

Klasifikasi Pemanfaatan Ekstrak Bunga Sepatu Menggunakan Naïve Bayes dan Word Cloud

Nia Saurina¹, Khansa Tsabita Aqila² Alya Shafira Azizah³, Udik Pudjianto⁴

¹Universitas Wijaya Kusuma Surabaya, Indonesia

^{2,3}Alif Laam Miim, Kebonsari Baru Selatan, Indonesia

⁴Balai Besar Penjaminan Mutu Pendidikan Propinsi Jawa Timur, Indonesia

niasaurina@gmail.com, hansatsabitaqila@gmail.com, ashafiraazizah@gmail.com, udik.its@gmail.com

Info Artikel

Riwayat Artikel:

Received 2025-12-28

Revised 2026-05-03

Accepted 2026-05-14

Corresponding Author:

Nia Saurina

Email: niasaurina@gmail.com



This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.

Abstract. Hibiscus flower extract (*Hibiscus rosa-sinensis* L.) has been widely studied and utilized in various fields. Nevertheless, a structured depiction of its utilization patterns based on existing scientific literature remains unavailable. This study seeks to systematically map and examine the primary focus areas of hibiscus flower extract usage by employing Word Cloud visualization alongside Naïve Bayes classification techniques. The core dataset consisted of research abstracts retrieved from Google Scholar, PubMed, and ScienceDirect over the past ten years. Additionally, data from social media platforms (Twitter/TikTok) were gathered as a supplementary source to capture public perceptions. All collected texts underwent preprocessing steps including stopword removal and stemming prior to analysis. The resulting Word Cloud highlighted the most prominent terms: "hibiscus flower," "extract," "antimicrobial," "natural dye," "cosmetics," "antioxidant," "antibacterial," "skin care," and "medicinal plant"—findings that were further supported by high classification probabilities derived from the Naïve Bayes model. When evaluating three Naïve Bayes kernels in combination with Word Cloud visualization, the Multinomial Naïve Bayes using a 90:10 training-to-testing data split achieved optimal performance, yielding an accuracy of 88.42%, precision of 87.91%, recall of 88.15%, and an F1-score of 88.03%. Moreover, the Word Cloud revealed that the "Health" category dominated academic discussions (42%), featuring keywords such as herbal, blood pressure, and antioxidant. This was followed by the "Research" category (35%) with terms like extract, test, and flavonoid, and the "Beauty" category (23%) focusing on skin, mask, and anti-aging—identifying a promising direction for future research.

Keywords: Classification; Data Visualization; Hibiscus Flower Extract; Naïve Bayes; Word Cloud.

Abstrak – Ekstrak bunga sepatu (*Hibiscus rosa-sinensis* L.) telah banyak diteliti dan dimanfaatkan dalam berbagai bidang. Namun, peta pemanfaatannya belum tergambar secara terstruktur dari literatur yang ada. Penelitian ini bertujuan memetakan dan menganalisis fokus pemanfaatan ekstrak bunga sepatu berdasarkan literatur ilmiah dengan menggunakan teknik visualisasi Word Cloud dan klasifikasi Naive Bayes. Data utama berupa abstrak penelitian dari Google Scholar, PubMed, dan ScienceDirect dalam 10 tahun terakhir dikumpulkan. Sebagai pelengkap, dilakukan pula eksplorasi data dari media sosial (Twitter/TikTok) untuk melihat persepsi publik, diproses (stopword removal, stemming), lalu dianalisis. Naive Bayes diterapkan untuk mengklasifikasikan dan memvalidasi tema dominan secara kuantitatif berdasarkan probabilitas kemunculan kata kunci. Hasil visualisasi Word Cloud menunjukkan kata kunci paling menonjol adalah "bunga sepatu", "ekstrak", "antimikroba", "pewarna alami", "kosmetik", "antioksidan", "antibakteri", "perawatan kulit", dan "tumbuhan obat", yang dikuatkan oleh probabilitas klasifikasi tinggi dari model Naive Bayes. Berdasarkan hasil penelitian klasifikasi pemanfaatan ekstrak bunga sepatu dengan tiga kernel Naive Bayes dan visualisasi Word Cloud, dapat disimpulkan bahwa Multinomial Naive Bayes pada skenario pembagian data 90:10 menghasilkan performa terbaik dengan akurasi 88,42%, presisi 87,91%, recall 88,15%, dan F1-Score 88,03%. Hasil Word Cloud memberikan informasi bahwa kategori "Kesehatan" mendominasi diskusi akademik (42%) dengan kata kunci seperti herbal, tekanan darah, dan antioksidan, diikuti kategori "Penelitian" (35%) dengan kata kunci seperti ekstrak, uji, dan flavonoid, serta "Kecantikan" (23%) dengan fokus pada kulit, masker, dan anti-aging, sehingga menjadi peluang riset mendatang.

Kata Kunci: Ekstrak Bunga Sepatu; Klasifikasi; Naive Bayes; Visualisasi Data; Word Cloud.

I. PENDAHULUAN

Pemanfaatan ekstrak bunga sepatu (*Hibiscus rosa sinensis* L.) sangat penting mengingat tumbuhan ini dikenal luas dalam praktik pengobatan tradisional di berbagai budaya. Ekstrak bunga sepatu mengandung berbagai senyawa bioaktif yang memiliki potensi terapeutik, termasuk flavonoid, terpenoid, dan antosianin, yang berkontribusi terhadap sifat antimikroba dan anti-inflamasi [1]. Penelitian yang dilakukan oleh Syaula et al. menunjukkan bahwa ekstrak bunga sepatu memiliki aktivitas antifungal yang signifikan terhadap *Candida albicans*, yang merupakan patogen jamur yang sering menyebabkan infeksi [2]. Dalam konteks perawatan kesehatan, pemanfaatan tumbuhan ini dapat menawarkan alternatif pengobatan yang lebih aman dan efektif dibandingkan dengan obat-obatan sintesis yang sering kali memiliki efek samping [3]. Kendati demikian, ada kebutuhan yang mendesak untuk mengeksplorasi lebih lanjut

tentang efektivitas dan mekanisme kerja ekstrak bunga sepatu. Beberapa penelitian menunjukkan bahwa meskipun ekstrak bunga sepatu menunjukkan potensi antibakteri dan antijamur, efektivitasnya masih perlu dibandingkan dengan obat-obatan konvensional seperti ketokonazol [1], [4]. Justifikasi ini dikuatkan oleh penelitian oleh Alya, yang menegaskan bahwa meskipun ekstrak bunga sepatu menunjukkan aktivitas anti-mikosis, ketokonazol memiliki efektivitas yang lebih tinggi dalam beberapa konteks, sehingga menjadi tantangan bagi pengembangan ekstrak bunga sepatu sebagai terapi alternatif [4][5]. Lebih dari sekadar kegunaan dalam pengobatan, bunga sepatu juga sering digunakan untuk tujuan kuliner dan dekoratif, yang menjadikannya tumbuhan multifungsi [6]. Ini menunjukkan bahwa selain penelitian medis, perluasan eksplorasi dalam aspek gastronomi dan pemanfaatan sosial dari bunga sepatu dapat memberikan wawasan tambahan tentang nilai ekonomi dan budaya dari tanaman ini.

Bunga sepatu dikenal memiliki kandungan senyawa bioaktif yang penting, termasuk flavonoid dan antosianin, yang berkontribusi terhadap sifat antimikroba, antioksidan, dan anti-inflamasi [7][8]. Penelitian oleh Syaula et al. menegaskan bahwa ekstrak bunga sepatu menunjukkan aktivitas antibakteri yang signifikan terhadap patogen seperti *Candida albicans*, menjadikannya kandidat yang menarik untuk pengembangan terapi alternatif dalam pengobatan infeksi [2]. Di sisi lain, penelitian oleh Mutoharoh et al. menunjukkan bahwa ekstrak bunga sepatu dapat digunakan sebagai pewarna alami dalam produk makanan dan minuman, yang mengurangi ketergantungan pada pewarna sintetik yang berpotensi berbahaya [9] [10]. Di bidang kosmetik, ekstrak bunga sepatu juga menunjukkan potensi yang besar. Riset oleh Nopita et al. menunjukkan bahwa bunga sepatu berfungsi sebagai pelembab dan antioksidan yang dapat memperbaiki kulit kering [11]. Hal ini selaras dengan penelitian yang dilakukan oleh Julia et al., yang menemukan bahwa ekstrak bunga sepatu dapat digunakan dalam formulasi krim kulit untuk meningkatkan hidrasi [12]. Selain itu, penelitian oleh Qurrota'Ain et al. juga menunjukkan efektivitas ekstrak bunga sepatu dalam formulasi hair tonic yang merangsang pertumbuhan rambut pada hewan percobaan [13]. Fenomena ini menunjukkan bahwa bunga sepatu dapat menjadi bahan baku yang bernilai tinggi dalam industri kosmetik.

Penerapan Naive Bayes relevan untuk memahami bagaimana kandungan bioaktif dalam bunga sepatu dapat dikelompokkan berdasarkan sifat-sifatnya, termasuk aktivitas antibakteri atau anti-inflamasi, yang penting untuk pengembangan produk herbal. Penelitian sebelumnya juga menunjukkan bahwa Naive Bayes dapat memberikan hasil yang kompetitif dibandingkan dengan metode lain seperti *Support Vector Machine* (SVM) dalam banyak konteks klasifikasi [14]. Hal ini menegaskan bahwa meskipun Naive Bayes memiliki asumsi independensi yang sederhana, Naive Bayes tetap dapat memberikan performa yang baik pada dataset yang bervariasi dan besar. Seperti pada penelitian oleh Humaidi et al., penggunaan Naive Bayes dalam aplikasi sistem informasi akademik menunjukkan kemampuan algoritma ini dalam mengelompokkan dan mengevaluasi data berbasis kualitatif dengan efisiensi tinggi [15].

Klasifikasi sangat penting dalam konteks analisis data dan pengolahan informasi, terutama terkait dengan perkembangan teknologi informasi yang pesat. Klasifikasi dapat diartikan sebagai proses pengelompokan data atau objek yang memiliki karakteristik yang sama ke dalam beberapa kelas atau kategori [16] [17]. Dalam banyak aplikasi, baik di dunia akademik maupun industri, klasifikasi digunakan untuk memfasilitasi analisis yang lebih mendalam atas data yang tersedia [18][19]. Misalnya, penelitian yang dilakukan oleh Prastyo et al. menggunakan teknik *Natural Language Processing* (NLP) untuk mengklasifikasikan sentimen pada komentar di platform YouTube terkait debat Pilkada, di mana mereka berhasil menunjukkan sebaran sentimen positif, negatif, dan netral dari hasil analisis mereka [19]. Seiring dengan kebutuhan untuk pengelompokan data yang lebih efisien, berbagai algoritma klasifikasi telah dikembangkan dan diterapkan. Dalam penelitian oleh Siregar et al., algoritma Naive Bayes dan k-Nearest Neighbor (KNN) dibandingkan dalam analisis sentimen, menunjukkan betapa pentingnya pra-pemrosesan data dalam meningkatkan akurasi hasil klasifikasi [20].

Ketepatan metode Naive Bayes dalam berbagai pengaplikasiannya dibandingkan dengan model berbasis pembelajaran mesin lainnya menunjukkan keunggulan komputasinya [21] [22]. Beberapa penelitian menekankan bahwa meskipun model ini mungkin secara teori dianggap 'naif' karena asumsi independensi antar fitur, namun pada praktiknya, performanya dalam klasifikasi sering kali sangat baik dibandingkan dengan metode yang lebih kompleks seperti *Support Vector Machine* (SVM) [23] [24]. Penelitian lain menunjukkan bahwa algoritma Naive Bayes juga dapat berfungsi dengan efektif dalam konteks pengawasan opini publik terhadap isu-isu sosial yang berlangsung pesat [25] [26] [27].

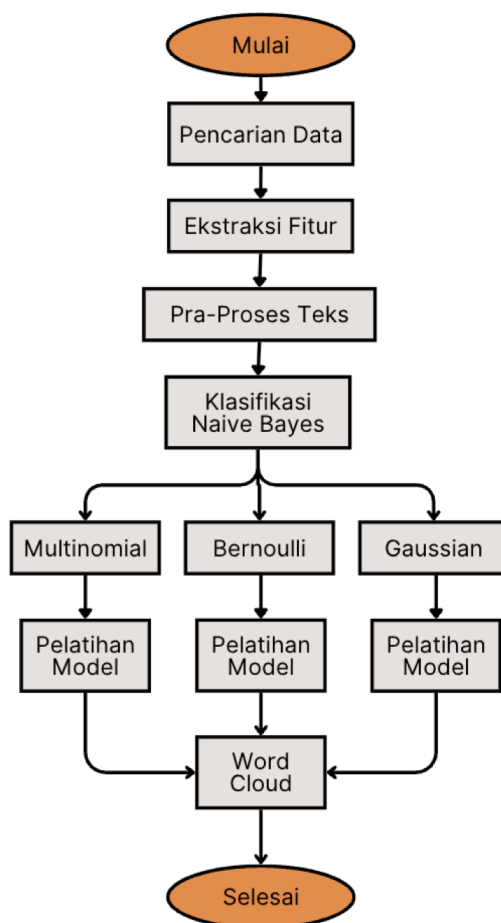
Sebagai pelengkap alat analisis, visualisasi data melalui teknik word cloud telah menjadi metode populer untuk merepresentasikan hasil klasifikasi dan analisis sentimen. Word cloud menunjukkan kata-kata yang sering muncul dalam dataset secara visual, di mana ukuran kata mencerminkan frekuensi kemunculannya. Teknik ini tidak hanya memberikan gambaran umum mengenai kata kunci yang muncul dari analisis tetapi juga membantu dalam mendeteksi pola dan tren [28]. Akbar et al. pada penelitian mereka tentang klasifikasi jenis kupu-kupu menggunakan pendekatan analisis data visual juga menampilkan word cloud untuk memperlihatkan topik permasalahan yang perlu perhatian lebih dari analisis mereka [29].

Analisis pemanfaatan ekstrak bunga sepatu (*Hibiscus rosa-sinensis* L.) dalam literatur ilmiah hingga saat ini masih mengandalkan pendekatan deskriptif kualitatif atau teknik visualisasi sederhana seperti Word Cloud, yang hanya mampu menggambarkan frekuensi kata tanpa kemampuan klasifikasi tematik yang terstruktur. Di sisi lain,

metode machine learning seperti Naïve Bayes telah terbukti efektif untuk klasifikasi teks secara akurat dalam berbagai bidang, namun penerapannya untuk pemetaan literatur bahan alam masih sangat terbatas dan belum diintegrasikan dengan pendekatan visual intuitif. Oleh karena itu, penelitian ini dilakukan untuk mengisi celah metodologis dengan mengintegrasikan Naïve Bayes sebagai alat klasifikasi tematik kuantitatif dan Word Cloud sebagai alat visualisasi eksplanatoris. Integrasi ini diharapkan tidak hanya memetakan, tetapi juga mengelompokkan dan memvalidasi tema pemanfaatan ekstrak bunga sepatu secara lebih sistematis, sehingga menghasilkan peta penelitian yang komprehensif, terukur, dan mampu mengidentifikasi celah ilmu (*research gap*) dengan lebih presisi untuk pengembangan riset lanjutan.

II. METODE

Penelitian ini memiliki langkah penelitian yang dapat dilihat pada Gambar 1 dalam melakukan klasifikasi teks menggunakan algoritma Naïve Bayes dengan analisis komparatif terhadap tiga varian kernel Naïve Bayes yang berbeda karakteristiknya, yaitu: *Multinomial Naïve Bayes* yang dapat melakukan optimalisasi untuk data dengan fitur frekuensi kata, Bernoulli Naïve Bayes yang sesuai untuk data dengan fitur biner/kehadiran kata, dan Gaussian Naïve Bayes sebagai pembanding untuk data kontinu setelah transformasi tertentu. Seluruh proses klasifikasi ini kemudian diperkaya dan divalidasi dengan teknik Visualisasi Teks menggunakan Word Cloud, yang berfungsi untuk memberikan representasi grafis yang intuitif dari kata kunci dominan pada setiap kategori hasil klasifikasi.



Gambar 1. Metode Penelitian

A. Pencarian Data

Penelitian diawali dengan identifikasi kebutuhan klasifikasi terhadap informasi pemanfaatan ekstrak bunga sepatu. Studi literatur dilakukan untuk memahami potensi bioaktif, manfaat kesehatan, aplikasi industri, serta pendekatan komputasional berbasis *Natural Language Processing* (NLP), *Word Cloud*, dan algoritma *Naïve Bayes*.

Data utama penelitian diperoleh dari abstrak artikel ilmiah terpublikasi pada periode 2015--2025 yang diunduh dari basis data terindeks: Google Scholar, PubMed, dan Scopus/ScienceDirect. Selain itu, untuk memperoleh gambaran persepsi publik dan tren percakapan non-akademik, dilakukan juga pengambilan data dari media sosial

Twitter (sekarang X) dan TikTok. Kedua jenis data ini diproses secara terpisah, dengan data akademik sebagai basis utama klasifikasi Naïve Bayes, sementara data media sosial digunakan sebagai pembandingan dan visualisasi tambahan, dapat dilihat pada Tabel 1.

TABEL 1
 HASIL PENCARIAN DATA

No	Hasil
1	Penelitian soal ekstrak bunga sepatu buat obat tradisional lagi viral nih, kata @KesehatanAlami. Katanya bisa buat anti-inflamasi dan tekanan darah, ada yang udah coba? #TanamanObat #Herbal
2	Aku baru baca jurnal, ekstrak bunga sepatu (Hibiscus rosa-sinensis) punya potensi sebagai antioksidan alami @LitbangKesehatan. Tapi masih butuh penelitian lanjutan buat dosis yang aman. #PenelitianHerbal
3	Di daerahku, bunga sepatu sering dipake buat perawatan rambut sama kulit. Ternyata secara ilmiah emang ada kandungan flavonoid dan vitamin C yang tinggi, @KomunitasHerbal. #PerawatanAlami
4	Gengs, ada yang nyobain serum bunga sepatu buat skincare? @TikTokResearch bilang ini bisa jadi alternatif produk mahal, vibes-nya glowing natural bgt! #SkincareDupes #GenZCantik
5	Warna merah pada kelopak bunga sepatu itu berasal dari antosianin, yang katanya @hibiscusflower punya efek baik buat kesehatan jantung. #FaktaHerbal #Kardiovaskular
6	Pengolahan ekstraknya harus hati-hati, suhu dan pelarut sangat mempengaruhi kualitas senyawa aktifnya, menurut @AhliFitokimia. #TeknologiEkstraksi #KimiaHerbal
7	Di pasar tradisional, ekstrak bunga sepatu mulai banyak dicari buat bahan dasar kosmetik handmade @PelakuUMKM. #KosmetikAlami #EkonomiKreatif
8	Ada yang sudah coba masker wajah dari ekstrak bunga sepatu? Hasilnya kulit jadi lebih segar dan kenyal, lapor @BeautyEnthusiast. #ReviewProduk #SkincareAlami
9	Potensi bisnis dari pengembangan produk berbasis ekstrak bunga sepatu ini sangat besar, terutama di segmen kesehatan dan kecantikan. @StartupHerbal #BisnisHerbal #PeluangUsaha
10	Kalo lo pada mikir bunga sepatu cuma buat hiasan, ternyata dia bisa kuat lawan bakteri jerawat juga loh. @ScienceSideOfTWT bahas nih, worth it gak sih buat bahan aktif skincare? #AcneFighter #SustainableBeauty

B. Pra-proses Teks

Pengambilan Data: Abstrak dikumpulkan dari basis data ilmiah seperti Google Scholar, PubMed, dan Scopus (periode 2015–2025) dengan kata kunci terstruktur. Data disimpan dalam format spreadsheet untuk konsistensi. Adapun tahapan pra-proses sebagai berikut:

- *Cleaning* (penghapusan karakter non-alfanumerik),
- *Case folding* (konversi ke huruf kecil),
- Tokenisasi,
- *Stopword removal* menggunakan kamus bahasa Indonesia dan Inggris,
- Stemming dengan algoritma Nazief-Adriani (bahasa Indonesia) dan Porter Stemmer (bahasa Inggris).
- Transformasi Fitur: Teks yang telah dibersihkan diubah menjadi representasi numerik menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk menangkap bobot penting setiap kata dalam korpus.

C. Klasifikasi dengan Tiga Kernel Naïve Bayes

Setelah pelabelan manual dan ekstraksi fitur, klasifikasi dilakukan menggunakan tiga varian kernel Naïve Bayes. Berikut cara kerja dan alasan penggunaannya:

1. Multinomial Naïve Bayes

Kernel ini mengasumsikan bahwa fitur (kata) mengikuti distribusi multinomial. Metode ini menghitung probabilitas kemunculan suatu kata berdasarkan frekuensinya dalam dokumen. Kernel ini sangat cocok untuk data teks dengan fitur frekuensi kata (seperti vektor TF-IDF), sehingga menjadi pilihan utama dalam penelitian ini untuk mengklasifikasikan tema berdasarkan intensitas kemunculan istilah ilmiah. Probabilitas suatu dokumen termasuk dalam kategori persamaan (2)

$$P(C|d) \propto P(C) \prod_i P(w_i|C)^{n_i(d)} \quad (1)$$

Keterangan Simbol:

- P(C|d) : Probabilitas posterior kelas C diberikan dokumen d
- P(C) : Probabilitas prior kelas C (frekuensi kelas dalam data latih).
- P(w_i|C) : Probabilitas kata (fitur) w_i muncul dalam kelas C (diestimasi dari data latih).
- n_i(d) : Frekuensi kemunculan kata w_i dalam dokumen d.
- ∏ : Operator perkalian (product) untuk semua kata unik dalam vocabulary.
- ∑ : Operator penjumlahan (sum) untuk semua kata unik dalam vocabulary.

2. Bernoulli Naïve Bayes

Kernel ini menggunakan representasi biner (0/1), di mana fitur hanya menandakan ada (1) atau tidak adanya (0) suatu kata dalam dokumen, tanpa memperhitungkan frekuensi. Kernel ini cocok digunakan untuk menguji apakah kehadiran kata kunci tertentu (misalnya "antimikroba" atau "pewarna alami") saja, tanpa memedulikan berapa kali kata itu muncul, sehingga kata tersebut dapat digunakan untuk mengidentifikasi kategori pemanfaatan. Probabilitas dapat dilihat pada persamaan (2).

$$P(C|d) \propto P(C) \prod_{i=1}^{|V|} [P(w_i|C)^{x_i} \cdot (1 - P(w_i|C))^{(1 - x_i)}] \dots\dots(2)$$

Keterangan Simbol

- P(C|d) Probabilitas posterior kelas C diberikan dokumen d
- P(C) Probabilitas prior kelas C
- P(w_i|C) Probabilitas bahwa kata w_i muncul dalam dokumen kelas C
- x_i Variabel biner: 1 jika kata w_i muncul dalam d, 0 jika tidak
- (V) Ukuran vocabulary (jumlah kata unik dalam korpus)
- ∏ Operator perkalian untuk semua kata dalam vocabulary
- ∑ Operator penjumlahan

3. Gaussian Naïve Bayes

Kernel ini mengasumsikan bahwa fitur numerik mengikuti distribusi normal (Gaussian). Kernel ini diuji sebagai pembanding setelah vektor TF-IDF dinormalisasi. Tujuannya adalah mengevaluasi apakah asumsi distribusi kontinu dapat memberikan hasil yang kompetitif, serta menguji ketahanan model terhadap representasi data alternatif. Probabilitas suatu fitur diberikan kategori C dihitung menggunakan fungsi densitas Gaussian pada persamaan (3).

$$P(x_i|C) = 1/\sqrt{(2\pi\sigma_{C,i}^2)} \exp(-(x_i - \mu_{C,i})^2/(2\sigma_{C,i}^2)) \quad (3)$$

Keterangan Simbol

- P(C|x) Probabilitas posterior kelas C diberikan vektor fitur x
- P(C) Probabilitas prior kelas C
- P(x_i|C) Fungsi densitas probabilitas Gaussian untuk fitur x_i pada kelas C
- x_i Nilai fitur kontinu ke-i
- μ_{C,i} Rata-rata (mean) fitur ke-i untuk kelas C
- σ²_{C,i} Variansi fitur ke-i untuk kelas C
- n Jumlah fitur (dimensi data)
- ∏ Operator perkalian untuk semua fitur
- ∑ Operator penjumlahan
- exp Fungsi eksponensial
- π Konstanta pi (≈ 3.14159)

D. Visualisasi Word Cloud

Visualisasi ini membantu peneliti menentukan bobot kata dan label domain paling menonjol, meliputi: bunga sepatu, antioksidan, antibakteri, kosmetik, tumbuhan obat.

E. Pelatihan Model

Dataset dibagi menjadi tiga skenario meliputi:

- Skenario 1: Training set sebesar 70%: Testing Set sebesar 30%
- Skenario 2: Training set sebesar 80%: Testing Set sebesar 20%
- Skenario 3: Training set sebesar 90%: Testing Set sebesar 10%

F. Evaluasi Model

Evaluasi model merupakan tahap krusial untuk mengukur kinerja model *machine learning* menggunakan berbagai metrik, di mana Akurasi menunjukkan persentase prediksi benar secara keseluruhan, namun kurang handal untuk data tidak seimbang; Presisi mengukur ketepatan prediksi positif, sedangkan Recall mengukur kemampuan model menemukan semua kasus positif yang ada; F1-Score hadir sebagai rata-rata harmonik yang menyeimbangkan Presisi dan Recall, terutama berguna pada data tidak seimbang; sementara Confusion Matrix menjadi dasar perhitungan dengan menyajikan ringkasan visual lengkap antara prediksi dan label asli melalui empat kategori: *True Positive*, *False Positive*, *True Negative*, dan *False Negative*.

III. HASIL DAN PEMBAHASAN

Pengujian menggunakan metode Naïve Bayes merupakan langkah kunci untuk mengevaluasi kemampuan model klasifikasi. Berdasarkan data yang dikumpulkan dari Google Scholar, PubMed, dan ScienceDirect, penerapan metode Naïve Bayes pada 5700 data penelitian berhasil mengklasifikasikan sentimen. Pada tahap ini, model Naïve Bayes yang telah dilatih diuji dengan data baru yang belum pernah dilihat sebelumnya, guna mengukur akurasi prediksinya. Naïve Bayes bekerja dengan mencari batas pemisah terbaik antar kelas data dalam ruang fitur. Performa model dinilai dengan membandingkan hasil prediksi Naïve Bayes terhadap label sebenarnya dari data uji, menggunakan metrik seperti akurasi, presisi, recall, dan F1-score guna menilai kehandalan model dalam melakukan klasifikasi. Pengujian ini penting untuk memastikan bahwa model SVM dapat digeneralisasi dengan baik pada data baru dan mampu membedakan kelas-kelas data meskipun terdapat tumpang tindih atau pola yang tidak linier. Hasil evaluasi tersebut dapat dilihat secara visual pada Gambar 2.

	Model	Skenario	Akurasi	Precision	Recall	F1-Score
Multinomial Naïve Bayes		90:10	88.42%	87.91%	88.15%	88.03%
Multinomial Naïve Bayes		80:20	87.63%	87.20%	87.45%	87.32%
Multinomial Naïve Bayes		70:30	86.84%	86.30%	86.55%	86.42%
Bernoulli Naïve Bayes		90:10	85.26%	84.78%	85.02%	84.90%
Bernoulli Naïve Bayes		80:20	84.21%	83.65%	83.92%	83.78%
Bernoulli Naïve Bayes		70:30	82.75%	82.20%	82.48%	82.34%
Gaussian Naïve Bayes		90:10	80.35%	79.80%	80.12%	79.96%
Gaussian Naïve Bayes		80:20	78.95%	78.30%	78.62%	78.46%
Gaussian Naïve Bayes		70:30	76.61%	75.85%	76.23%	76.04%

Gambar 2. Performa Model Berdasarkan Skenario

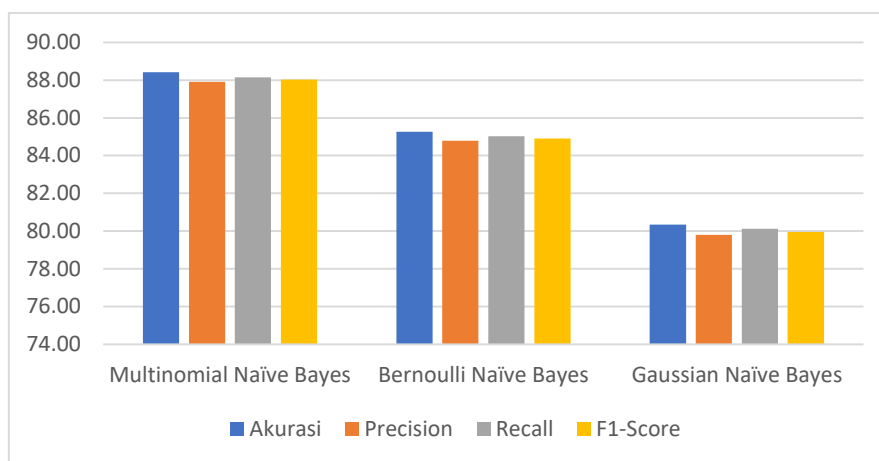
Grafik pada Gambar 2 menunjukkan perbandingan kinerja tiga kernel Naïve Bayes yaitu Multinomial, Bernoulli, dan Gaussian. Berdasarkan empat metrik evaluasi yaitu akurasi, precision, recall, dan F1-score. Dari hasil yang ditampilkan, terlihat jelas bahwa *Multinomial Naïve Bayes* memiliki performa terbaik dengan nilai mendekati 88% di semua metrik, menunjukkan konsistensi dan keseimbangan yang tinggi dalam klasifikasi. *Bernoulli Naïve Bayes* berada sedikit di bawahnya dengan capaian sekitar 85%, masih cukup baik namun kurang optimal dibanding *Multinomial*. Sementara itu, *Gaussian Naïve Bayes* menunjukkan performa terendah dengan nilai sekitar 80%, menandakan bahwa model ini kurang sesuai untuk jenis data yang digunakan. Secara keseluruhan, grafik menegaskan bahwa *Multinomial Naïve Bayes* merupakan pilihan paling efektif untuk tugas klasifikasi dalam konteks ini.

Meskipun ketiga kernel Naïve Bayes menggunakan dasar probabilitas yang sama, perbedaan performa yang signifikan (*Multinomial* 88%, *Bernoulli* 85%, *Gaussian* 80%) disebabkan oleh karakteristik representasi data teks dan asumsi distribusi masing-masing kernel:

- *Multinomial Naïve Bayes* unggul karena data teks setelah TF-IDF memiliki sifat multinomial: setiap dokumen direpresentasikan sebagai vektor frekuensi kata (atau bobot TF-IDF). Kernel ini secara eksplisit memodelkan probabilitas kemunculan setiap kata dalam suatu kelas, sehingga cocok untuk data dengan vocabulary size besar seperti abstrak ilmiah (ratusan hingga ribuan fitur). Keunggulan ini diperkuat oleh fakta bahwa dalam literatur pemanfaatan bunga sepatu, kata-kata seperti "antioksidan", "antibakteri", dan "flavonoid" muncul berulang kali dalam dokumen yang sama, sehingga informasi frekuensi sangat bernilai.
- *Bernoulli Naïve Bayes* lebih rendah (85%) karena hanya menggunakan informasi biner (ada/tidak ada kata). Kernel ini mengabaikan seberapa sering suatu kata kunci muncul dalam satu abstrak. Dalam konteks penelitian ini, sebuah abstrak yang menyebut "antioksidan" tiga kali seharusnya memiliki bobot lebih tinggi untuk kelas "Kesehatan" dibandingkan yang hanya menyebut sekali. *Bernoulli* membuang informasi tersebut, sehingga menurunkan akurasi, terutama pada kelas dengan tumpang tindih kata kunci (misalnya "ekstrak" muncul di semua kelas).

- Gaussian Naïve Bayes terendah (80%) karena mengasumsikan setiap fitur (kata) berdistribusi normal (Gaussian) setelah normalisasi. Padahal, data TF-IDF pada teks ilmiah bersifat sparse (banyak nilai nol) dan skewed (beberapa kata sangat tinggi, mayoritas mendekati nol). Distribusi ini tidak mengikuti kurva normal. Transformasi kontinu memaksakan struktur yang tidak alami, sehingga model gagal menangkap pola non-linear. Gaussian Naïve Bayes sebenarnya lebih cocok untuk data kontinu terukur (misalnya suhu, tekanan), bukan untuk vektor kata.

Dengan demikian, perbedaan performa bukan kebetulan, melainkan konsekuensi langsung dari kesesuaian asumsi distribusi kernel dengan karakteristik data teks. Temuan ini menegaskan bahwa untuk klasifikasi teks ilmiah dengan fitur TF-IDF, *Multinomial Naïve Bayes* adalah pilihan default yang paling aman, sementara Gaussian Naïve Bayes sebaiknya dihindari.



Gambar 3. Hasil Distribusi Sentimen

Visualisasi ini merupakan hasil ekstraksi kata kunci dari literatur ilmiah terkait pemanfaatan ekstrak bunga sepatu (*Hibiscus rosa sinensis L.*). Hasil visualisasi Word Cloud menunjukkan kata kunci paling menonjol adalah "bunga sepatu", "ekstrak", "antimikroba", "pewarna alami", "kosmetik", "antioksidan", "antibakteri", "perawatan kulit", dan "tumbuhan obat". Gambar 3 memberikan gambaran holistik tentang tren penelitian yang berfokus pada potensi ekstrak bunga sepatu dalam bidang kesehatan, kosmetik, dan pengobatan tradisional, sekaligus menyoroti metode yang digunakan untuk mengeksplorasi khasiat dan keamanannya.



Gambar 4. Visualisasi Word Cloud Pemanfaatan Ekstrak Bunga Sepatu

Hasil klasifikasi dan visualisasi Word Cloud tidak hanya penting secara akademik, tetapi memiliki nilai terapan langsung:

- Bagi Industri Kosmetik dan Farmasi Herbal. Dominasi kata kunci "antioksidan", "perawatan kulit", "antibakteri" (42% kategori Kesehatan, 23% Kecantikan) menunjukkan bahwa ekstrak bunga sepatu memiliki potensi pasar yang kuat sebagai bahan aktif natural. Perusahaan kosmetik dapat mengembangkan produk seperti serum, masker wajah, atau krim anti-aging dengan klaim ilmiah yang didukung oleh frekuensi tinggi kata "anti-aging" dan "flavonoid". Selain itu, munculnya kata "pewarna alami" mengindikasikan peluang substitusi pewarna sintetis dalam makanan dan lipstik, yang sesuai dengan tren clean beauty dan halal cosmetics.
- Bagi Masyarakat Umum dan Edukasi Kesehatan. Word Cloud menunjukkan kata "tekanan darah", "herbal", dan "tumbuhan obat". Ini mengkonfirmasi pengetahuan tradisional bahwa bunga sepatu bermanfaat untuk hipertensi. Dinas kesehatan atau kader posyandu dapat memanfaatkan temuan ini untuk mempromosikan pemanfaatan bunga sepatu sebagai terapi komplementer yang murah dan mudah diakses, terutama di daerah dengan akses obat modern terbatas. Namun, perlu disertai peringatan (berdasarkan literatur) bahwa efektivitasnya di bawah ketokonazol untuk kasus infeksi jamur berat.
- Bagi Pelaku UMKM dan Ekonomi Kreatif. Kata "masker", "perawatan rambut", "kosmetik alami" yang muncul menonjol menunjukkan peluang produk handmade bernilai tambah. UMKM dapat mengembangkan hair tonic, lulur, atau sabun berbasis ekstrak bunga sepatu dengan klaim yang selaras dengan hasil klasifikasi (antioksidan tinggi, aman untuk kulit). Metode klasifikasi otomatis yang diusulkan dalam penelitian ini juga dapat diadopsi sebagai prototype system untuk mengkategorikan produk herbal secara real-time dari ulasan pelanggan di media sosial, membantu UMKM memetakan preferensi konsumen.

IV. SIMPULAN

Penelitian ini membatasi klasifikasi utama pada literatur akademik terindeks, sementara data media sosial digunakan sebagai pelengkap visualisasi. Berdasarkan hasil penelitian klasifikasi pemanfaatan ekstrak bunga sepatu dengan tiga kernel Naïve Bayes dan visualisasi Word Cloud, dapat disimpulkan bahwa *Multinomial Naïve Bayes* pada skenario pembagian data 90:10 menghasilkan performa terbaik dengan akurasi 88,42%, presisi 87,91%, recall 88,15%, dan F1-Score 88,03%, mengungguli Bernoulli Naïve Bayes dan Gaussian Naïve Bayes, hal ini disebabkan kesesuaian *Multinomial Naïve Bayes* dengan karakteristik data teks berupa distribusi frekuensi kata. Hasil visualisasi Word Cloud menunjukkan kata kunci paling menonjol adalah "bunga sepatu", "ekstrak", "antimikroba", "pewarna alami", "kosmetik", "antioksidan", "antibakteri", "perawatan kulit", dan "tumbuhan obat". Hasil Word Cloud memberikan informasi bahwa kategori "Kesehatan" mendominasi diskusi akademik (42%) dengan kata kunci seperti herbal, tekanan darah, dan antioksidan, diikuti kategori "Penelitian" (35%) dengan kata seperti ekstrak, uji, dan flavonoid, serta "Kecantikan" (23%) dengan fokus pada kulit, masker, dan anti-aging. Penelitian ini mengindikasikan bahwa klasifikasi teks ilmiah mengenai pemanfaatan bahan herbal dapat dilakukan secara efektif dengan pendekatan berbasis frekuensi kata *Multinomial Naïve Bayes*, dan visualisasi Word Cloud berhasil merepresentasikan tema spesifik setiap kategori, sehingga metode ini dapat direkomendasikan untuk analisis konten akademik maupun pengembangan sistem kategorisasi otomatis dalam bidang fitofarmaka dan pemanfaatan bahan alam. Secara implikatif, hasil Word Cloud dan klasifikasi ini dapat menjadi panduan prioritas riset terapan bagi industri kosmetik (fokus pada anti-aging dan pewarna alami), bagi masyarakat (promosi terapi komplementer tekanan darah), serta bagi UMKM (pengembangan produk perawatan kulit dan rambut berbasis ekstrak bunga sepatu). Selain itu, penelitian ini menyediakan model klasifikasi otomatis yang dapat diadopsi untuk analisis sentimen produk herbal di media sosial.

Untuk pengembangan penelitian selanjutnya, disarankan memperluas sumber data dengan memasukkan konten dari platform media sosial seperti Twitter, Instagram, atau forum kesehatan dan kecantikan, agar analisis tidak hanya terbatas pada literatur akademik namun juga mencakup persepsi publik. Selain itu, penggunaan metode klasifikasi lain seperti *Support Vector Machine* atau Random Forest dapat dibandingkan dengan Naïve Bayes untuk memperoleh perspektif performa yang lebih beragam.

UCAPAN TERIMAKASIH

Terima Kasih kepada Universitas Wijaya Kusuma Surabaya dan Pondok Pesantren Alif Laam Mim sehingga penelitian ini dapat terlaksana dengan baik.

DAFTAR PUSTAKA

- [1] E. I. Zulkurnain, S. Ramli, A. A. Ali, R. J. James, I. S. Kamarazaman, And H. Halim, "The Phytochemical And Pharmacological Effects Of Hibiscus Rosa-Sinensis: A Review," *Int J Pharm Investig*, Vol. 13, No. 3, Pp. 422–431, Jul. 2023, Doi: 10.5530/Ijpi.13.3.053.
- [2] Y. Syaula, A. L. Antari, And D. A. Purbaningrum, "Pengaruh Perendaman Ekstrak Bunga Sepatu (Hibiscus Rosa Sinensis L.) Terhadap Pertumbuhan Candida Albicans Pada Plat Resin Akrilik," *E-Gigi*, Vol. 9, No. 2, P. 159, Jun. 2021, Doi: 10.35790/Eg.9.2.2021.34104.
- [3] A. Razi *Et Al.*, "Faktor Yang Memengaruhi Penggunaan Pengobatan Herbal Pada Pasien Di Puskesmas Pekan Labuhan Tahun 2023," *Scripta Score Scientific Medical Journal*, Vol. 5, No. 1, Pp. 1–8, Sep. 2023, Doi: 10.32734/Scripta.V5i1.13656.

- [4] Q. A. Alya, "Ketokonazol 2% Lebih Efektif Dibanding Ekstrak Bunga Sepatu (*Hibiscus Rosa Sinensis L.*) Sebagai Herbal Potensial Anti Mikosis," *Jurnal Kedokteran Raflesia*, Vol. 6, No. 2, Pp. 10–18, Dec. 2020, Doi: 10.33369/Juke.V6i2.13829.
- [5] A. Soyata, S. Hodijah, And M. Suhendri, "Formulasi Sediaan Lip Cream Dari Ekstrak Bunga Kembang Sepatu (*Hibiscus Rosa Sinensis L.*) Sebagai Zat Warna Alami," *Majalah Farmasetika*, Vol. 9, No. 3, P. 216, Apr. 2024, Doi: 10.24198/Mfarmasetika.V9i3.51674.
- [6] R. M. Syahadat, I. Saleh, And A. A. F. A. Christalista, "Tren Riset Pascapanen Edible Flower," *Sentri: Jurnal Riset Ilmiah*, Vol. 1, No. 2, Pp. 488–497, Oct. 2022, Doi: 10.55681/Sentri.V1i2.241.
- [7] S. Andriyani, L. Kusmita, And Y. Franyoto, "Aktivitas Antibakteri *Staphylococcus Aureus* Dan Penetapan Kadar Antosianin Total Dari Bunga Kembang Sepatu (*Hibiscus Rosa-Sinensis L.*)," *Media Farmasi Indonesia*, Vol. 18, No. 1, Pp. 17–26, Mar. 2023, Doi: 10.53359/Mfi.V18i1.214.
- [8] R. A. Ratte, Nur Herlina Nasir, And Silviana Hasanuddin, "Uji Aktivitas Antijamur Ekstrak Etanol Bunga Kembang Sepatu (*Hibiscus Rosa-Sinensis L.*) Terhadap *Candida Albicans* Dan *Malassezia Furfur*," *Jurnal Pharmacia Mandala Waluya*, Vol. 4, No. 3, Pp. 156–167, Jun. 2025, Doi: 10.54883/Jpmw.V4i3.141.
- [9] L. Mutoharoh, S. D. Santoso, And A. A. Mandasari, "Pemanfaatan Ekstrak Bunga Sepatu (*Hibiscus Rosa-Sinensis L.*) Sebagai Alternatif Pewarna Alami Sediaan Sitologi Pengganti Eosin Pada Pengecatan Diff Quik," *Jurnal Sainhealth*, Vol. 4, No. 2, P. 21, Nov. 2020, Doi: 10.51804/Jsh.V4i2.770.21-26.
- [10] I. Trisnajayanti, "Pengaruh Basis Salep Hidrokarbon Dan Basis Salep Serap Terhadap Sediaan Salep Ekstrak Daun Bunga Sepatu (*Hibiscus Rosa-Sinensis L.*)," *Parapemikir: Jurnal Ilmiah Farmasi*, Vol. 2, No. 3, Oct. 2014, Doi: 10.30591/Pjif.V2i3.53.
- [11] R. A. D. Nopita, E. A. Ihsan, And P. Hariadi, "Formulasi Dan Evaluasi Krim Pelembab Kulit Ekstrak Mahkota Bunga Sepatu (*Hibiscus Rosa-Sinensis L.*)," *Sinteca*, Vol. 2, No. 1, Pp. 29–39, Feb. 2022, Doi: 10.29408/Sinteca.V2i1.4391.
- [12] D. Julia, S. Salni, And S. Nita, "Pengaruh Ekstrak Bunga Kembang Sepatu (*Hibiscus Rosa-Sinensis Linn.*) Terhadap Jumlah, Motilitas, Morfologi, Viabilitas Spermatozoa Tikus Jantan," *Biomedical Journal Of Indonesia*, Vol. 3, No. 3, Pp. 104–112, Jun. 2019, Doi: 10.32539/Bji.V3i3.8600.
- [13] R. Qurrota'ain, S. M. Wicahyo, And A. Fitriawati, "Formulasi Hair Tonic Kombinasi Ekstrak Bunga Sepatu (*Hibiscus Rossa Sinensis L.*) Dan Daun Seledri (*Apium Graveolens L.*) Sebagai Perangsang Pertumbuhan Rambut Pada Kelinci (New Zealand White)," *Jurnal Buana Farma*, Vol. 5, No. 3, Pp. 560–572, Sep. 2025, Doi: 10.36805/Jbf.V5i3.1515.
- [14] L. B. Ilmawan And M. A. Mude, "Perbandingan Metode Klasifikasi *Support Vector Machine* Dan *Naïve Bayes* Untuk Analisis Sentimen Pada Ulasan Tekstual Di Google Play Store," *Ilkom Jurnal Ilmiah*, Vol. 12, No. 2, Pp. 154–161, Aug. 2020, Doi: 10.33096/Ilkom.V12i2.597.154-161.
- [15] M. Haris Humaidi, Sutrisno, And P. Widyo Laksono, "Implementation Of Machine Learning For Text Classification Using The *Naïve Bayes* Algorithm In Academic Information Systems At Sebelas Maret University Indonesia," *E3s Web Of Conferences*, Vol. 465, P. 02048, Dec. 2023, Doi: 10.1051/E3sconf/202346502048.
- [16] M. F. F. Mardianto, B. S. Pratama, M. Audilla, And E. Pusporani, "Sentiment Analysis Of Suicide On X Using *Support Vector Machine* And *Naïve Bayes* Classifier Algorithms," *Intensif: Jurnal Ilmiah Penelitian Dan Penerapan Teknologi Sistem Informasi*, Vol. 9, No. 1, Pp. 60–75, Feb. 2025, Doi: 10.29407/Intensif.V9i1.23742.
- [17] F. Al Farikhi And R. W. D. Pramono, "Perbandingan Algoritma Classification And Regression Tree (Cart) Dan Random Forest (Rf) Untuk Klasifikasi Penggunaan Lahan Pada Google Earth Engine," *Jurnal Spatial Wahana Komunikasi Dan Informasi Geografi*, Vol. 23, No. 2, Pp. 170–179, Sep. 2023, Doi: 10.21009/Spatial.232.09.
- [18] J. Asher And E. P. Rachmawati, "Analisis Sentimen Ulasan Bintang Lima Aplikasi Instagram Di Google Play Store Menggunakan Indobert," *Dinamik*, Vol. 30, No. 2, Pp. 318–330, Jul. 2025, Doi: 10.35315/Dinamik.V30i2.10192.
- [19] D. Prastyo, D. Irawan, And I. H. Mursyidin, "Klasifikasi Sentimen Komentar Youtube Dengan Nlp Pada Debat Pilkada Banten 2024," *Bit-Tech*, Vol. 7, No. 2, Pp. 413–421, Dec. 2024, Doi: 10.32877/Bt.V7i2.1833.
- [20] M. E. Siregar, S. Dermawan, And A. A. Hisyam, "Perbandingan Kinerja *Naïve Bayes* Dan *Knn* Dalam Analisis Sentimen Komentar X Dengan Dan Tanpa Text Preprocessing (Studi Kasus: Danantara)," *Jurnal Informatika Dan Teknik Elektro Terapan*, Vol. 13, No. 3, Jul. 2025, Doi: 10.23960/Jitet.V13i3.6612.
- [21] Ire Puspa Wardhani, Yudi Irawan Chandra, And Ferri Yusra, "Application Of The *Naïve Bayes* Classifier Algorithm To Analyze Sentiment For The Covid-19 Vaccine On Twitter In Jakarta," *International Journal Of Innovation In Enterprise System*, Vol. 7, No. 1, Pp. 1–18, Oct. 2024, Doi: 10.25124/Ijies.V7i01.171.
- [22] Ire Puspa Wardhani, Yudi Irawan Chandra, And Ferri Yusra, "Application Of The *Naïve Bayes* Classifier Algorithm To Analyze Sentiment For The Covid-19 Vaccine On Twitter In Jakarta," *International Journal Of Innovation In Enterprise System*, Vol. 7, No. 1, Pp. 1–18, Oct. 2024, Doi: 10.25124/Ijies.V7i01.171.
- [23] M. F. F. Mardianto, B. S. Pratama, M. Audilla, And E. Pusporani, "Sentiment Analysis Of Suicide On X Using *Support Vector Machine* And *Naïve Bayes* Classifier Algorithms," *Intensif: Jurnal Ilmiah Penelitian Dan Penerapan Teknologi Sistem Informasi*, Vol. 9, No. 1, Pp. 60–75, Feb. 2025, Doi: 10.29407/Intensif.V9i1.23742.
- [24] Arun Padmanabhan And Dr. Devasenapathy. K, "Detection Of Partisan Bias In Political Social Media Posts Using *Naïve Bayes* Algorithm," *International Journal Of Engineering Technology And Management Sciences*, Vol. 7, No. 4, Pp. 647–652, 2023, Doi: 10.46647/Ijetms.2023.V07i04.090.
- [25] Arun Padmanabhan And Dr. Devasenapathy. K, "Detection Of Partisan Bias In Political Social Media Posts Using *Naïve Bayes* Algorithm," *International Journal Of Engineering Technology And Management Sciences*, Vol. 7, No. 4, Pp. 647–652, 2023, Doi: 10.46647/Ijetms.2023.V07i04.090.
- [26] M. Kaur, R. Sandhu, And R. Mohana, "A Framework For Scheduling Iot Application Jobs On Fog Computing Infrastructure Based On Qos Parameters," *International Journal Of Pervasive Computing And Communications*, Vol. 19, No. 3, Pp. 364–385, May 2023, Doi: 10.1108/Ijpc-08-2020-0108.
- [27] Aldo, D., Nur, Y. S. R., & Fathoni, M. Y. Sistem Pakar Diagnosis Penyakit Pada Ikan Bawal Bintang Dengan Pendekatan *Naïve Bayes*. *Jurnal Informatika: Jurnal Pengembangan IT*, 8(2). 2023.
- [28] C. Bao And Y. Wang, "A Survey Of Word Cloud Visualization," *Journal Of Computer-Aided Design & Computer Graphics*, Vol. 33, No. 4, Pp. 532–544, Apr. 2021, Doi: 10.3724/Sp.J.1089.2021.18811.
- [29] S. Chairunnissa Deliva Akbar, D. Agus Salim, R. Rahmad, A. Khomsi, And R. Sovia, "Perbandingan Kinerja PCA-KNN Dan LDA Dalam Klasifikasi Jenis Kupu-Kupu: Analisis Akurasi Dan Efektivitas," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 3, pp. 5014–5022, May 2025, doi: 10.36040/jati.v9i3.13863.