

## COMPARISON OF ACCURACY OF VARIOUS TEXT CLASSIFICATION METHODS IN SENTIMENT ANALYSIS OF E-STAMPS AT X

Dimas Bagus Reynaldi<sup>\*1</sup>, Ryan Randy Suryono<sup>\*2</sup>

<sup>1,2</sup>Information Systems, Faculty of Engineering and Computer Science, Universitas Teknokrat Indonesia  
Email: <sup>1</sup>[dimas\\_bagus\\_reynaldi@teknokrat.ac.id](mailto:dimas_bagus_reynaldi@teknokrat.ac.id), <sup>2</sup>[ryan@teknokrat.ac.id](mailto:ryan@teknokrat.ac.id)

(Article received: October 12, 2024; Revision: November 24, 2024; published: February 20, 2025)

### Abstract

*In the rapidly evolving digital era, technological innovations are applied in various fields, including law and administration, to improve efficiency. One such innovation in Indonesia is e-stamps, which are designed to facilitate electronic transactions in a legal and secure manner. Although it aims to improve efficiency and security, the public's perception of the application of this technology is still mixed. Although e-stamps aim to improve efficiency and security in transactions, there are still various perceptions from the public that reflect their views and experiences regarding the implementation of this technology. In this case, sentiment analysis is an effective method to evaluate public opinion generated from text data, such as user reviews and comments on social media. This research aims to analyze the sentiment towards e-metallocations in X app, using text classification methods to separate positive and negative sentiments. After collecting 3282 datasets and performing preprocessing that includes case folding, data cleaning, tokenizing, and stemming, the difference in the accuracy results of Naïve Bayes, SVM, and Random Forest is influenced by the characteristics of each method. Naïve Bayes, which assumes feature independence, achieved an average accuracy of 92.26%. SVM excels in separating classes with an optimal hyperplane, resulting in an accuracy of 96.74%. Meanwhile, Random Forest, with its robust ensemble method in the face of data variability, achieved the highest accuracy of 98.48%. These results contribute to e-stamp research and policy by showing the most accurate model for sentiment analysis, supporting decision-making in digital policy in Indonesia.*

**Keywords:** *e-stamps, Modeling, Sentiment Analysis.*

## KOMPARASI AKURASI BERBAGAI METODE KLASIFIKASI TEKS DALAM ANALISIS SENTIMEN E-MATERAI DI X

### Abstrak

Dalam era digital yang berkembang pesat, inovasi teknologi diterapkan di berbagai bidang, termasuk hukum dan administrasi, untuk meningkatkan efisiensi. Salah satu inovasi di Indonesia adalah e-meterai, yang dirancang untuk memfasilitasi transaksi elektronik secara sah dan aman. Meskipun bertujuan meningkatkan efisiensi dan keamanan, persepsi masyarakat terhadap penerapan teknologi ini masih beragam. Dalam hal ini, analisis sentimen menjadi metode efektif untuk mengevaluasi opini publik yang dihasilkan dari data teks, seperti ulasan pengguna dan komentar di media sosial. Penelitian ini bertujuan untuk menganalisis sentimen terhadap e-meterai di aplikasi X, dengan menggunakan metode klasifikasi teks untuk memisahkan sentimen positif dan negatif. Setelah mengumpulkan 3282 dataset dan melaksanakan preprocessing yang mencakup case folding, cleaning data, tokenizing, dan stemming, perbedaan hasil akurasi Naïve Bayes, SVM, dan Random Forest dipengaruhi oleh karakteristik metode masing-masing. Naïve Bayes yang mengasumsikan independensi fitur, mencapai akurasi rata-rata sebesar 92,26%. SVM unggul dalam memisahkan kelas dengan hyperplane optimal, menghasilkan akurasi 96,74%. Sementara itu, Random Forest, dengan metode ensemble yang kuat dalam menghadapi variabilitas data, mencapai akurasi tertinggi, yaitu 98,48%. Hasil ini berkontribusi dalam penelitian dan kebijakan e-meterai dengan menunjukkan model paling akurat untuk analisis sentimen, mendukung pengambilan keputusan dalam kebijakan digital di Indonesia.

**Kata kunci:** *Analisis Sentimen, e-meterai, Modeling,*

### 1. PENDAHULUAN

Dalam era digital yang terus berkembang pesat, berbagai inovasi teknologi telah diterapkan di

berbagai bidang, termasuk hukum dan administrasi, dengan tujuan untuk meningkatkan efisiensi dan efektivitas proses yang ada [1], [2]. Salah satu inovasi terbaru yang telah diperkenalkan di Indonesia adalah penerapan e-meterai atau meterai elektronik, yang dirancang khusus untuk memfasilitasi transaksi elektronik secara sah dan aman, serta untuk memenuhi kebutuhan masyarakat yang semakin bergantung pada teknologi digital dalam aktivitas sehari-hari [3]. Meskipun e-meterai bertujuan untuk meningkatkan efisiensi dan keamanan dalam bertransaksi, masih terdapat beragam persepsi dan tanggapan dari masyarakat yang mencerminkan pandangan serta pengalaman mereka mengenai penerapan teknologi ini [4]. Dalam hal ini, analisis sentimen menjadi metode yang efektif untuk mengevaluasi opini publik yang dihasilkan dari data teks, seperti ulasan pengguna, posting di media sosial, atau komentar yang muncul di berbagai forum online [5]. Dengan melakukan analisis sentimen terhadap e-meterai di aplikasi X, kita dapat memperoleh wawasan mendalam mengenai sikap masyarakat, baik yang bersifat positif maupun negatif, terhadap kebijakan tersebut, yang pada gilirannya dapat memberikan informasi penting bagi pembuat kebijakan dan pengembang teknologi untuk lebih memahami respons publik dan memperbaiki implementasi kebijakan di masa mendatang [6]. Untuk melakukan analisis ini, diperlukan penggunaan berbagai metode klasifikasi teks yang dapat memisahkan sentimen positif dan negatif dengan tingkat akurasi yang tinggi [7].

Analisis sentimen Twitter berbahasa Indonesia dilakukan untuk memahami opini publik dan mendukung riset pasar. Data dikumpulkan melalui pelabelan manual, preprocessing, dan pemodelan dengan mencari istilah “kendaraan listrik” di Kaggle.com. Model klasifikasi dibangun menggunakan algoritma Naïve Bayes dan SVM, dengan hasil menunjukkan akurasi Naïve Bayes sebesar 63,02% dan SVM 70,82%, sehingga SVM memiliki akurasi tertinggi [8]. Penelitian ini menganalisis sentimen masyarakat Indonesia terhadap Luhut Binsar Pandjaitan di media sosial X dengan mengumpulkan 4008 data melalui web scraping dan menganalisisnya menggunakan Naïve Bayes dan SVM yang dioptimasi dengan SMOTE. Hasilnya, Naïve Bayes mencapai akurasi 95%, sedangkan SVM unggul dengan akurasi 99%, presisi 99%, recall 100%, dan F1-Score 99%, menjadikannya pilihan terbaik untuk analisis sentimen [9].

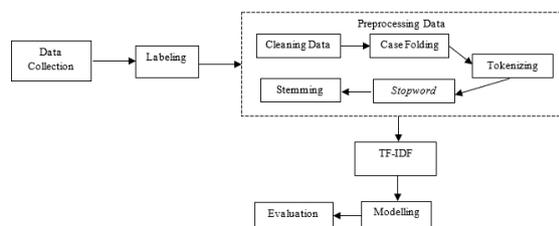
Analisis sentimen terhadap aplikasi Spotify digunakan untuk mengelompokkan opini menjadi negatif atau positif dengan teknik Support Vector Machines (SVM) dan Naïve Bayes. Naïve Bayes unggul dalam kesederhanaan dan akurasi, sedangkan SVM efektif dalam mengidentifikasi hyperplane antara kelas. Hasil penelitian menunjukkan SVM mencapai akurasi 84%, sementara Naïve Bayes lebih

tinggi, yaitu 86,4% [10]. Pegawai adalah aset penting bagi instansi, termasuk pemerintah, tetapi kesulitan dalam perekrutan memunculkan tenaga honorer. Setelah puluhan tahun, pemerintah mengumumkan penghapusan tenaga honorer, yang memicu pro dan kontra di masyarakat. Analisis sentimen terhadap opini publik di Twitter dilakukan dengan algoritma Random Forest, Naïve Bayes, dan SVM, menunjukkan Random Forest sebagai metode terbaik dengan akurasi 66,67% setelah menangani data imbalanced menggunakan *random oversampling* [11].

Berbagai metode klasifikasi, seperti Naïve Bayes, SVM, dan Random Forest, telah digunakan secara luas dalam analisis sentimen. Setiap metode memiliki kekuatan dan kelemahan masing-masing dalam hal akurasi [12],[13]. Oleh karena itu, penelitian ini bertujuan untuk membandingkan akurasi dari berbagai metode klasifikasi teks dalam menganalisis sentimen terkait e-meterai di aplikasi “X” [14], [15]. Melalui komparasi ini, diharapkan dapat ditemukan metode klasifikasi yang paling efektif dalam memahami opini masyarakat terhadap e-meterai, yang nantinya dapat menjadi acuan bagi pengambilan keputusan yang lebih baik oleh para pembuat kebijakan dan pengembang teknologi terkait. Dengan demikian, tujuan utama dari penelitian ini adalah untuk mengidentifikasi metode klasifikasi yang paling akurat dan efisien dalam menganalisis sentimen masyarakat terhadap e-meterai, serta memberikan rekomendasi berdasarkan temuan tersebut untuk meningkatkan implementasi kebijakan terkait.

## 2. METODE PENELITIAN

Metode penelitian ini meliputi tahapan yang terstruktur, dimulai dari pengumpulan data, persiapan data, ekstraksi fitur, penyusunan dataset latih dan uji, pemodelan, hingga evaluasi. Dengan pendekatan ini, memastikan bahwa setiap langkah dalam proses analisis sentimen dikembangkan secara sistematis dan terkoordinasi untuk mencapai hasil yang akurat dan relevan dalam penelitian ini. Berikut adalah tahapan penelitian pada gambar 1.



Gambar 1. Tahapan Penelitian

### a. *Data Collection (Pengumpulan Data)*

Dalam penelitian ini, data dikumpulkan melalui data scraping melalui ekstraksi ulasan pengguna aplikasi X. Dataset yang berhasil dikumpulkan sebanyak 3282 tweet, berisikan empat kolom yang terdiri: 'no', 'created\_at', 'username', dan 'full\_text'

menggunakan pengembang *Google Colab* Kolom 'full\_text' digunakan untuk menyimpan teks lengkap dari setiap entri, yang kemudian dianalisis menggunakan pendekatan berbasis sentimen. Analisis sentimen dilakukan dengan memanfaatkan fungsi *getPolarity*, yang didasarkan pada *TextBlob* untuk mengekstraksi skor polaritas dari setiap teks. Skor polaritas ini digunakan untuk mengukur sentimen positif atau negatif yang terkandung dalam teks. Setiap skor polaritas dihitung dan disimpan dalam kolom baru bernama '*Polarity*', sehingga dapat dianalisis lebih lanjut dalam penelitian ini [16].

**b. Pelabelan**

Pelabelan data merupakan teknik dalam text mining yang melibatkan pemberian tag atau anotasi pada entitas teks seperti kata, frasa, atau kalimat. Tujuannya adalah untuk memberikan struktur dan konteks pada data teks sehingga dapat digunakan sebagai input untuk model pembelajaran mesin [17]. *TextBlob* adalah pustaka pemrosesan bahasa alami yang digunakan untuk menganalisis sentimen teks, di mana setiap teks dalam kolom 'full\_text' diberi label berdasarkan nilai polaritas yang dihasilkan. *Polarity* berkisar antara -1 (sentimen negatif) hingga +1 (sentimen positif). Berdasarkan skor ini, label sentimen diberikan dengan aturan sebagai berikut: jika polaritas lebih besar dari 0, teks diklasifikasikan sebagai sentimen positif; jika kurang dari 0, teks diklasifikasikan sebagai sentimen negatif; dan jika polaritas sama dengan 0, teks diberi label netral. Proses pelabelan ini dilakukan secara otomatis dengan menambahkan kolom baru bernama 'Sentiment' yang memuat label sentimen

Tabel 1. Pelabelan

Username	Tweet	Sentimen
PeruriDigital	Sobat Peruri! Butuh meterai cepat? Sekarang nggak perlu repot! Pakai e-Meterai aja. E-Meterai bisa dibeli dengan mudah di distributor resmi Peruri. Praktis aman dan legal! #SobatPeruri #PeruriDigital #eMeterai	Positif
alfiah67	#PosIND mengajukan izin untuk menjadi distributor E-Meterai dan bekerjasama dengan Peruri. @PosIndonesia #PosIndonesia https://t.co/kmo16UdHuJ	Netral
Nigikege	Sobat Peruri! Butuh meterai cepat? Sekarang nggak perlu repot! Pakai e-Meterai aja. E-Meterai bisa dibeli dengan mudah di distributor resmi Peruri. Praktis aman dan legal! #SobatPeruri #PeruriDigital #eMeterai	Negatif

**c. Preprocessing Data**

*Preprocessing* data adalah langkah penting dalam persiapan data untuk analisis, yang mencakup beberapa tahapan penting [13]. Dengan menerapkan langkah-langkah preprocessing ini, data akan lebih siap untuk analisis selanjutnya, meningkatkan akurasi

dan relevansi hasil yang diperoleh [18]. Pada tahap ini akan dilakukan 5 tahapan, yaitu :

**1. Cleaning Data**

*Cleaning data* atau pembersihan data teks merupakan langkah awal dalam analisis teks yang bertujuan untuk menghilangkan karakter-karakter yang tidak relevan seperti tautan URL, mention pengguna, hashtag, tanda kutip, dan emotikon. Proses ini dilakukan untuk menghasilkan teks yang lebih bersih dan terstruktur sehingga memudahkan proses analisis selanjutnya.

Tabel 2. *Cleaning Data*

Tweet	Cleaning Data
Sobat Peruri! Butuh meterai cepat? Sekarang nggak perlu repot! Pakai e-Meterai aja. E-Meterai bisa dibeli dengan mudah di distributor resmi Peruri. Praktis aman dan legal! #SobatPeruri #PeruriDigital #eMeterai	Sobat Peruri Butuh meterai cepat Sekarang nggak perlu repot Pakai e Meterai aja E Meterai bisa dibeli dengan mudah di distributor resmi Peruri Praktis aman dan legal SobatPeruri PeruriDigital eMeterai
#PosIND mengajukan izin untuk menjadi distributor E-Meterai dan bekerjasama dengan Peruri. @PosIndonesia #PosIndonesia https://t.co/kmo16UdHuJ	PosIND mengajukan izin untuk menjadi distributor E Meterai dan bekerjasama dengan Peruri PosIndonesia

**2. Case Folding**

*Case folding* merupakan teknik yang mengubah semua karakter dalam teks menjadi bentuk huruf kecilnya. Tujuannya adalah untuk membuat perbandingan teks menjadi lebih mudah karena tidak perlu mempertimbangkan apakah huruf pertama suatu kata ditulis dengan huruf besar atau kecil [19].

Tabel 3. *Case Folding*

Tweet	Case Folding
Sobat Peruri Butuh meterai cepat Sekarang nggak perlu repot Pakai e Meterai aja E Meterai bisa dibeli dengan mudah di distributor resmi Peruri Praktis aman dan legal SobatPeruri PeruriDigital eMeterai	sobat peruri butuh meterai cepat sekarang nggak perlu repot pakai e meterai aja e meterai bisa dibeli dengan mudah di distributor resmi peruri praktis aman dan legal sobatperuri peruridigital emeterai
PosIND mengajukan izin untuk menjadi distributor E Meterai dan bekerjasama dengan Peruri PosIndonesia	posind mengajukan izin untuk menjadi distributor e meterai dan bekerjasama dengan peruri posindonesia

**3. Tokenizing**

Tokenisasi merupakan langkah penting dalam analisis teks yang bertujuan untuk mengubah teks menjadi representasi numerik yang dapat diproses oleh komputer. Dengan memecah teks menjadi token, kita dapat menganalisis frekuensi kemunculan kata, melakukan stemming, atau membangun model bahasa [20].

Tabel 4. *Tokenizing*

<i>Tweet</i>	<i>Tokenizing</i>
sobat peruri butuh meterai cepat sekarang nggak perlu repot pakai e meterai aja e meterai bisa dibeli dengan mudah di distributor resmi peruri praktis aman dan legal sobatperuri peruridigital emeterai	['sobat', 'peruri', '', 'butuh', 'meterai', 'cepat', '', 'sekarang', 'nggak', 'perlu', 'repot', '', 'pakai', 'e', 'meterai', 'aja', '', 'e', 'meterai', 'bisa', 'dibeli', 'dengan', 'mudah', 'di', 'distributor', 'resmi', 'peruri', '', 'praktis', 'aman', 'dan', 'legal', '', 'sobatperuri', '', 'peruridigital', 'emeterai']
posind mengajukan izin untuk menjadi distributor e meterai dan bekerjasama dengan peruri posindonesia	['', 'posind', 'mengajukan', 'izin', 'untuk', 'menjadi', 'distributor', 'e', 'meterai', 'dan', 'bekerjasama', 'dengan', 'peruri', 'posindonesia', '']

4. **Stopword**

Proses *stopword removal* biasanya dilakukan dengan menggunakan daftar kata-kata stop (*stoplist*) yang telah ditentukan sebelumnya. Kata-kata dalam teks yang cocok dengan kata-kata dalam *stoplist* akan dihilangkan. Selain menggunakan *stoplist*, kita juga dapat menggunakan *wordlist* yang berisi kata-kata penting yang ingin kita pertahankan [21].

Tabel 5. *StopWord*

<i>Tweet</i>	<i>StopWord</i>
['sobat', 'peruri', 'butuh', 'meterai', 'cepat', 'sekarang', 'nggak', 'perlu', 'repot', 'pakai', 'meterai', 'aja', 'meterai', 'bisa', 'dibeli', 'dengan', 'mudah', 'distributor', 'resmi', 'peruri', 'praktis', 'aman', 'dan', 'legal', 'sobatperuri', 'peruridigital', 'emeterai']	sobat peruri butuh meterai cepat nggak repot pakai e meterai aja e meterai dibeli mudah distributor resmi peruri praktis aman legal sobatperuri peruridigital emeterai
['', 'posind', 'mengajukan', 'izin', 'untuk', 'menjadi', 'distributor', 'e', 'meterai', 'dan', 'bekerjasama', 'dengan', 'peruri', 'posindonesia', '']	posind mengajukan izin distributor e meterai bekerjasama peruri posindonesia

5. **Stemming**

Dalam tahap *stemming*, setiap kata akan dipecah menjadi bagian akarnya dengan menghilangkan imbuhan seperti awalan dan akhiran. Proses ini dilakukan menggunakan library Sastrawi yang khusus dirancang untuk bahasa Indonesia.

Tabel 6. *Stemming*

<i>Tweet</i>	<i>Stemming</i>
['sobat', 'peruri', 'butuh', 'meterai', 'cepat', 'sekarang', 'nggak', 'perlu', 'repot', 'pakai', 'meterai', 'aja', 'meterai', 'bisa', 'dibeli', 'dengan', 'distributor', 'resmi', 'peruri', 'praktis', 'aman', 'dan', 'legal', 'sobatperuri', 'peruridigital', 'emeterai']	['sobat', 'uri', '', 'butuh', 'meterai', 'cepat', '', 'sekarang', 'nggak', 'perlu', 'repot', '', 'pakai', 'e', 'meterai', 'aja', '', 'e', 'meterai', 'bisa', 'beli', 'dengan', 'mudah', 'di', 'distributor', 'resmi', 'uri', '', 'praktis', 'aman', 'dan', 'legal', 'sobatperuri', '', 'peruridigital', '', 'emeterai']
['', 'posind', 'mengajukan', 'izin', 'untuk', 'menjadi', 'distributor', 'e', 'meterai', 'dan', 'bekerjasama', 'dengan', 'peruri', 'posindonesia', '']	['', 'posind', 'aju', 'izin', 'untuk', 'jadi', 'distributor', 'e', 'meterai', 'dan', 'bekerjasama', 'dengan', 'uri', 'posindonesia', '']

d. **Feature Extraction**

Dalam penelitian ini, fitur teks diekstraksi menggunakan metode *Bag of Words* (BoW) dan *Term Frequency-Inverse Document Frequency* (TF-IDF). Tujuan dari proses *Bag of Words* (BoW) adalah untuk mengubah data teks mentah menjadi representasi numerik yang dapat diproses oleh algoritma pembelajaran mesin [22]. Proses BoW menghasilkan sparse matrix dengan ukuran (3282, 5216), yang berarti terdapat 3.282 dokumen teks dan 5.216 fitur atau kata unik yang teridentifikasi dari keseluruhan korpus. Dari hasil ini, sebanyak 42.664 elemen dalam matriks memiliki nilai non-nol, menunjukkan frekuensi kemunculan kata dalam dokumen tertentu. Representasi BoW ini memberikan gambaran awal tentang distribusi kata dalam data yang digunakan. Setelah memperoleh representasi BoW, dilakukan transformasi TF-IDF untuk memberi bobot pada kata-kata yang lebih relevan dalam analisis sentimen. TF-IDF menghitung bobot berdasarkan frekuensi kemunculan kata dalam sebuah dokumen dibandingkan dengan seluruh koleksi dokumen, sehingga mengurangi pengaruh kata-kata umum yang kurang signifikan.

e. **Modelling**

Modelling melibatkan penggunaan model statistik atau algoritma pembelajaran mesin tertentu untuk menganalisis pola-pola yang terdapat dalam data dan membuat prediksi atau klasifikasi. Dalam penelitian ini, metode Naïve Bayes, Support Vector Machine (SVM), dan Random Forest dipilih karena masing-masing algoritma memiliki keunggulan yang sesuai untuk analisis sentimen. Naïve Bayes dipilih karena kesederhanaannya dan efisiensinya dalam memproses data teks dengan asumsi independensi antar fitur, sehingga cocok untuk klasifikasi cepat dengan hasil yang memadai. SVM dipilih karena kemampuannya untuk menangani dataset yang lebih besar dan kompleks dengan memaksimalkan margin antara kelas-kelas, yang dapat meningkatkan akurasi klasifikasi. Sementara itu, Random Forest dipilih karena kemampuannya untuk mengurangi overfitting melalui ensemble learning dan memberikan hasil yang lebih stabil dengan akurasi tinggi. Kombinasi ketiga metode ini memberikan perspektif yang komprehensif dalam mengevaluasi dan membandingkan performa klasifikasi sentimen.

1. **Random Forest**

Random Forest merupakan algoritma ensemble learning yang menggabungkan sejumlah besar pohon keputusan yang dibangun secara acak untuk meningkatkan kinerja prediksi [18]. 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 yang berbasis pada Teorema Bayes dengan asumsi sederhana bahwa fitur-fitur dalam data adalah independen satu sama lain. Meskipun asumsi ini terkadang tidak realistis, Naïve Bayes sering memberikan kinerja yang baik, terutama pada dataset dengan dimensi tinggi. Keunggulan Naïve Bayes termasuk kecepatan pelatihan dan prediksi yang tinggi serta kinerja yang baik pada dataset yang cukup besar [24]. Dengan menggunakan Random Forest dan 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.

$$P(X|Y) = \frac{P(Y|X) \times P(X)}{P(Y)} \quad (1)$$

$P(X|Y)$  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) adalah algoritma pembelajaran mesin yang kuat dalam melakukan klasifikasi dan regresi. Prinsip kerjanya berfokus pada pencarian hyperplane optimal yang memisahkan data menjadi kelas-kelas yang berbeda [25]. Hyperplane ini dipilih sedemikian rupa sehingga jarak antara hyperplane dengan data terdekat dari masing-masing kelas (disebut margin) maksimal. Dengan memaksimalkan margin, SVM menghasilkan model yang lebih generalisasi dan robust terhadap noise dalam data. Berikut rumusnya:

$$f(x) = \text{sign}(w^T x + b) \quad (2)$$

Dimana:

- $w$ : Vektor bobot yang menentukan arah dan kemiringan hyperplane. Bobot ini diperoleh selama proses pelatihan model.
- $x$ : Vektor fitur dari data baru yang ingin diklasifikasikan.
- $b$ : Bias yang menyesuaikan posisi hyperplane dalam ruang fitur. Bias ini juga dihasilkan selama pelatihan.

- $w^T x + b$ : Ini adalah fungsi linear yang menghitung jarak data xxx dari hyperplane. Hasilnya bisa positif, negatif, atau nol.
- $\text{sign}()$ : Fungsi ini mengembalikan tanda dari nilai yang dihitung. Jika hasilnya lebih besar dari nol, fungsi ini mengembalikan +1 (menunjukkan bahwa data tersebut termasuk dalam kelas positif). Jika hasilnya kurang dari nol, fungsi ini mengembalikan -1 (menunjukkan bahwa data tersebut termasuk dalam kelas negatif).

## f. Evaluation

Pengujian model untuk menilai kinerja metode *Random Forest*, *Support Vector Machine*, dan *Naïve Bayes*. Dalam hal ini, pengujian model menggunakan data uji yang terpisah. Metrik evaluasi utama yang digunakan dalam penelitian ini adalah akurasi. [10]. Akurasi adalah suatu ukuran yang menunjukkan seberapa sering prediksi yang dihasilkan oleh model sesuai dengan hasil yang sebenarnya. Dalam konteks klasifikasi, akurasi mengukur proporsi data yang diklasifikasikan dengan benar oleh model [26]. Hal ini tergambar dalam kemampuan model untuk memprediksi kelas atau label dengan tepat, yang mengindikasikan seberapa baik model mampu mengklasifikasikan data secara keseluruhan. Dengan kata lain, akurasi mencerminkan persentase keberhasilan model dalam memprediksi kelas yang sesuai dengan data yang telah diberikan. Persamaan evaluasi model adalah sebagai berikut:

$$\text{Akurasi} = \frac{(\text{Jumlah Prediksi Benar})}{(\text{Total Jumlah Data})} \quad (3)$$

## 3. HASIL DAN PEMBAHASAN

### 3.1. Pembagian Data

Dari total 3282 data yang tersedia, dataset dibagi menjadi dua bagian dengan rasio 80:20, yaitu 2625 data digunakan untuk melatih model (*training data*), dan 657 data digunakan untuk menguji performa model (*testing data*). Pembagian ini bertujuan untuk memastikan bahwa model memiliki cukup data untuk belajar dan dapat dievaluasi secara objektif menggunakan data yang belum dilihat selama proses pelatihan.

### 3.2. Tahap Pengujian

Dalam tahap pengujian mengevaluasi kinerja tiga algoritma klasifikasi, yaitu Naïve Bayes, SVM, dan Random Forest, dengan menggunakan metode cross-validation. Tabel 6 menunjukkan bahwa Naïve Bayes menghasilkan skor cross-validation dengan nilai 0,9391, 0,9072, 0,9055, 0,9345, dan 0,9268, yang memberikan akurasi rata-rata sebesar 92,26%. Di sisi lain, SVM menunjukkan performa yang lebih baik dengan skor 0,9680, 0,9787, 0,9680, 0,9588, dan 0,9634, menghasilkan akurasi rata-rata sebesar 96,74%. Sementara itu, Random Forest berhasil

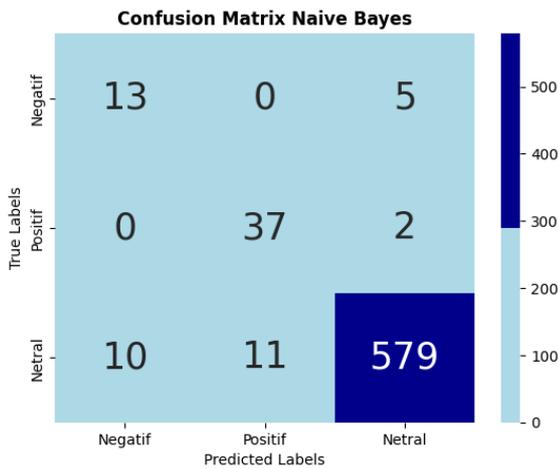
menunjukkan kinerja terbaik dengan skor 0,9909, 0,9817, 0,9863, 0,9832, dan 0,9817, menghasilkan akurasi rata-rata tertinggi sebesar 98,48%.

Tabel 6. Cross Validation Modelling

Percobaan	Random Forest	SVM	Naïve Bayes
1	99,08%	96,80%	93,91%
2	98,17%	97,86%	90,71%
3	98,62%	96,79%	90,54%
4	98,32%	95,88%	93,44%
5	98,17%	96,34%	92,68%

Hasil ini mengindikasikan bahwa Random Forest adalah model yang paling efektif untuk klasifikasi data dalam penelitian ini, diikuti oleh SVM dan Naïve Bayes. Meskipun Naïve Bayes memiliki akurasi yang lebih rendah, model ini tetap mampu memberikan kinerja yang baik dalam konteks analisis sentimen. Keunggulan Random Forest dalam hal akurasi dapat dijelaskan oleh kemampuannya untuk mengurangi overfitting melalui agregasi keputusan dari berbagai pohon keputusan. Temuan ini memberikan wawasan penting dalam pemilihan model yang tepat untuk aplikasi analisis sentimen, serta menunjukkan bahwa pemilihan algoritma yang sesuai berperan signifikan terhadap kinerja model dalam pengolahan data.

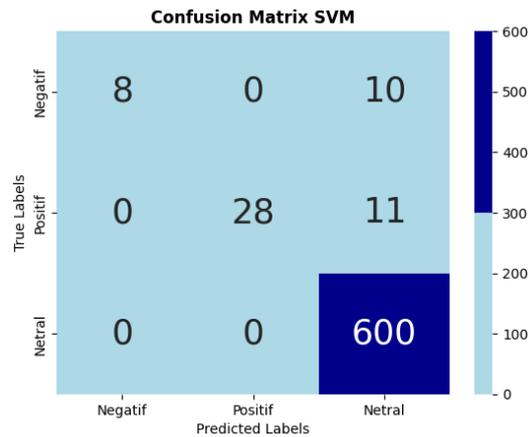
Dalam menentukan model algoritma yang paling optimal, peneliti melakukan eksperimen dengan membandingkan nilai-nilai dalam *confusion matrix* dari ketiga algoritma yang diterapkan. Hasil perbandingan *confusion matrix* untuk masing-masing model algoritma dapat dilihat pada gambar 2 sebagai berikut:



Gambar 2. Confusion Matrix Naïve Bayes

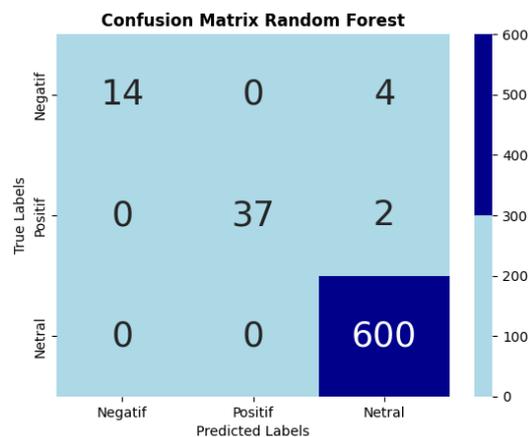
*Confusion matrix* untuk model Naïve Bayes menunjukkan bahwa dari total data yang diuji, model berhasil mengklasifikasikan 13 teks sebagai negatif dengan benar, tanpa ada teks positif yang salah diklasifikasikan sebagai negatif (*false positives*). Namun, terdapat 5 teks negatif yang salah dikategorikan sebagai netral atau positif (*false negatives*). Untuk kategori positif, model berhasil mengidentifikasi 37 teks dengan benar, tetapi terdapat 2 teks positif yang salah diklasifikasikan sebagai

netral atau negatif. Sementara itu, pada kategori netral, model menunjukkan kinerja yang sangat baik dengan 579 teks yang benar-benar diidentifikasi sebagai netral, meskipun terdapat 14 teks negatif dan 19 teks positif yang salah dikategorikan sebagai netral. Hasil dari *confusion matrix* ini mengindikasikan bahwa model Naïve Bayes lebih efektif dalam mengenali teks positif dan netral dibandingkan dengan teks negatif.



Gambar 3. Confusion Matrix SVM

*Confusion matrix* untuk model *Support Vector Machine* menunjukkan kinerja klasifikasi pada tiga kategori sentimen: negatif, positif, dan netral. Dari hasil pengujian, model berhasil mengklasifikasikan 8 data negatif dengan benar, tanpa kesalahan dalam mengidentifikasi data negatif sebagai positif. Namun, terdapat 10 data negatif yang salah diklasifikasikan sebagai netral. Model juga menunjukkan akurasi yang baik dalam mengidentifikasi data positif, dengan 28 data positif terklasifikasi dengan tepat, meskipun 11 data positif salah dikategorikan sebagai netral. Kategori netral memiliki performa yang sangat baik, dengan 600 prediksi benar, dan tidak ada kesalahan klasifikasi untuk kategori ini. Hasil ini menunjukkan bahwa meskipun *Support Vector Machine* efektif dalam mengenali sentimen positif dan netral, ada ruang untuk perbaikan dalam mengidentifikasi sentimen negatif.

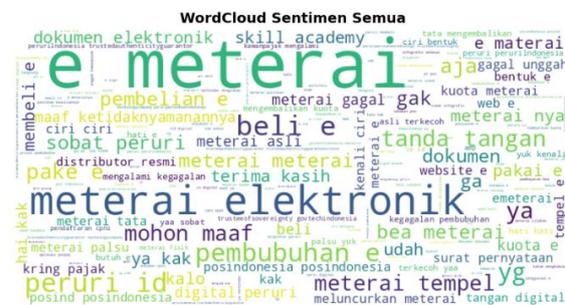


Gambar 4. Confusion Matrix Random Forest

Gambar 4 Confusion matrix untuk model *Random Forest* menunjukkan hasil klasifikasi pada tiga kategori sentimen: negatif, positif, dan netral. Dari hasil pengujian, model berhasil mengklasifikasikan 14 data negatif dengan benar, tanpa kesalahan dalam mengidentifikasi data negatif sebagai positif. Namun, terdapat 4 data negatif yang salah dikategorikan sebagai netral. Model juga menunjukkan kinerja yang baik dalam mengidentifikasi data positif, dengan 37 data positif terklasifikasi dengan tepat, meskipun 2 data positif salah dikategorikan sebagai netral. Kategori netral memiliki performa yang sangat baik, dengan 600 prediksi benar dan tanpa kesalahan klasifikasi untuk kategori ini. Hasil ini menunjukkan bahwa meskipun model *Random Forest* efektif dalam mengenali sentimen positif dan netral, ada kebutuhan untuk perbaikan dalam mengidentifikasi sentimen negatif.

### 3.3. Visualisasi Data

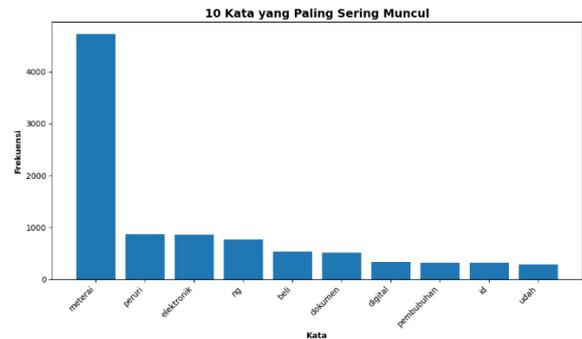
Penelitian ini menggunakan visualisasi *wordcloud* sebagai alat untuk menganalisis kumpulan data teks. Visualisasi *wordcloud* dalam analisis sentimen merupakan representasi grafis yang menampilkan kata-kata yang paling sering muncul dalam sekumpulan teks [27]. Setiap kata ditampilkan dengan ukuran font yang bervariasi, di mana kata yang lebih besar menunjukkan frekuensi kemunculan yang lebih tinggi dibandingkan dengan kata yang lebih kecil.



Gambar 5. *Wordcloud* Sentimen meterai elektronik

Kata-kata yang sering muncul dalam analisis ini meliputi “meterai”, “elektronik”, “beli meterai”, “pembubuhan meterai”, dan “kuota meterai” yang menunjukkan tingginya perhatian masyarakat terhadap isu-isu terkait penggunaan meterai, terutama dalam konteks transaksi elektronik dan peraturan yang mengaturnya. Kemunculan kata “e meterai” yang dominan mencerminkan relevansi tema ini dalam percakapan publik, sementara istilah “elektronik” menunjukkan pergeseran menuju digitalisasi dan penggunaan teknologi dalam transaksi resmi. Frasa “beli meterai” menandakan adanya permintaan yang signifikan untuk pembelian meterai, yang dapat berhubungan dengan kebijakan terbaru atau perubahan regulasi yang mendorong masyarakat untuk memahami cara mendapatkan meterai secara mudah dan efisien. Selain itu, kata

“pembubuhan meterai” dan “kuota meterai” menunjukkan bahwa masyarakat tidak hanya tertarik pada aspek transaksi itu sendiri, tetapi juga pada prosedur dan ketentuan yang harus dipatuhi dalam menggunakan meterai, mengindikasikan adanya keingintahuan dan kebutuhan untuk memahami implikasi hukum dan administratif dari penggunaan meterai dalam konteks yang lebih luas.



Gambar 6. *Wordcloud* Sentimen meterai elektronik

Dalam analisis frekuensi kata pada dataset, visualisasi *wordcloud* digunakan untuk menampilkan hasil. Sepuluh kata yang paling sering muncul dalam dataset ini adalah meterai dengan 4720 kemunculan, peruri sebanyak 870 kali, elektronik sebanyak 853 kali, ng dengan 774 kemunculan, beli muncul 537 kali, dokumen sebanyak 515 kali, digital muncul 330 kali, pembubuhan sebanyak 326 kali, id sebanyak 315 kemunculan, dan udah dengan 286 kemunculan. Hasil ini mencerminkan fokus utama pada penggunaan meterai elektronik dalam transaksi digital, yang diwakili oleh kata-kata seperti meterai, elektronik, dan digital, serta keterlibatan Peruri sebagai otoritas terkait. Selain itu, kata-kata seperti beli, dokumen, dan ya mengindikasikan proses pembelian dan pengesahan dokumen yang umum terjadi dalam konteks digital ini. Hasil ini memberikan gambaran yang jelas tentang tema-tema sentral yang dibahas dalam analisis ini.

### 4. DISKUSI

Hasil penelitian ini menunjukkan hasil analisis sentimen terkait penerapan e-meterai di Indonesia, di mana model *Random Forest* terbukti menjadi yang paling efektif dibandingkan dengan model *Naïve Bayes* dan *SVM*. Dengan akurasi rata-rata mencapai 98,48%, *Random Forest* menunjukkan kemampuannya yang superior dalam menangkap pola data dan memberikan prediksi yang lebih tepat. Kekuatan *Random Forest* terletak pada kemampuannya mengurangi *overfitting* dan menangani data hilang dengan baik, namun kelemahannya adalah memerlukan lebih banyak memori dan sulit diinterpretasikan. Sementara itu, *SVM* menempati posisi kedua dengan akurasi rata-rata sebesar 96,74%, yang menunjukkan bahwa metode ini juga memiliki kinerja yang solid dalam mengklasifikasikan sentimen. *SVM* efektif dalam

ruang berdimensi tinggi dan robust terhadap overfitting, tetapi kinerjanya sangat bergantung pada tuning parameter. Di sisi lain, model Naïve Bayes, meskipun kurang efektif dibandingkan kedua model lainnya dengan akurasi rata-rata sebesar 92,26%, tetap memiliki nilai dalam konteks analisis sentimen. Kekuatan Naïve Bayes terletak pada kesederhanaan dan kecepatan prosesnya, sementara kelemahannya adalah asumsi independensi fitur yang dapat mengurangi akurasi pada data yang saling terkait. Temuan ini menggarisbawahi pentingnya pemilihan algoritma yang tepat dalam analisis sentimen, yang dapat berkontribusi pada pengembangan kebijakan yang lebih responsif terhadap harapan dan kebutuhan masyarakat.

Jika dibandingkan dengan penelitian terdahulu, hasil ini selaras dengan beberapa studi yang menggunakan metode pembelajaran mesin untuk analisis sentimen. Misalnya, penelitian Ningsih (2024) yang menganalisis sentimen tentang kendaraan listrik di Twitter berbahasa Indonesia, menemukan bahwa SVM mengungguli Naïve Bayes dengan akurasi masing-masing 70,82% dan 63,02%. Penelitian lain oleh Ardiyansah dan Parjito (2024), yang menganalisis sentimen terhadap Luhut Binsar Pandjaitan di media sosial X, menunjukkan hasil serupa di mana SVM juga menunjukkan performa lebih baik daripada Naïve Bayes, dengan akurasi 99% dan F1-Score 99% [7].

Selain itu, penelitian oleh Miftahusalam (2022), yang mengkaji opini publik terkait penghapusan tenaga honorer di Twitter, menunjukkan bahwa Random Forest memberikan akurasi terbaik setelah menangani data imbalanced, dengan akurasi mencapai 66,67%. Hal ini memperkuat temuan dalam penelitian ini bahwa Random Forest dapat menjadi pilihan yang lebih baik dalam analisis sentimen, terutama ketika berhadapan dengan dataset yang kompleks atau imbalanced [11].

Dengan membandingkan hasil ini dengan penelitian sebelumnya, terlihat bahwa baik Naïve Bayes, SVM, maupun Random Forest memiliki keunggulan masing-masing, namun Random Forest sering kali menunjukkan performa yang lebih baik dalam berbagai skenario analisis sentimen. Penelitian ini berkontribusi dalam memperkuat pemahaman tentang efektivitas model pembelajaran mesin dalam analisis sentimen, khususnya dalam konteks kebijakan publik seperti penerapan e-meterai di Indonesia.

## 5. KESIMPULAN

Penelitian ini menunjukkan bahwa penerapan e-meterai di Indonesia telah menarik perhatian masyarakat dan memicu beragam respons yang dapat dievaluasi melalui analisis sentimen. Dengan menggunakan metode klasifikasi teks, penelitian ini membandingkan akurasi tiga model pembelajaran mesin *Naïve Bayes*, *Support Vector Machine* (SVM), dan *Random Forest* dalam menganalisis sentimen

terkait e-meterai. Dari hasil perbandingan ketiga metode, dapat disimpulkan bahwa model *Random Forest* memberikan performa terbaik dengan akurasi rata-rata sebesar 92,26%, menunjukkan kemampuan yang lebih unggul dalam menangkap pola dari data. Model SVM berada di posisi kedua dengan rata-rata akurasi 96,74%, yang menunjukkan kinerja yang stabil. Sementara itu, model *Naïve Bayes* memiliki akurasi yang lebih rendah dibandingkan keduanya yakni 92,26%, meskipun tetap menunjukkan hasil yang solid. Dengan demikian, *Random Forest* terbukti paling efektif dalam analisis sentimen terkait e-meterai, diikuti oleh SVM dan *Naïve Bayes*. Penelitian ini tidak hanya memberikan wawasan mengenai opini publik, tetapi juga menawarkan kontribusi untuk pengembangan kebijakan yang lebih responsif terhadap kebutuhan dan ekspektasi masyarakat. Dari hasil penelitian ini, pemangku kebijakan disarankan untuk menggunakan model *Random Forest* dalam pengambilan keputusan yang berkaitan dengan e-meterai, serta mempertimbangkan penerapan SVM dan *Naïve Bayes* sebagai alternatif dalam situasi di mana kecepatan dan kesederhanaan proses menjadi prioritas.

## SARAN

Disarankan untuk penelitian selanjutnya dataset yang digunakan lebih besar dan beragam, sehingga hasil analisis sentimen terhadap e-meterai menjadi lebih representatif. Selain itu, eksplorasi algoritma lain seperti deep learning dapat dipertimbangkan untuk meningkatkan akurasi model.

## DAFTAR PUSTAKA

- [1] F. Rumimpunu and M. Mangowal, "Analisis Yuridis Terhadap Meterai Elektronik Ditinjau Dari Undang-Undang Nomor 10 Tahun 2020," *J. Tek. Ind. Mesin, Elektro, dan Ilmu Komput.*, vol. 10, no. 2, pp. 48–58, 2020, doi: <https://doi.org/10.24176/simet.v10i2.3487>.
- [2] D. W. Dachmar, "Kekuatan Hukum e-Meterai Pada Dokumen Elektronik," *J. Educ. Dev.*, vol. 9, no. 1, pp. 558–562, 2021, doi: 10.37081/ed.v9i1.2322.
- [3] R. Mursyid, A. D. Indriyanti, J. Teknik, I. Fakultas, T. Universitas, and N. Surabaya, "Perbandingan Akurasi Metode Analisis Sentimen Untuk Evaluasi Opini Pengguna Pada Platform Media Sosial ( Studi Kasus : Twitter )," vol. 06, pp. 371–383, 2024.
- [4] Cucu Ika Agustyaningrum, W. Gata, Ridan Nurfalalah, Ummu Radiah, and Mawadatul Maulidiah, "Komparasi Algoritma Naive Bayes, Random Forest Dan SVM Untuk Memprediksi Niat Pembelanja Online," *J. Inform.*, vol. 20, no. 2, 2020.

- [5] F. F. Mailoa, "Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas Di Indonesia," *J. Inf. Syst. Public Heal.*, vol. 6, no. 1, p. 44, 2021, doi: 10.22146/jisph.44455.
- [6] Y. Yusra and M. Fikry, "Klasifikasi Tweet E-Commerce dengan Menggunakan Metode Support Vector Machine," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 2, p. 50, 2018, doi: 10.24014/coreit.v4i2.5205.
- [7] W. Ningsih, B. Alfianda, Rahmaddeni, and D. Wulandari, "Perbandingan Algoritma SVM dan Naive Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia," vol. 4, no. 2, pp. 556–562, 2024.
- [8] 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.
- [9] Ardiyansah and Parjito, "Perbandingan Metode Naive Bayes dan Support Vector Machine Dalam Analisis Sentimen Terhadap Tokoh Publik," *Media Online*, vol. 4, no. 6, pp. 2813–2821, 2024, doi: 10.30865/klik.v4i6.1928.
- [10] A. S. Rahayu, A. Fauzi, and R. Rahmat, "Komparasi Algoritma Naive 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.
- [11] A. Miftahusalam, A. F. Nuraini, A. A. Khoirunisa, and H. Pratiwi, "Perbandingan Algoritma Random Forest, Naive Bayes, dan Support Vector Machine Pada Analisis Sentimen Twitter Mengenai Opini Masyarakat Terhadap Penghapusan Tenaga Honorer," *Semin. Nas. Off. Stat.*, vol. 2022, no. 1, pp. 563–572, 2022, doi: 10.34123/semnasoffstat.v2022i1.1410.
- [12] D. A. Farook, M. Yusuf, and G. Syatauw, "Analisis Sentimen Terhadap Popularitas Partai Pendukung Capres 2024 Di Twitter Menggunakan Algoritma Naive Bayes Classifier," *J. Ilm. Tek. Inform.*, vol. 17, no. 2, pp. 216–226, 2024.
- [13] Y. Ansori and K. F. H. Holle, "Perbandingan Metode Machine Learning dalam Analisis Sentimen Twitter," *J. Sist. dan Teknol. Inf.*, vol. 10, no. 4, p. 429, 2022, doi: 10.26418/justin.v10i4.51784.
- [14] A. Maulana and A. Yuliana, "Analisis Sentimen Opini Publik Terkait Judi Online Pada Pengguna Aplikasi X Menggunakan Algoritma Naive Bayes Dan Support Vector Mechine," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, Oct. 2024, doi: 10.23960/jitet.v12i3S1.5187.
- [15] N. M. Farhan and B. Setiaji, "Analisis Sentimen Komentar Childfree di Aplikasi X Menggunakan Naive Bayes," *Indones. J. Comput. Sci.*, vol. 12, no. 2, pp. 284–301, 2023, doi: <https://doi.org/10.33022/ijcs.v13i3.3883>.
- [16] S. Rita and A. Pambudi, "Penggunaan Support Vector Machine Untuk Analisis Sentimen Ulasan Aplikasi Truecaller Dan Getcontact Using Support Vector Machine for Sentiment Analysis of Truecaller and Getcontact App Reviews," vol. 20, no. 2, pp. 131–140, 2023.
- [17] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.
- [18] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 9, pp. 4305–4313, 2022, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11562>
- [19] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.
- [20] W. Hidayat, M. Ardiansyah, and A. Setyanto, "Pengaruh Algoritma ADASYN dan SMOTE terhadap Performa Support Vector Machine pada Ketidakseimbangan Dataset Airbnb," *Edumatic J. Pendidik. Inform.*, vol. 5, no. 1, pp. 11–20, 2021, doi: 10.29408/edumatic.v5i1.3125.
- [21] V. Sari, F. Firdausi, and Y. Azhar, "Perbandingan Prediksi Kualitas Kopi Arabika dengan Menggunakan Algoritma SGD, Random Forest dan Naive Bayes," *Edumatic J. Pendidik. Inform.*, vol. 4, no. 2, pp. 1–9, 2020, doi: 10.29408/edumatic.v4i2.2202.
- [22] V. I. Santoso, G. Virginia, and Y. Lukito, "Penerapan Sentiment Analysis Pada Hasil Evaluasi Dosen Dengan Metode Support Vector Machine," *J. Transform.*, vol. 14, no.

- 2, p. 72, 2017, doi: 10.26623/transformatika.v14i2.439.
- [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] 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.
- [25] Sri Diantika, Windu Gata, and Hiya Nalatisifa, "Komparasi Algoritma SVM Dan Naive Bayes Untuk Klasifikasi Kestabilan Jaringan Listrik," *Elkom J. Elektron. dan Komput.*, vol. 14, no. 1, pp. 10–15, 2021, doi: 10.51903/elkom.v14i1.319.
- [26] 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.
- [27] A. Sudrajat, N. Mulyani, and N. Marpaung, "Sistem Pendukung Keputusan Penentuan Kelayakan Penangguhan Kredit Nasabah menggunakan Naïve Bayes," *Edumatic J. Pendidik. Inform.*, vol. 6, no. 2, pp. 205–214, 2022, doi: 10.29408/edumatic.v6i2.6298.