 'banyak', 'hal', 'thanks', 'all']

pembelajaran digital materi pelajaran dapat diakses dari mana saja dan kapan saja menggunakan media interaktif seperti video simulasi kuis online dan forum pembelajaran dapat disesuaikan dengan kebutuhan dan kecepatan belajar individu nbsoft elearning pagii

['pembelajaran', 'digital', 'materi', 'pelajaran', 'dapat', 'diakses', 'dari', 'mana', 'saja', 'dan', 'kapan', 'saja', 'menggunakan', 'media', 'interaktif', 'seperti', 'video', 'simulasi', 'kuis', 'online', 'dan', 'forum', 'pembelajaran', 'dapat', 'disesuaikan', 'dengan', 'kebutuhan', 'dan', 'kecepatan', 'belajar', 'individu', 'nbsoft', 'elearning', 'pagii']

#### 4. Stopword

Tabel 4 dilakukan proses *stopword removal* dilakukan untuk menghilangkan kata-kata yang sering muncul namun kurang relevan dalam analisis teks [19].

Tabel 4. StopWord

Tweet	StopWord
['', 'the', 'end', 'the', 'day', 'aku', 'bersyukur', 'banget', 'bisa', 'ketemu', 'banyak', 'orang', 'tahun', 'ini', 'mau', 'online', 'kek', 'mau', 'offline', 'dari', 'mereka', 'semua', 'aku', 'belajar', 'banyak', 'hal', 'thanks', 'all']	the end the day bersyukur banget ketemu orang online kek offline belajar thanks all
['pembelajaran', 'digital', 'materi', 'pelajaran', 'dapat', 'diakses', 'dari', 'mana', 'saja', 'dan', 'kapan', 'saja', 'menggunakan', 'media', 'interaktif', 'seperti', 'video', 'simulasi', 'kuis', 'online', 'dan', 'forum', 'pembelajaran', 'dapat', 'disesuaikan', 'dengan', 'kebutuhan', 'dan', 'kecepatan', 'belajar', 'individu', 'nbsoft', 'elearning', 'pagii']	pembelajaran digital materi pelajaran diakses media interaktif video simulasi kuis online forum pembelajaran disesuaikan kebutuhan kecepatan belajar individu nbsoft elearning pagii

#### 5. Stemming

Tabel 5 proses *stemming* dengan *library Sastrawi* sangat penting untuk mengurangi variasi bentuk kata dan memfokuskan pada makna dasar dari setiap kata

Tabel 5. Stemming

Tweet	Stemming
['', 'the', 'end', 'the', 'day', 'aku', 'bersyukur', 'banget', 'bisa', 'ketemu', 'banyak', 'orang', 'tahun', 'ini', 'mau', 'online', 'kek', 'mau', 'offline', 'dari', 'mereka', 'semua', 'aku', 'belajar', 'banyak', 'hal', 'thanks', 'all']	['', 'the', 'end', 'the', 'day', 'aku', 'syukur', 'banget', 'bisa', 'ketemu', 'banyak', 'orang', 'tahun', 'ini', 'mau', 'online', 'kek', 'mau', 'offline', 'dari', 'mereka', 'semua', 'aku', 'ajar', 'banyak', 'hal', 'thanks', 'all']
['pembelajaran', 'digital', 'materi', 'pelajaran', 'dapat', 'diakses', 'dari', 'mana', 'saja', 'dan', 'kapan', 'saja', 'menggunakan', 'media', 'interaktif', 'seperti', 'video', 'simulasi', 'kuis', 'online', 'dan', 'forum', 'pembelajaran', 'dapat', 'disesuaikan', 'dengan', 'kebutuhan', 'dan', 'kecepatan', 'belajar', 'individu', 'nbsoft', 'elearning', 'pagii']	['ajar', 'digital', 'materi', 'ajar', 'dapat', 'akses', 'dari', 'mana', 'saja', 'dan', 'kapan', 'saja', 'guna', 'media', 'interaktif', 'seperti', 'video', 'simulasi', 'kuis', 'online', 'dan', 'forum', 'ajar', 'dapat', 'sesuai', 'dengan', 'butuh', 'dan', 'cepat', 'ajar', 'individu', 'nbsoft', 'elearning', 'pagi']

#### 2.4. Feature Extraction

TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah sebuah pendekatan dalam pengolahan teks yang digunakan untuk mengidentifikasi dan mengukur sejauh mana suatu kata penting dalam satu dokumen atau di antara sejumlah dokumen [20]. Dalam pendekatan TF-IDF, setiap kata dalam dokumen diberikan nilai yang menggambarkan tingkat kepentingannya dalam konteks dokumen tersebut, dibandingkan dengan seluruh kumpulan dokumen. Nilai ini dihitung dengan mengintegrasikan dua elemen: TF (Term Frequency), yang merepresentasikan frekuensi kata dalam dokumen, dan IDF (Inverse Document Frequency), yang mengukur seberapa jarang atau eksklusif kata tersebut di seluruh koleksi dokumen [21]. Rumus TF dapat dilihat sebagai berikut:

$$TF(t_i, d_j) = f(t_i, d_j) \quad (2)$$

$f(t_i, d_j)$  merupakan jumlah kemunculan term  $i$  pada dokumen  $j$ . Dengan menggabungkan TF dan IDF, TF-IDF memberikan skor yang tinggi untuk kata-kata yang sering muncul dalam sebuah dokumen, tetapi jarang muncul dalam kumpulan dokumen lainnya. Metode ini berfungsi untuk menyoroti istilah-istilah yang spesifik atau unik dalam dokumen. Nilai TF-IDF diterapkan sebagai atribut dalam analisis teks, seperti pada pengklasifikasian teks dan dalam sistem pencarian informasi.

#### 2.5. Modelling

Pemodelan melibatkan penerapan model statistik atau algoritma machine learning tertentu untuk menganalisis pola dalam data dan membuat prediksi atau klasifikasi. Beberapa model yang sering digunakan dalam konteks ini adalah *Random Forest*, *Support Vector Machine (SVM)*, dan *Naïve Bayes*.

##### 1. Random Forest

Random Forest adalah algoritma pembelajaran mesin ensemble yang menggunakan sejumlah besar pohon keputusan secara independen untuk membuat prediksi [22]. Setiap pohon keputusan dibangun dengan menggunakan dataset latihan yang diambil secara acak dengan penggantian (*bootstrap sampling*) dan pemilihan fitur secara acak pada setiap simpul pemisahan. Prediksi akhir dari *Random Forest* diperoleh melalui agregasi hasil prediksi dari semua pohon keputusan, seperti voting mayoritas untuk klasifikasi atau rata-rata untuk regresi, sehingga menghasilkan model yang kuat, tangguh, dan mampu menangani berbagai jenis data [23]. Beberapa kelebihan dari algoritma Random Forest adalah kinerja yang tinggi dalam klasifikasi dan regresi, ketangguhan terhadap overfitting karena penggunaan *ensemble learning*, kemampuan untuk mengatasi data yang tidak seimbang, serta kemudahan penggunaan dan disesuaikan dengan hyperparameter yang intuitif.

## 2. Naïve Bayes

Algoritma klasifikasi *Naïve Bayes* didasarkan pada *Teorema Bayes* dengan mengasumsikan bahwa fitur-fitur dalam data bersifat independen satu sama lain [24]. Walaupun asumsi ini bisa jadi tidak selalu mencerminkan kenyataan, *Naïve Bayes* sering kali menunjukkan kinerja yang memuaskan, terutama pada dataset yang memiliki banyak dimensi. Beberapa keuntungan dari *Naïve Bayes* adalah kecepatan dalam pelatihan dan prediksi, serta efektivitasnya pada dataset yang relatif besar [25]. Dengan menggunakan *Naïve Bayes* sebagai model dalam proses modelling, kita dapat memanfaatkan keunggulan masing-masing model untuk mendapatkan prediksi atau klasifikasi yang akurat dan berguna dalam analisis data. Selain itu, perbandingan kinerja kedua model ini juga dapat memberikan wawasan tambahan dalam memilih model yang paling sesuai untuk dataset dan tujuan analisis yang spesifik.

$$P(X|Y) = \frac{P(Y|X) \times P(X)}{P(Y)} \quad (2)$$

$P(X|Y)$  adalah probabilitas bahwa X terjadi jika Y telah terjadi, sementara  $P(Y|X)$  adalah probabilitas bahwa Y terjadi jika X telah terjadi, dengan  $P(X)$  dan  $P(Y)$  sebagai probabilitas sebelumnya dari X dan Y secara berturut-turut.

## 3. Support Vector Machine

*Support Vector Machine* (SVM) merupakan algoritma dalam pembelajaran mesin yang diterapkan untuk tugas klasifikasi dan regresi. Algoritma ini beroperasi dengan mencari hyperplane terbaik yang memisahkan data ke dalam kategori-kategori yang berbeda dalam ruang fitur [26]. Hyperplane ini ditentukan sedemikian rupa untuk memaksimalkan jarak (margin) antara data dari setiap kategori dengan hyperplane tersebut. Berikut rumusnya :

$$f(x) = \text{sign}(w^T x + b) \quad (3)$$

Dalam penelitian ini, algoritma *Naïve Bayes*, *Support Vector Machine* (SVM), dan *Random Forest* dipilih sebagai metode klasifikasi karena relevansinya dalam analisis sentimen. *Naïve Bayes* menawarkan kecepatan pelatihan dan prediksi yang tinggi, serta efektivitas dalam menangani dataset besar dengan banyak dimensi. Sementara itu, SVM dikenal karena kemampuannya dalam menemukan hyperplane optimal untuk pemisahan kelas, sehingga sangat efektif dalam mengklasifikasikan data kompleks. Di sisi lain, *Random Forest* memberikan ketahanan terhadap overfitting dan kinerja yang tinggi melalui pendekatan ensemble, yang membuatnya cocok untuk menangani data yang tidak seimbang. Kombinasi ketiga algoritma ini diharapkan dapat memberikan wawasan yang mendalam tentang sentimen pengguna terhadap pembelajaran daring, serta memfasilitasi perbandingan yang komprehensif dalam konteks pendidikan di kalangan Generasi Z.

## 2.6. Evaluation

Model diuji untuk mengevaluasi performa metode *Support Vector Machine*, *Random Forest*, dan *Naïve Bayes*. Dalam konteks ini, evaluasi dilakukan dengan menggunakan akurasi sebagai metrik utama untuk menilai kinerja [27]. Akurasi adalah parameter kunci yang digunakan untuk menilai seberapa baik model klasifikasi atau prediksi dalam menghasilkan hasil yang akurat berdasarkan data yang ada [28]. Tahap evaluasi dilaksanakan dengan memeriksa nilai akurasi, presisi, dan recall. Akurasi merupakan indikator seberapa sering algoritma memberikan prediksi yang tepat, yang dihitung sebagai rasio antara jumlah prediksi yang benar dan total keseluruhan prediksi yang dibuat. Dengan kata lain, akurasi mencerminkan persentase keberhasilan model dalam memprediksi kelas yang sesuai dengan data yang telah disediakan. Persamaan evaluasi model adalah sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

*Precision* mengukur ketepatan prediksi model, yaitu berapa proporsi prediksi positif yang benar dibandingkan dengan semua prediksi positif yang dibuat oleh model. *Precision* tinggi berarti sedikit prediksi positif yang salah.

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

*Recall* untuk mengukur sensitivitas model, yaitu seberapa baik model menangkap semua kasus positif yang ada. *Recall* tinggi berarti model berhasil menemukan sebagian besar kasus positif.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (6)$$

*TP (True Positive)*: Jumlah prediksi yang akurat untuk kelas positif, di mana model mengidentifikasi sebagai positif dan label yang sebenarnya juga positif.

*TN (True Negative)*: Jumlah prediksi yang akurat untuk kelas negatif, di mana model mengidentifikasi sebagai negatif dan label yang sebenarnya juga negatif.

*FP (False Positive)*: Jumlah prediksi yang tidak tepat untuk kelas positif, di mana model mengidentifikasi sebagai positif tetapi label yang sebenarnya negatif.

*FN (False Negative)*: Jumlah prediksi yang tidak tepat untuk kelas negatif, di mana model mengidentifikasi sebagai negatif namun label yang sebenarnya positif.

## 3. HASIL DAN PEMBAHASAN

### 3.1. Pembagian Data

Proses pembagian data menjadi langkah yang sangat penting sebelum menerapkan metode *Random Forest*, SVM, dan *Naïve Bayes*, untuk memastikan efektivitas model. Data tersebut terbagi menjadi dua

kategori: data training yang digunakan untuk melatih algoritma, serta data testing yang berfungsi untuk menguji performa model. Dengan menggunakan rasio pembagian 80:20, total dataset yang terdiri dari 4922 data mencakup 3937 data untuk pelatihan dan 984 data untuk pengujian. Tabel 6 menunjukkan hasil data training dan data testing.

Tabel 6. Hasil Data *Training* dan *Data Testing*

Hasil Data Training	Hasil Data Testing
3833 koramil pasar kliwon	4205 semoga aman ga dapat
memantau kegiatan pengemu...	surat cinta daring mang ...
571 last but not least sabar tekun	2899 link youtube panel
marketing onli...	Lembaga pendidikan
31 pendaftaran kursus sd kelas	perlawanan...
mapel bahasa inggr...	3157 advokat interview pintar
1034 selasa september	mencari diharapkan kl...
bappedalitbang provinsi kalim...	4275 karantina kaltim jg tuan
40 malam peserta indonesia	rumah bg pelantikan p...
besok melanjutkan rang...	1622 iya udah dobel tp daring

### 3.2. Model Klasifikasi

Model klasifikasi adalah algoritma yang digunakan untuk mengelompokkan data ke dalam kategori atau kelas tertentu berdasarkan fitur yang dimiliki data tersebut. Dalam penelitian ini, model klasifikasi digunakan adalah *Naïve Bayes*, *SVM*, dan *Random Forest* untuk menganalisis dan memprediksi sentimen atau persepsi terhadap pendidikan daring. Beberapa algoritma yang sering digunakan dalam klasifikasi. *Confusion Matrix* adalah alat yang digunakan untuk mengevaluasi performa model klasifikasi dengan cara membandingkan prediksi model dengan data yang sebenarnya. Sebelum melakukan evaluasi, dataset dibagi menjadi dua bagian: data training, yang berfungsi untuk melatih model, dan data testing, yang digunakan untuk menguji kinerja model. Pada penelitian ini hasil akurasi model pada data training dan testing untuk tiga algoritma adalah sebagai berikut

Tabel 7. Hasil Perbandingan *Data Training* dan *Data Testing*

Modelling	Data Training	Data Testing
<i>Random Forest</i>	99,97%	92,21%
<i>SVM</i>	94,77%	87,33%
<i>Naïve Bayes</i>	95,49%	76,05%

Pada tabel 7 menunjukkan bahwa model *Random Forest* memiliki akurasi tertinggi dengan 99,97% pada data training dan 92,21% pada data testing. Model *SVM* mencapai akurasi 94,77% pada data training dan 87,33% pada data testing. Sementara itu, model *Naïve Bayes* mencatat akurasi 95,49% pada data training, namun mengalami penurunan signifikan dengan hanya 76,05% pada data testing.

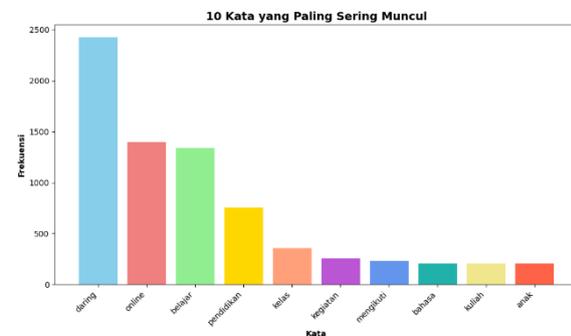
Tabel 8. Hasil *Confusion Matrix Modelling*

Modelling	Accuracy	Precision	Recall
<i>Random Forest</i>	91,74%	95,66%	68,66%
<i>SVM</i>	87,33%	95,66%	46,00%
<i>Naïve Bayes</i>	76,05%	46,66%	50,33%

Tabel 8 menunjukkan hasil evaluasi model menggunakan metrik akurasi, precision, dan recall. Model *Random Forest* menghasilkan akurasi 91,74%,

precision 95,66%, dan recall 68,66%. Model *SVM* mencatat akurasi 87,33% dengan precision 95,66%, namun memiliki recall rendah sebesar 46,00%. Sementara itu, *Naïve Bayes* menunjukkan akurasi 76,05%, precision 46,66%, dan recall 50,33%. Dari hasil ini, meskipun *Random Forest* dan *SVM* menunjukkan precision yang tinggi, performa recall terutama pada *SVM* jauh lebih rendah, sedangkan *Naïve Bayes* memiliki keseimbangan lebih baik antara precision dan recall meskipun dengan akurasi yang lebih rendah.

Berikut ini disajikan sepuluh kata yang paling umum ditemukan dalam dataset, yang akan ditampilkan dalam grafik untuk memberikan visualisasi yang lebih jelas mengenai frekuensi kemunculan setiap kata. Grafik ini bertujuan untuk membantu dalam memahami pola dan tren dalam data, serta mengidentifikasi kata-kata kunci yang mungkin relevan untuk analisis lebih lanjut.



Gambar 2. Kata yang paling sering muncul

Dari total 4299 dataset, terdapat 10 kata yang paling sering muncul. Gambar 2 menunjukkan frekuensi kata bahwa kata daring muncul sebanyak 2425 kali, online muncul 1394 kali, belajar muncul 1340 kali, pendidikan muncul 756 kali, kelas muncul 355 kali, kegiatan muncul 258 kali, mengikuti muncul 234 kali, bahasa muncul 206 kali, kuliah muncul 205 kali, dan anak muncul 203 kali. Hasil ini mengindikasikan bahwa topik utama dalam dataset berkaitan erat dengan pembelajaran daring, pendidikan online, dan keterlibatan siswa dalam kegiatan kelas digital.

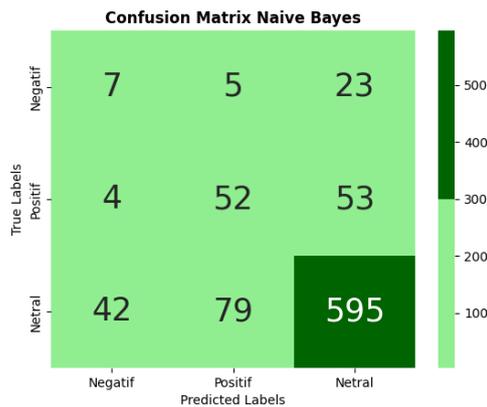
### 3.3. Confusion Matrix

*Confusion Matrix* adalah alat yang menunjukkan bagaimana model klasifikasi memprediksi label dibandingkan dengan label sebenarnya. Tabel ini memetakan hubungan antara prediksi dan kondisi nyata dari tiga kategori, yaitu '0', '1', dan '2'. Masing-masing kategori dapat diartikan sebagai berikut:

#### 1. *Confusion Matrix Naïve Bayes*

*Confusion matrix* berikut menunjukkan hasil klasifikasi model dalam tiga kategori: negatif, positif, dan netral. Dari data yang tersedia, terdapat 7 prediksi benar untuk kategori negatif, 23 data yang salah

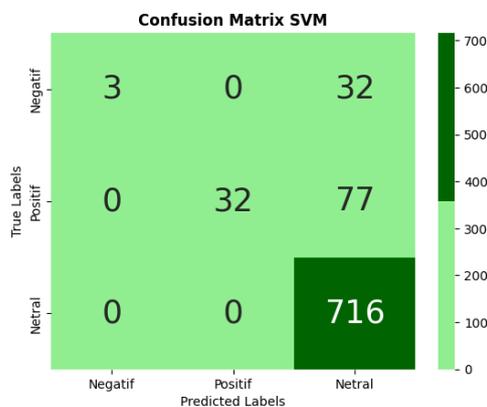
diprediksi sebagai netral, dan 5 data yang salah diklasifikasikan sebagai positif. Untuk kategori positif, model berhasil memprediksi 52 data dengan benar, tetapi terdapat 53 data yang salah dikategorikan sebagai netral dan 4 data yang salah diklasifikasikan sebagai negatif. Sedangkan untuk kategori netral, model memprediksi 595 data dengan benar, tetapi mengalami kesalahan pada 79 data yang salah dikategorikan sebagai positif dan 42 data yang negatif. Analisis ini memberikan wawasan penting mengenai performa model dalam mengklasifikasikan setiap kategori.



Gambar 3. *Confusion Matrix Naive Bayes*

### 2. Confusion Matrix SVM

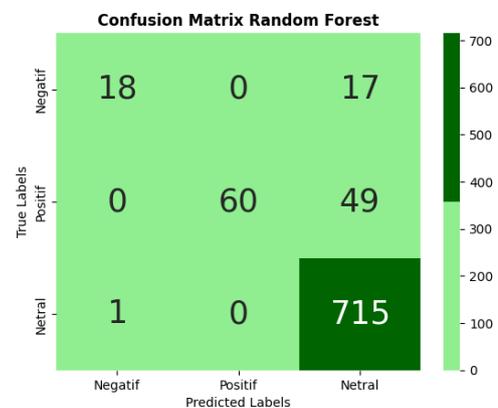
*Confusion matrix* yang dihasilkan dari total data yang diklasifikasikan, model berhasil memprediksi 3 data sebagai negatif dengan benar, sementara 0 data salah diprediksi sebagai positif dan 32 data dikategorikan sebagai netral. Untuk kategori positif, model mengidentifikasi 32 data dengan tepat, tetapi tidak ada data yang diprediksi sebagai negatif dan 77 data salah diklasifikasikan sebagai netral. Di sisi lain, untuk kategori netral, model berhasil memprediksi 716 data dengan benar tanpa kesalahan. Hasil ini memberikan pemahaman yang jelas tentang kinerja model dalam mengklasifikasikan setiap kategori dan menunjukkan area yang perlu diperbaiki dalam prediksi positif dan negatif.



Gambar 4. *Confusion Matrix Support Vector Machine*

### 3. Confusion Matrix Random Forest

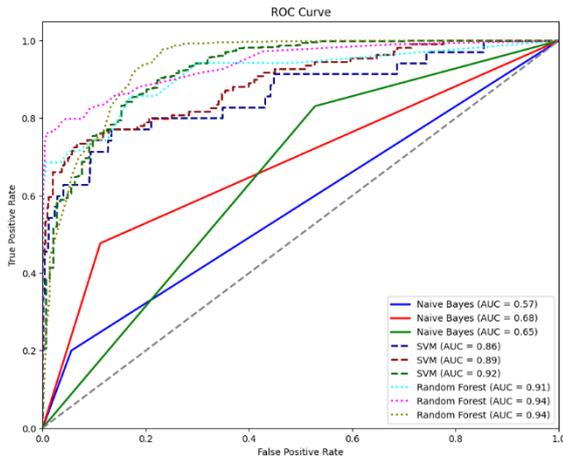
*Confusion Matrix* menggambarkan hasil prediksi model dibandingkan dengan data sebenarnya. Dari hasil tersebut, model berhasil memprediksi kategori negatif dengan benar sebanyak 18 kali, namun tidak ada yang salah memprediksi negatif sebagai positif dan sebagai netral 17 kali. Untuk kategori netral, model memprediksi dengan benar sebanyak 60 kali, tetapi tidak ada yang salah memprediksi netral sebagai negatif, dan prediksi positif sebagai netral sebanyak 49 kali. Sedangkan untuk kategori netral, model berhasil memprediksi dengan benar sebanyak 715 kali, namun salah memprediksi sebagai negatif 1 kali, tanpa kesalahan prediksi ke positif. Hasil ini menunjukkan kekuatan model dalam memprediksi kategori netral dengan sangat baik, namun ada beberapa kesalahan dalam memprediksi kategori positif dan netral.



Gambar 5. *Confusion Matrix Random Forest*

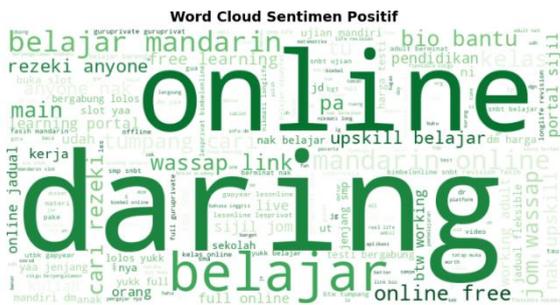
### 3.4. Visualisasi Data

Penelitian ini menyajikan Grafik ROC Curve (*Receiver Operating Characteristic Curve*), yang merupakan alat visual untuk mengevaluasi kinerja model klasifikasi biner dengan menunjukkan hubungan antara *False Positive Rate* (FPR) dan *True Positive Rate* (TPR) pada berbagai threshold. Gambar 6 menggambarkan hasil evaluasi kinerja model klasifikasi yang bervariasi antara algoritma yang diuji. Model Naive Bayes menunjukkan nilai AUC sebesar 0.57, 0.68, dan 0.65, yang menunjukkan kinerja yang rendah dan kemampuan terbatas dalam membedakan antara kelas positif dan negatif. Sebaliknya, model SVM menunjukkan performa yang lebih baik dengan nilai AUC sebesar 0.86, 0.89, dan 0.92, menandakan bahwa model ini mampu memisahkan kelas dengan sangat efektif. Demikian pula, model Random Forest menunjukkan nilai AUC sebesar 0.91, 0.94, dan 0.94, menunjukkan kinerja yang konsisten dan solid dalam klasifikasi. Hasil ini menegaskan bahwa SVM dan *Random Forest* jauh lebih unggul dibandingkan *Naive Bayes* dalam konteks dataset ini, dan memberikan wawasan penting untuk pengembangan model yang lebih akurat dalam analisis data.



Gambar 6. ROC Curve

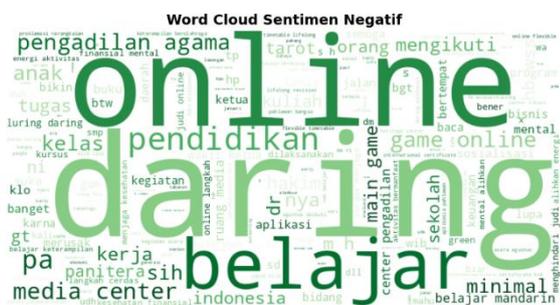
Beberapa *wordcloud* untuk menggambarkan kata-kata yang paling sering muncul dalam sentimen positif, negatif, dan netral, yang membantu memberikan pemahaman yang lebih baik mengenai distribusi kata di setiap kategori sentimen.



Gambar 7. WordCloud Sentimen Positif



Gambar 8. WordCloud Sentimen Netral



Gambar 9. WordCloud Sentimen Negatif

#### 4. DISKUSI

Berdasarkan penelitian yang telah dilakukan, ditemukan bahwa pembelajaran daring telah menjadi metode utama pasca-pandemi COVID-19, dengan Generasi Z menunjukkan persepsi yang beragam terkait pengalaman belajar mereka. Analisis sentimen menggunakan algoritma *Naive Bayes*, SVM, dan *Random Forest* mengungkapkan perbedaan kinerja yang signifikan antara model. *Naive Bayes* menunjukkan akurasi 95,49% pada data training, namun turun drastis menjadi 76,05% pada data testing. SVM mencatat akurasi 94,77% pada data training dan 87,33% pada data testing. Sementara itu, *Random Forest* mencapai akurasi sempurna 99,97% pada data training dan tetap stabil dengan 92,21% pada data testing. Hasil ini menunjukkan bahwa, meskipun semua model menunjukkan kinerja yang baik pada data training, penurunan performa pada data testing, terutama pada *Naive Bayes*, mengindikasikan potensi *overfitting*.

Penelitian sebelumnya menganalisis dampak pandemi COVID-19 pada pendidikan di Indonesia melalui 953.378 tweet dari Januari 2020 hingga Mei 2022. Dengan menerapkan pembelajaran daring, respons masyarakat dieksplorasi melalui analisis sentimen yang mengklasifikasikan data menjadi tiga kelas: negatif, positif, dan netral, menggunakan metode TF-IDF serta algoritma *Naive Bayes* dan SVM. Hasilnya menunjukkan akurasi tertinggi 84% dengan SVM, dengan sentimen netral mendominasi 43,28%, diikuti sentimen negatif 32,91% dan positif 23,82% [10]. Selain itu, penelitian yang dilakukan oleh Prakoso dan Cholissodin (2021) juga menemukan bahwa *Naive Bayes* menghasilkan kinerja yang lebih rendah dalam analisis sentimen pembelajaran daring, dengan rata-rata akurasi sebesar 0,59, precision 0,61, recall 0,60, dan F-measure 0,58 [7].

Secara keseluruhan, penelitian ini memberikan wawasan berharga tentang pandangan Generasi Z terhadap pendidikan daring dan menyoroti perlunya strategi yang lebih baik untuk meningkatkan pengalaman pembelajaran online di masa depan, terutama mengingat perbedaan performa antara data training dan testing pada beberapa algoritma.

#### 5. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan menunjukkan bahwa pendidikan daring telah menjadi metode utama pasca-pandemi COVID-19, dan Generasi Z memiliki persepsi yang beragam terhadap pengalaman belajar mereka. Melalui analisis sentimen yang dilakukan, ditemukan bahwa meskipun model klasifikasi seperti *Naive Bayes*, SVM, dan *Random Forest* menunjukkan akurasi yang tinggi dalam memprediksi sentimen, terdapat perbedaan signifikan dalam kinerja model, terutama antara data training dan testing. Hasil akurasi *Naive Bayes* menunjukkan potensi *overfitting*, dengan

akurasi yang lebih tinggi pada *data training* 95,49% tetapi turun signifikan pada data testing 76,05% pada data testing, SVM mencatat akurasi 94,77% pada data training dan 87,33% pada data testing; sementara Random Forest memperoleh akurasi sempurna 99,97% pada data training dan 92,21% pada data testing. Penelitian ini memberikan wawasan penting tentang bagaimana Generasi Z merasakan pendidikan online, serta menyoroti perlunya strategi yang lebih efektif dalam mengoptimalkan pengalaman pembelajaran daring di masa depan.

## SARAN

Penelitian lanjutan dapat mempertimbangkan penggunaan dataset yang lebih besar untuk meningkatkan generalisasi model dan mengurangi overfitting. Selain itu, menerapkan algoritma lain seperti deep learning dapat memberikan perspektif baru untuk meningkatkan keterlibatan Generasi Z. Selain itu, penelitian selanjutnya dapat mengeksplorasi penggunaan model ensemble untuk meningkatkan akurasi klasifikasi sentimen serta menerapkan metode anotasi manual untuk mendapatkan pemahaman yang lebih mendalam tentang konteks dan nuansa dalam data sentimen yang dianalisis.

## DAFTAR PUSTAKA

- [1] K. Amboro, "Kontekstualisasi Pandemi Covid-19 dalam Pembelajaran Sejarah," *Yupa Hist. Stud. J.*, vol. 3, no. 2, pp. 90–106, 2020, doi: 10.30872/yupa.v3i2.203.
- [2] A. F. Rochim, K. Widyaningrum, and D. Eridani, "Performance Comparison of Support Vector Machine Kernel Functions in Classifying COVID-19 Sentiment," *2021 4th Int. Semin. Res. Inf. Technol. Intell. Syst. ISRITI 2021*, pp. 224–228, 2021, doi: 10.1109/ISRITI54043.2021.9702845.
- [3] A. A. Munandar, F. Farikhin, and C. E. Widodo, "Sentimen Analisis Aplikasi Belajar Online Menggunakan Klasifikasi SVM," *JOINTECS (Journal Inf. Technol. Comput. Sci.)*, vol. 8, no. 2, p. 77, 2023, doi: 10.31328/jointecs.v8i2.4747.
- [4] P. Elisa and A. Rahman Isnain, "Comparison of Random Forest, Support Vector Machine and Naïve Bayes Algorithms To Analyze Sentiment Towards Mental Health Stigma," *J. Tek. Inform.*, vol. 5, no. 1, pp. 321–329, 2024, [Online]. Available: <https://doi.org/10.52436/1.jutif.2024.5.1.1817>
- [5] R. Andriyanty, "Rasionalitas dan Altruisme Generasi Z," *Mediastima*, vol. 30, no. 1, pp. 1–12, 2024, doi: <https://doi.org/10.55122/mediastima.v30i1.1207>.
- [6] S. Lestari and S. Berliani, "Analisis Sentimen Masyarakat Terhadap Isu Pecat Sri Mulyani Pada Twitter Menggunakan Metode Naive Bayes Dan Support Vector Machine," vol. 5, no. 3, pp. 951–960, 2024, doi: <https://doi.org/10.55338/saintek.v5i3.2746>.
- [7] I. Moch Bima Prakoso, Imam Cholissodin, "Analisis Sentimen Masyarakat Terhadap Sistem Pembelajaran Online Selama Pandemi Covid-19 Berdasarkan Dari Twitter Menggunakan Metode Naïve Bayes," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 12, pp. 5376–5383, 2021.
- [8] H. Sujadi, "Analisis Sentimen Pengguna Media Sosial Twitter Terhadap Wabah Covid-19 Dengan Metode Naive Bayes Classifier Dan Support Vector Machine," *INFOTECH J.*, vol. 8, no. 1, pp. 22–27, 2022, doi: 10.31949/infotech.v8i1.1883.
- [9] L. K. Sihombing, T. Tugiono, and U. F. Sari, "Implementasi Data Mining Dalam Menganalisa Pola Penjualan Roti Menggunakan Algoritma Fp-Growth," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 3, p. 228, 2022, doi: 10.53513/jursi.v1i3.5288.
- [10] Ketut Mediana Ayu Candrayani, I Made Agus Dwi Suarjaya, and Anak Agung Ketut Agung Cahyawan Wiranatha, "Analisis Sentimen Pembelajaran Daring Era Pandemi Covid-19 Menggunakan Naive Bayes dan SVM," *Tematik*, vol. 10, no. 1, pp. 47–53, 2023, doi: 10.38204/tematik.v10i1.1274.
- [11] H. Setiawan, E. Utami, and S. Sudarmawan, "analisis sentimen twitter kuliah online pasca covid-19 menggunakan algoritma support vector machine dan naive bayes," *J. Komtika (Komputasi dan Inform.)*, vol. 5, no. 1, pp. 43–51, 2021, doi: 10.31603/komtika.v5i1.5189.
- [12] R. R. S. Putri Kumala Sari, "Komparasi Algoritma Support Vector Machine dan Random Forest Untuk Analisis Sentimen Metaverse," *J. Mnemon.*, vol. 7, no. 1, pp. 31–39, 2024.
- [13] A. Ulfah, "Analisis Sentimen Hate Speech Pada Portal Berita Online Menggunakan Support Vector Machine (SVM)," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, pp. 1–10, Apr. 2020, doi: 10.35957/jatisi.v7i1.196.
- [14] W. Ningsih, B. Alfianda, R. Rahmaddeni, and D. Wulandari, "Perbandingan Algoritma SVM dan Naive Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 556–562, 2024, doi: 10.57152/malcom.v4i2.1253.

- [15] H. Apriyani and K. Kurniati, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus," *J. Inf. Technol. Ampera*, vol. 1, no. 3, pp. 133–143, 2020, doi: 10.51519/journalita.volume1.issue3.year2020.page133-143.
- [16] R. Wijanarko, D. E. Ratnawati, and P. P. Adikara, "analisis sentimen dampak perkembangan artificial intelligence (ai) pada media sosial x/twitter menggunakan metode random forest," vol. 1, no. 1, pp. 2548–964, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [17] H. Herlawati, R. T. Handayanto, P. D. Atika, F. N. Khasanah, A. Y. P. Yusuf, and D. Y. Septia, "Analisis Sentimen Pada Situs Google Review dengan Naïve Bayes dan Support Vector Machine," *J. Komtika (Komputasi dan Inform.,* vol. 5, no. 2, pp. 153–163, 2021, doi: 10.31603/komtika.v5i2.6280.
- [18] D. U. Suhendra, J. Jondri, and I. Indwiarti, "Sentiment Analysis of Hate Speech on Twitter Public Figures with AdaBoost and XGBoost Methods," *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1484, 2022, doi: 10.30865/mib.v6i3.4394.
- [19] R. Julianto, E. D. Bintari, and I. Indrianti, "Analisis Sentimen Layanan Provider Telepon Seluler pada Twitter Menggunakan Metode Naïve Bayesian Classification," *J. Big Data Anal. Artif. Intell.*, vol. 3, no. 1, pp. 23–30, 2018.
- [20] M. H. Wicaksono, M. D. Purbolaksono, and S. Al Faraby, "Perbandingan Algoritma Machine Learning untuk Analisis Sentimen Berbasis Aspek pada Review Female Daily," *eProceedings Eng.*, vol. 10, no. 3, pp. 3591–3600, 2023.
- [21] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, "Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 153–160, 2023, doi: 10.57152/malcom.v3i2.897.
- [22] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," ... *Teknol. Inf. dan ...*, vol. 6, no. 9, pp. 4305–4313, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [23] D. Ahmad Dzulhijjah, H. Sanjaya, A. Said Wahyudi Hidayat, A. Yulistia Alwanda, and E. Utami, "Perbandingan Metode Random Forest dan KNN pada Analisis Sentimen Twitter Review BCA Mobile," *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 12, no. 3, pp. 767–772, 2023, doi: 10.30591/smartcomp.v12i3.5106.
- [24] V. Sari, F. Firdausi, and Y. Azhar, "Perbandingan Prediksi Kualitas Kopi Arabika dengan Menggunakan Algoritma SGD, Random Forest dan Naïve Bayes," *Edumatic J. Pendidik. Inform.*, vol. 4, no. 2, pp. 1–9, 2020, doi: 10.29408/edumatic.v4i2.2202.
- [25] K. S. Putri, I. R. Setiawan, and A. Pambudi, "Analisis Sentimen Terhadap Brand Skincare Lokal Menggunakan Naïve Bayes Classifier," *Technol. J. Ilm.*, vol. 14, no. 3, p. 227, 2023, doi: 10.31602/tji.v14i3.11259.
- [26] G. A. Mursianto, D. Widiyanto, and B. T. Wahyono, "Analisis Sentimen Ulasan Pengguna Pada Aplikasi Google Classroom Menggunakan Metode SVM Dan Seleksi Fitur PSO," *Inform. J. Ilmu Komput.*, vol. 18, no. 3, p. 221, 2022, doi: 10.52958/iftk.v18i3.4685.
- [27] A. S. Rahayu, A. Fauzi, and R. Rahmat, "Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Spotify," *J. Sist. Komput. dan Inform.*, vol. 4, no. 2, p. 349, 2022, doi: 10.30865/json.v4i2.5398.
- [28] M. I. Fikri, T. S. Sabrila, Y. Azhar, and U. M. Malang, "Comparison of the Naïve Bayes Method and Support Vector Machine on Twitter Sentiment Analysis," *SMATIKA J. STIKI Inform. J.*, vol. 10, no. 2, pp. 71–76, 2020.