

Analisis Sentiment Pengguna X Terhadap Hilirisasi Kemenyan Menggunakan Algoritma Naïve Bayes

Farrel Rizki Utama*, Suaidah

Fakultas Teknik dan Ilmu Komputer, Program Studi Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ¹*farrel_rizki_utama@teknokrat.ac.id, ²suaidah@teknokrat.ac.id

Email Penulis Korespondensi: farrel_rizki_utama@teknokrat.ac.id

Submitted: 03/11/2025; Accepted: 09/12/2025; Published: 11/12/2025

Abstrak—Isu hilirisasi kemenyan di Indonesia menjadi perhatian penting karena berpotensi meningkatkan nilai ekonomi komoditas lokal serta kesejahteraan masyarakat, khususnya petani. Namun, persepsi publik terhadap kebijakan hilirisasi ini masih beragam dan belum banyak dianalisis secara ilmiah, terutama di media sosial. Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen pengguna media sosial X terhadap isu hilirisasi kemenyan dengan menggunakan algoritma Naïve Bayes. Data penelitian diperoleh melalui proses crawling menggunakan API Twitter dengan kata kunci “Hilirisasi Kemenyan” dan “Hilirisasi”, menghasilkan sebanyak 1.844 tweet. Data tersebut kemudian melalui tahap preprocessing yang meliputi cleaning, case folding, normalisasi, tokenizing, stopword removal, dan stemming, sehingga tersisa 1.790 tweet yang siap dianalisis. Proses pelabelan sentimen dilakukan menggunakan pendekatan lexicon-based dengan tiga kategori, yaitu positif, negatif, dan netral. Representasi fitur dilakukan dengan metode TF-IDF, kemudian data diklasifikasikan menggunakan algoritma Naïve Bayes. Hasil pengujian menunjukkan bahwa algoritma Naïve Bayes mampu mengklasifikasikan sentimen dengan baik, dengan presisi tertinggi pada kelas negatif sebesar 0,90 dan recall tertinggi pada kelas netral sebesar 0,92. Mayoritas pengguna X menunjukkan sentimen netral terhadap isu hilirisasi kemenyan sebesar 55,20%, diikuti positif 26,03% dan negatif 18,77%.

Kata Kunci: Analisis Sentiment; Hilirisasi Kemenyan; Media Sosial X; Naïve Bayes; TF-IDF.

Abstract—The issue of frankincense downstreaming in Indonesia is a significant concern because it has the potential to increase the economic value of local commodities and the welfare of the community, especially farmers. However, public perception of this downstreaming policy is still diverse and has not been widely analyzed scientifically, especially on social media. Therefore, this study aims to analyze the sentiment of social media users X towards the issue of frankincense downstreaming using the Naïve Bayes algorithm. The research data was obtained through a crawling process using the Twitter API with the keywords "Frankincense Downstreaming" and "Downstreaming", resulting in 1,844 tweets. The data then went through a preprocessing stage including cleaning, case folding, normalization, tokenizing, stopword removal, and stemming, leaving 1,790 tweets ready for analysis. The sentiment labeling process was carried out using a lexicon-based approach with three categories: positive, negative, and neutral. Feature representation was carried out using the TF-IDF method, then the data was classified using the Naïve Bayes algorithm. The test results show that the Naïve Bayes algorithm is able to classify sentiment well, with the highest precision in the negative class at 0.90 and the highest recall in the neutral class at 0.92. The majority of X users showed neutral sentiment towards the issue of frankincense downstreaming at 55.20%, followed by positive at 26.03% and negative at 18.77%.

Keywords: Sentiment Analysis; Frankincense Downstreaming; X Social Media; Naïve Bayes; TF-IDF.

1. PENDAHULUAN

Media sosial telah menjadi sangat penting bagi masyarakat kontemporer untuk berkomunikasi dan berbagi informasi. Salah satu media sosial yang populer di dunia termasuk di Indonesia adalah X. X juga dikenal sebagai tempat untuk berbagi pesan singkat, X tersebut memiliki berbagai jenis informasi, seperti opini dan penilaian mengenai produk serta layanan [1]. Banyaknya tweet di X membuat sulit untuk mengetahui arti atau maksud dari setiap tweet, apakah bernuansa negatif, positif, atau netral [2]. Sehingga, diperlukan penggunaan metode analisis sentiment untuk mengetahui makna dari setiap tweet, apakah bersifat positif, negatif, atau netral. *Analisis sentiment* adalah Metode pemrosesan data yang tidak memiliki struktur yang jelas diubah menjadi bentuk data yang terstruktur agar dapat dianalisis untuk mengetahui *sentiment*. *Sentiment* umumnya berkaitan dengan perasaan, pendapat, atau komentar yang disampaikan dalam bentuk teks [3]. Hilirisasi adalah proses pengolahan bahan mentah menjadi produk yang memiliki nilai tambah lebih tinggi, sering kali melibatkan transformasi dari bahan mentah menjadi bahan setengah jadi atau barang jadi. Bertujuan untuk meningkatkan nilai ekonomi sumber daya alam suatu negara dengan mengolahnya menjadi produk yang lebih bernilai. Hilirisasi adalah proses fundamental dari pengolahan bahan baku menjadi barang setengah jadi atau jadi, penting untuk menambah nilai ekonomi dan kesejahteraan bangsa [4]. Kemenyan merupakan getah atau resin aromatic yang dihasilkan dari pohon (*Genius Styrax*), terutama *Styrax benzoin*. Getah ini berasal dari menyadap batang pohon, lalu setelah mengeras akan dipanen. Pohon kemenyan banyak tumbuh di daerah hutan tropis dan subtropis terutama di Sumatera. Resin atau getah digunakan sebagai bahan untuk parfum, aromaterapi, pengawet, dan campuran kosmetik [5] [6]. Algoritma *Naive Bayes* merupakan salah satu metode yang digunakan dalam *text mining* atau *data mining* untuk mengelompokkan sejumlah data berdasarkan kategori tertentu [7]. *Naive Bayes* memandang setiap data sebagai kumpulan kata-kata, kemudian menghitung probabilitas dokumen tersebut termasuk ke dalam kelas tertentu berdasarkan seberapa sering kata-kata tertentu muncul dalam dokumen tersebut [8].

Berikut ini beberapa penelitian telah dilakukan untuk membandingkan kemampuan algoritma klasifikasi *Naive Bayes* dalam berbagai kasus. Aprilia dkk. (2021) menggunakan algoritma *Naive Bayes* dalam klasifikasi *event flash sale* di twitter dengan akurasi yang optimal 83.53% dan 81.48% pada kasus *flash sale* shopee [9]. Penelitian yang

dilakukan oleh Moch Bima Prakoso dkk (2021) menggunakan metode Naive Bayes dalam menganalisis *sentiment* masyarakat mengenai sistem pembelajaran online selama pandemi Covid-19, mereka melakukan pengujian dengan 5 *fold* dengan akurasi 0,55, presisi 0,54, dan recall 0,54 dan 10 *fold* dengan akurasi 0,59, presisi 0,61, recall 0,60 [10].

Nilai Hardi dkk menggunakan algoritma *Naive Bayes* pada klasifikasi data untuk analisis *sentiment physical distancing* di twitter dengan akurasi 50,26% [11]. Dalam analisis *sentiment* di twitter terhadap penggunaan mobil listrik di Indonesia pada tahun 2024, Widia Ningsih dkk (2024) mengemukakan bahwa *Naive Bayes* menampilkan hasil yang cukup baik dengan akurasi 63.02%, presisi 71.05%, recall 70.19% [12]. Aditya Hastami Ruger dkk meneliti *sentiment* terhadap pelanggan shopee dengan media sosial twitter menggunakan algoritma *Naive Bayes* dan hasilnya membuktikan bahwa algoritma Naive Bayes mencapai tingkat akurasi yang tinggi dengan akurasi mencapai 97% [13].

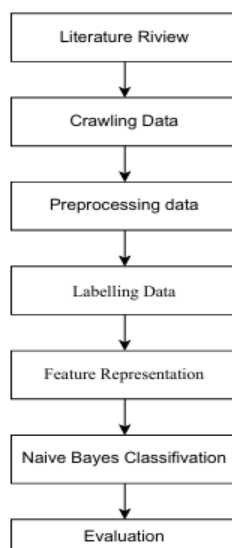
Penelitian ini berawal dari pidato Bapak Wakil Presiden Indonesia yaitu Bapak Gibran Rakabuming Raka yang menyoroti pentingnya hilirisasi kemenyan sebagai salah satu komoditas unggulan nasional yang berpotensi meningkatkan nilai tambah sektor agraris. Banyak penelitian sebelumnya telah membantu dalam memperkembangkan cara menganalisis *sentiment*. Khususnya dengan menggunakan algoritma *Naive Bayes*. Namun, penelitian yang secara khusus mengeksplorasi persepsi publik terhadap masalah hilirisasi kemenyan, khususnya di platform media sosial X, dan di bidang agraris masih terbatas sehingga peneliti membuka peluang untuk penelitian lanjutan. Analisis *sentiment* terhadap isu hilirisasi kemenyan menghadapi tantangan khusus akibat sifat data media sosial yang selalu berubah, tidak memiliki struktur yang jelas, serta menggunakan berbagai variasi bahasa yang rumit.

Penelitian ini membahas tentang analisis *sentiment* untuk mengklasifikasikan ulasan dari X mengenai hilirisasi kemenyan. Data dari X tersebut dapat diproses dengan menggunakan teknik *text mining*, lalu dilanjutkan dengan mengelompokkan setiap Tweet ke dalam tiga kategori, yakni positif, negatif, atau netral dan mengidentifikasi akurasi dengan algoritma *Naive Bayes* [8]. Tujuan dalam penelitian ini yaitu memberikan gambaran mengenai pola persepsi masyarakat yang menampilkan bahwa kemenyan tidak hanya memiliki banyak manfaat, baik dari segi kesehatan, budaya, maupun industri, tetapi juga pengembangan hilirisasi dan meningkatkan kesejahteraan masyarakat. Fokus utama dalam penelitian ini yaitu mengklasifikasikan *sentiment* masyarakat (positif, negatif, netral) dan mengidentifikasi algoritma *Naive Bayes* dengan akurasi terhadap isu hilirisasi kemenyan dengan menggunakan algoritma *Naive Bayes*. Dengan adanya penelitian ini masyarakat harus lebih bijak dalam menyampaikan opini di media sosial dan sebagai bahan evaluasi kebijakan untuk pemerintah.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian dilakukan secara terstruktur untuk menganalisis *sentiment* masyarakat terhadap isu hilirisasi kemenyan yang dibicarakan di media sosial X. Sebelum melakukan *crawling data* peneliti membaca *Literature review*, *Literatur review* adalah upaya untuk mengecek, menganalisis, dan menilai berbagai karya ilmiah yang sudah ada dan berkaitan dengan topik yang diteliti. Kemudian data dikumpulkan dengan teknik *crawling data* berdasarkan kata kunci yang sesuai dan relevan, yang berhasil menghasilkan 1.800 tweet. Data kemudian melewati beberapa tahapan *preprocessing data* yaitu *cleaning*, *case folding*, *tokenizing*, *stopword removal*, *normalisasi*, dan *stemming* sebelum dilakukan proses *labelling data*. Data tersebut dikategorikan menjadi tiga kelompok, yaitu positif, negatif, dan netral. Data yang sudah diberi label akan dilakukan representasi data menggunakan *TF-IDF* dan dibagi menjadi dua data yaitu *data training* dan *data testing*. Algoritma klasifikasi yang akan diterapkan yaitu *Naive Bayes* dan evaluasi menggunakan akurasi, presisi, recall dan *F1-score*. gambaran mengenai metodologi penelitian yang telah dibuat ditampilkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Penjelasan Gambar 1

- Literature Riview*, yaitu proses penelitian yang mencakup peninjauan serta evaluasi secara mendalam terhadap berbagai sumber referensi.
- Crawling Data*, yaitu pengumpulan data dari twitter menggunakan *Tweet-Harvest* yang berfokus pada Hilirisasi Kemenyan tahun 2025.
- Preprocessing data*, yaitu proses membersihkan data mentah menjadi data yang mudah dipahami. *Preprocessing* meliputi: *cleaning, casefolding, normalization, tokenizing, stopword removal, stemming*.
- Labelling Data*, yaitu proses pemberian label pada data.

2.2 Pengumpulan Data

Penelitian ini mengumpulkan data berupa opini masyarakat yang diunggah di media sosial X, dengan memanfaatkan teknik *crawling tweet harvest* dengan menggunakan *API (Application Programming Interface)*. Proses *crawling* dilakukan menggunakan Python dengan bantuan tools *Google Colaboratory*. [14]

2.3. Preprocessing Data

Proses *preprocessing data* bertujuan mengubah data teks yang tidak terstruktur menjadi bentuk data yang terstruktur [15]. Tahap ini dilakukan agar analisis *sentiment* dapat berjalan dengan lebih efektif menggunakan data yang bersih dan terstruktur. Berikut adalah tahap preprocessing data ditampilkan pada Gambar 2.



Gambar 2. Tahapan *Preprocessing*

Pada Gambar 2 menampilkan tahapan preprocessing data beberapa tahapan yang terdiri dari *cleaning, case folding, normalization, tokenizing, stopword removal, dan stemming*. Penjelasan untuk setiap tahapan adalah sebagai berikut.

2.3.1 Cleaning Data

Cleaning Data adalah tahap membersihkan karakter selain huruf, menghapus penunjuk (@), hastag (#), serta menghilangkan tautan dan emoji dari setiap respons [16].

2.3.2 Case Folding

Case Folding adalah tahapan mengubah semua huruf dalam dokumen menjadi huruf kecil [17]. Tujuannya adalah untuk menyamakan bentuk huruf dalam teks agar komputer dapat memperlakukan kata yang sama sebagai bentuk yang sama, tanpa memperhatikan perbedaan besar kecil hurufnya.

2.3.3 Normalization

Normalization adalah untuk menormalkan kata yang disingkat ataupun diperpanjang [18]. Tahapan ini merupakan proses untuk menyamakan bentuk kata dalam teks agar sesuai dengan aturan bahasa yang baku.

2.3.4 Tokenizing

Tokenizing adalah tahap membagi kata dalam suatu kalimat [19] Tahapan ini bertujuan menyederhanakan proses analisis teks dengan mengubah kalimat atau paragraf menjadi sekumpulan kata yang terpisah.

2.3.5 Stopword Removal

Stopword Removal adalah tahap dalam menghilangkan kata-kata yang terdapat dalam daftar kata stopwords Bahasa Indonesia. Tujuan dari stopwords removal yaitu untuk mengurangi penggunaan kata-kata umum yang tidak terlalu penting dalam kamus kata stopwords [20].

2.3.6 Stemming

Stemming adalah mengubah kata menjadi bentuk dasarnya agar memudahkan analisis data [21]. Tujuan dari tahap ini adalah mengurangi kerumitan teks serta memperkuat data dalam proses analisis.

2.4 Labelling Data

Labelling Data adalah cara memberi tanda atau kategori pada data di dalam dataset agar mudah digunakan dalam melatih model *machine learning*. Tanda atau kategori ini menjelaskan bagaimana data tersebut diklasifikasikan. Bertujuan memberikan label seperti positif, negatif, atau netral untuk menampilkan perasaan yang diungkapkan dalam teks tersebut[22].

2.5 Pembobotan TF-IDF

TF-IDF (*Term Frequency Inverse Document Frequency*) adalah metode yang digunakan untuk menentukan seberapa relevan sebuah kata dalam sebuah dokumen dengan bobot tertentu pada setiap kata tersebut. [23] Berikut ini adalah

rumus yang digunakan untuk menghitung nilai TF-IDF (*Term Frequency-Inverse Document Frequency*). Berikut ini adalah rumus TF-IDF ditunjukkan persamaan 1.

$$TF(t, d) = \frac{f_{t,d}}{\sum_k f_{k,d}} \times IDF(t) = \log \left(\frac{N}{df_t} \right) \tag{1}$$

2.6 Klasifikasi Naive Bayes

Algoritma *Naive Bayes* adalah salah satu metode yang digunakan dalam *text mining* atau *data mining* untuk mengelompokkan sejumlah data berdasarkan kategori tertentu. Algoritma ini sangat sering digunakan dalam analisis *sentiment* karena terkenal memiliki tingkat akurasi yang cukup baik, terutama dalam mengklasifikasikan teks. Berikut ini adalah rumus perhitungan *Naive Bayes* ditampilkan pada persamaan 2.

$$P(A | B) = \frac{P(B|A) \times P(A)}{P(B)} \tag{2}$$

$P(A | B)$ adalah peluang sebuah data termasuk kedalam kelas tertentu. $P(B | A)$ adalah peluang kata-kata dalam teks muncul jika memang termasuk kelas itu. $P(A)$ adalah seberapa sering kelas itu muncul di seluruh data. $P(B)$ adalah dimana peluang data itu muncul secara umum.

2.7 Evaluasi Model

Evaluasi model adalah nilai yang menampilkan perbandingan data tweet yang berhasil terdeteksi selama pengujian. Nilai akurasi dapat menggambarkan seberapa dekat hasil prediksi sistem dengan hasil prediksi manusia. Dalam proses klasifikasi, evaluasi dilakukan dengan menggunakan beberapa metrik seperti akurasi, presisi, recall, dan *F1-score*. Berikut ini adalah rumus perhitungan *Naive Bayes* ditampilkan pada persamaan 3, 4, 5, 6.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

$$\text{Presisi} = \frac{TP}{TP+FP} \tag{4}$$

$$\text{Recal} = \frac{TP}{TP+FN} \tag{5}$$

$$F1\text{-score} = 2 \times \frac{\text{Presisi} \times \text{recall}}{\text{Presisi} + \text{recall}} \tag{6}$$

True Positive (TP) merupakan jumlah data yang benar-benar termasuk dalam kelas positif dan berhasil diprediksi positif oleh model. *False Positive* (FP) menampilkan kesalahan model saat memprediksi data sebagai positif, padahal sebenarnya data itu berasal dari kelas negatif. *True Negative* (TN) adalah jumlah data yang memang termasuk kelas negatif dan berhasil dikenali dengan benar oleh model. Sementara itu, *False Negative* (FN) menggambarkan kesalahan ketika model memprediksi data sebagai negatif, padahal data tersebut sebenarnya termasuk ke dalam kelas positif.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Penelitian dimulai dengan mengumpulkan data yang didapat dari media sosial X dengan menggunakan kata kunci “Hilirisasi Kemenyan”, “Hilirisasi” dan komentar sebuah postingan yang berhubungan dengan hilirisasi kemenyan untuk mendapat *tweet* yang sesuai. *Tweet* sebanyak 1.844 berhasil dikumpulkan dengan bahasa pemrograman python, dengan memanfaatkan *library tweet harvest* untuk mengekstrak data dan *library pandas* untuk mengolah *dataset* tersebut. Berikut adalah *dataset* yang digunakan sebagai contoh penelitian yang dapat dilihat pada Tabel 1.

Tabel 1. Hasil *Crawling* Data

No	Tweet
1	Dengan mengolahnya jadi bahan setengah jadi saja yakni minyak benzoin industri dalam negeri bisa menikmati nilai tambah lebih dari 1.000 persen dari kemenyan. #Ekonomi #AdadiKompas https://t.co/L9HZkbTGvs
1844	@is_pelssy @taufik_q Spt melindungi rekan2 yg punya usaha besar... Coba lihat Sektor SDA sbg akibat Hilirisasi.. Yg spt ga sejalan spt disuntik mati... Lihat cukai rokok.. produsen rkk ilegal makin merajalela.. yg ditangkapin distributornya....

Dilihat dari Tabel 1 *tweet* tersebut diperoleh dari *tweet-tweet* yang dikumpulkan, tidak hanya kata kunci seperti “Hilirisasi Kemenyan” atau “Hilirisasi” tetapi juga menyampaikan berbagai opini masyarakat. Perbedaan dalam bahasa dan struktur kalimat dari *tweet* tersebut menjadi dasar utama dalam proses analisis *sentiment*, yang nantinya akan dilakukan melalui beberapa tahapan yaitu *preprocessing data*, *labelling data* dan klasifikasi *Naive Bayes*.



3.2 Preprocessing Data

Setelah data berhasil dikumpulkan, langkah selanjutnya adalah tahap *preprocessing* atau pemrosesan awal data, yang berperan penting dalam menyiapkan data teks agar siap digunakan untuk analisis *sentiment*, khususnya pada topik terkait Hilirisasi Kemenyan. Pada tahap ini dilakukan beberapa proses utama seperti *cleaning*, *case folding*, *normalisasi kata*, *tokenizing*, *stopword removal*, dan *stemming*. Setiap tahapan memiliki perannya masing-masing, mulai dari membersihkan data dari karakter yang tidak penting, menyeragamkan huruf, memperbaiki kata tidak baku, hingga mengubah kata menjadi bentuk dasarnya. Proses *preprocessing* juga bertujuan untuk mengurangi *noise* atau gangguan dalam data, memperbaiki nilai yang tidak konsisten, serta memastikan bahwa data yang digunakan benar-benar relevan dengan konteks penelitian. Dengan melakukan tahap ini, kualitas data dapat ditingkatkan sehingga model analisis *sentiment* dapat bekerja lebih optimal dan menghasilkan hasil klasifikasi yang lebih akurat. Setelah melalui proses pembersihan dan penyaringan ini, jumlah data yang siap digunakan dalam analisis *sentiment* berkurang menjadi 1790 tweet.

3.2.1 Cleaning Data

Pada tahap ini peneliti melakukan pembersihan data untuk membersihkan data dari bagian-bagian yang tidak memberikan dampak penting terhadap analisis *sentiment*. Hasil dari tahap *Cleaning Data* ditampilkan pada Tabel 2.

Tabel 2. Hasil *Cleaning Data*

Tweet	Cleaning
Dengan mengolahnya jadi bahan setengah jadi saja yakni minyak benzoin industri dalam negeri bisa menikmati nilai tambah lebih dari 1.000 persen dari kemenyan. #Ekonomi #AdadiKompas https://t.co/L9HZkbTGvs @is_pelssy @taufik_q Spt melindungi rekan2 yg punya usaha besar... Coba lihat Sektor SDA sbg akibat Hilirisasi.. Yg spt ga sejalan spt disuntik mati... Lihat cukai rokok.. produsen rkk ilegal makin merajalela.. yg ditangkapin distributornya....	Dengan mengolahnya jadi bahan setengah jadi saja yakni minyak benzoin industri dalam negeri bisa menikmati nilai tambah lebih dari persen dari kemenyan Spt melindungi rekan yg punya usaha besar Coba lihat Sektor SDA sbg akibat Hilirisasi Yg spt ga sejalan spt disuntik mati Lihat cukai rokok produsen rkk ilegal makin merajalela yg ditangkapin distributornya

Tabel 2 menampilkan hasil proses *Cleaning Data*. Pada tweet pertama, elemen seperti mention, tautan link, dan tanda baca dihilangkan, tetapi makna utama dalam tweet tidak dihilangkan. Sama halnya dengan tweet kedua, mention, tautan link dan tanda baca dihilangkan tetapi informasi penting mengenai Hilirisasi Kemenyan tidak dihilangkan.

3.2.2 Case Folding

Case folding adalah tahap mengubah semua huruf yang berada di dalam teks menjadi huruf kecil agar menghilangkan perbedaan penulisan karena penggunaan huruf kapital. Tujuan tahap ini adalah untuk menyederhanakan analisis teks serta meningkatkan akurasi dalam menganalisis *sentiment* dengan menggunakan kata-kata yang sama meskipun ditulis dalam format huruf yang berbeda. Hasil dari tahap *Case Folding* ditampilkan pada Tabel 3.

Tabel 2. Hasil *Cleaning Data*

Cleaning	Case Folding
Dengan mengolahnya jadi bahan setengah jadi saja yakni minyak benzoin industri dalam negeri bisa menikmati nilai tambah lebih dari persen dari kemenyan Spt melindungi rekan yg punya usaha besar Coba lihat Sektor SDA sbg akibat Hilirisasi Yg spt ga sejalan spt disuntik mati Lihat cukai rokok produsen rkk ilegal makin merajalela yg ditangkapin distributornya	dengan mengolahnya jadi bahan setengah jadi saja yakni minyak benzoin industri dalam negeri bisa menikmati nilai tambah lebih dari persen dari kemenyan spt melindungi rekan yg punya usaha besar coba lihat sektor sda sbg akibat hilirisasi yg spt ga sejalan spt disuntik mati lihat cukai rokok produsen rkk ilegal makin merajalela yg ditangkapin distributornya

Berdasarkan hasil dari Tabel 3 yaitu tahapan *Case Folding*, pada tweet pertama kata “Dengan” yang semula menggunakan huruf kapital menjadi “dengan” yang diubah ke dalam huruf kecil. Sama halnya dengan tweet kedua, tidak ada perubahan selain huruf kapital diubah menjadi huruf kecil di keseluruhan kata.

3.2.3 Normalization

Normalization adalah tahap penting yang digunakan untuk mengubah teks yang tidak terstruktur dan berbahasa *informal* menjadi bentuk yang lebih standar. Tahap ini sangat diperlukan ketika memproses data dari media sosial, karena teks tersebut biasanya mengandung ejaan, singkatan, serta gaya penulisan yang beragam. Hasil dari tahap normalisasi dapat dilihat pada Tabel 4.



Tabel 4. Hasil *Normalization*

<i>Case Folding</i>	<i>Normalization</i>
dengan mengolahnya jadi bahan setengah jadi saja yakni minyak benzoin industri dalam negeri bisa menikmati nilai tambah lebih dari persen dari kemenyan	dengan mengolahnya jadi bahan setengah jadi saja yakni minyak benzoin industri dalam negeri bisa menikmati nilai tambah lebih dari persen dari kemenyan
spt melindungi rekan yg punya usaha besar coba lihat sektor sda sbg akibat hilirisasi yg spt ga sejalan spt disuntik mati lihat cukai rokok produsen rkk ilegal makin merajalela yg ditangkalin distributornya	seperti melindungi rekan yang punya usaha besar coba lihat sektor sumber daya alam sebagai akibat hilirisasi yang seperti tidak sejalan seperti disuntik mati lihat cukai rokok produsen rokok ilegal makin merajalela yang ditangkalin distributornya

Tabel 4 menampilkan hasil dari tahap *Normalization*. Dalam tweet kedua, kata " spt " diubah menjadi "seperti", kata " sda " diubah menjadi " sumber daya alam ", kata "rkk" diubah menjadi " rokok".

3.2.4 Tokenizing

Tokenizing adalah tahap yang bertujuan membagi teks menjadi bagian yang lebih kecil, seperti kata. Langkah ini membantu dalam menganalisis dan memproses data dengan memberikan struktur yang mudah di mengerti oleh sistem, yaitu dengan membagi teks menjadi token. Hasil dari tahap *Tokenizing* ditampilkan dalam Tabel 5.

Tabel 5. Hasil *Tokenizing*

<i>Normalization</i>	<i>Tokenizing</i>
dengan mengolahnya jadi bahan setengah jadi saja yakni minyak benzoin industri dalam negeri bisa menikmati nilai tambah lebih dari persen dari kemenyan	['dengan', 'mengolahnya', 'jadi', 'bahan', 'setengah', 'jadi', 'saja', 'yakni', 'minyak', 'benzoin', 'industri', 'dalam', 'negeri', 'bisa', 'menikmati', 'nilai', 'tambah', 'lebih', 'dari', 'persen', 'dari', 'kemenyan']
seperti melindungi rekan yang punya usaha besar coba lihat sektor sumber daya alam sebagai akibat hilirisasi yang seperti tidak sejalan seperti disuntik mati lihat cukai rokok produsen rokok ilegal makin merajalela yang ditangkalin distributornya	['seperti', 'melindungi', 'rekan', 'yang', 'punya', 'usaha', 'besar', 'coba', 'lihat', 'sektor', 'sumber', 'daya', 'alam', 'sebagai', 'akibat', 'hilirisasi', 'yang', 'seperti', 'tidak sejalan', 'seperti', 'disuntik', 'mati', 'lihat', 'cukai', 'rokok', 'produsen', 'rokok', 'ilegal', 'makin', 'merajalela', 'yang', 'ditangkalin', 'distributornya']

Tabel 4 menampilkan hasil dari tahap *Normalization*. Dalam tweet pertama “dengan mengolahnya jadi bahan setengah jadi saja yakni” dipecah menjadi daftar kata, yaitu ['dengan', 'mengolahnya', 'jadi', 'bahan', 'setengah', 'jadi', 'saja', 'yakni', ...]. Hal yang sama juga terdapat pada tweet kedua, di mana teks dibagi menjadi token-token individual, seperti ['seperti', 'melindungi', 'rekan', 'yang', 'punya', 'usaha', 'besar', 'coba', 'lihat', ...].

3.2.5 Stopword Removal

Stopword removal adalah tahap menghilangkan kata umum yang tidak penting terhadap konteks hilirisasi kemenyan, seperti "dengan", "jadi", atau "bisa". Tujuan dari tahapan ini untuk memfokuskan proses pengolahan hanya pada kata-kata yang memiliki makna dan relevansi dengan hilirisasi kemenyan. Hasil dari tahap *stopword removal* ditampilkan pada Tabel 6.

Tabel 6. Hasil *Stopword removal*

<i>Tokenizing</i>	<i>Stopword removal</i>
['dengan', 'mengolahnya', 'jadi', 'bahan', 'setengah', 'jadi', 'saja', 'yakni', 'minyak', 'benzoin', 'industri', 'dalam', 'negeri', 'bisa', 'menikmati', 'nilai', 'tambah', 'lebih', 'dari', 'persen', 'dari', 'kemenyan']	['mengolahnya', 'bahan', 'minyak', 'benzoin', 'industri', 'negeri', 'menikmati', 'nilai', 'persen', 'kemenyan']
['seperti', 'melindungi', 'rekan', 'yang', 'punya', 'usaha', 'besar', 'coba', 'lihat', 'sektor', 'sumber', 'daya', 'alam', 'sebagai', 'akibat', 'hilirisasi', 'yang', 'seperti', 'tidak sejalan', 'seperti', 'disuntik', 'mati', 'lihat', 'cukai', 'rokok', 'produsen', 'rokok', 'ilegal', 'makin', 'merajalela', 'yang', 'ditangkalin', 'distributornya']	['melindungi', 'rekan', 'usaha', 'coba', 'lihat', 'sektor', 'sumber', 'daya', 'alam', 'akibat', 'hilirisasi', 'tidak sejalan', 'disuntik', 'mati', 'lihat', 'cukai', 'rokok', 'produsen', 'rokok', 'ilegal', 'merajalela', 'ditangkalin', 'distributornya']

Tabel 6 menampilkan hasil *stopword removal*, kata-kata seperti "dengan", "jadi", atau "bisa". dihilangkan, menyisakan kata-kata penting seperti “kemenyan,” dan “hilirisasi.”

3.2.6 Stemming

Stemming adalah tahap mengganti kata-kata yang memiliki makna menjadi bentuk dasar kata. Tahap ini memudahkan model dalam mengenali pola secara lebih cepat, meningkatkan efisiensi dalam proses analisis, serta membantu

menghasilkan evaluasi *sentiment* dengan lebih akurat. Dengan mengubah kata ke bentuk dasarnya, sistem menjadi lebih fokus pada unsur-unsur penting dalam teks. Hasil dari tahap *stemming* ditampilkan pada Tabel 7.

Tabel 7. Hasil *Stemming*

<i>Stopword removal</i>	<i>Stemming</i>
['mengolahnya', 'bahan', 'minyak', 'benzoin', 'industri', 'negeri', 'menikmati', 'nilai', 'persen', 'kemenyan']	olah bahan minyak benzoin industri negeri nikmat nilai persen kemenyan
['melindungi', 'rekan', 'usaha', 'coba', 'lihat', 'sektor', 'sumber', 'daya', 'alam', 'akibat', 'hilirisasi', 'tidak_sejalan', 'disuntik', 'mati', 'lihat', 'cukai', 'rokok', 'produsen', 'rokok', 'ilegal', 'merajalela', 'ditangkapin', 'distributornya']	indung rekan usaha coba lihat sektor sumber daya alam akibat hilir tidak jalan suntik mati lihat cukai rokok produsen rokok ilegal rajalela ditangkapin distributor

Tabel 7 menampilkan hasil stemming. Pada hasil stemming diatas , kata seperti “mengolahnya,” dan “menikmati,” diubah menjadi bentuk dasar “olah,” dan “nikmat,” untuk menyederhanakan teks dan memperjelas makna.

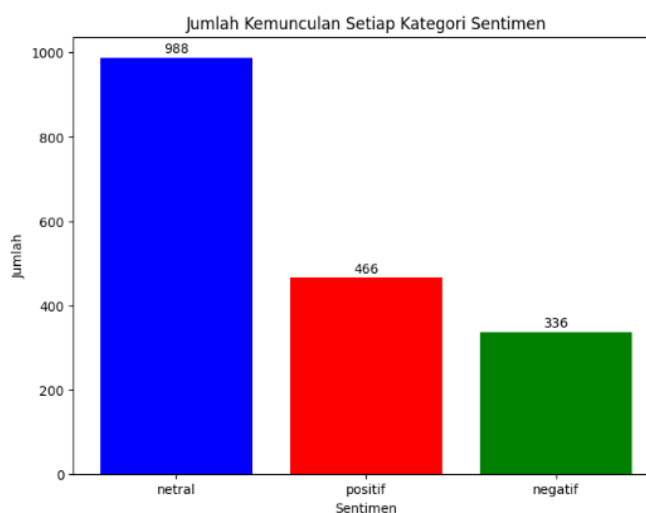
3.3 Labelling Data

Labelling data menggunakan metode kamus *lexicon based*, di mana sistem menganalisis setiap kata dalam komentar untuk menentukan *sentiment* yang sesuai berdasarkan kamus. Kamus ini dibagi menjadi tiga kategori yaitu positif, negatif, dan netral. Hasil *labelling data* tersebut ditampilkan pada Tabel 8.

Tabel 8. *Labelling Data*

<i>Stemming</i>	Score	Sentiment
olah bahan minyak benzoin industri negeri nikmat nilai persen kemenyan	1	positif
indung rekan usaha coba lihat sektor sumber daya alam akibat hilir tidak jalan suntik mati lihat cukai rokok produsen rokok ilegal rajalela ditangkapin distributor	-1	negatif
hilir kemenyan impor kemenyan	0	netral

Hasil *labelling data* menunjukkan bahwa terdapat 466 data dengan *sentiment* positif, 336 data dengan *sentiment* negatif, dan 988 data dengan *sentiment* netral. *Labelling data* ini dilakukan berdasarkan nilai, di mana nilai di atas 0 di kategorikan sebagai positif, nilai di bawah 0 sebagai negatif, dan nilai 0 dianggap netral. *Visualisasi distribusi sentiment* ditampilkan pada Gambar 3.



Gambar 3. *Visualisasi Hasil Labelling Data*

Berdasarkan hasil *labelling data* yang terlihat pada Gambar 3, mayoritas pengguna tweet yang menunjukkan *sentiment* netral, yaitu sebesar 55.20% dari total data. *Sentiment* negatif dengan *persentase* 18.77%, sedangkan *sentiment* positif 26.03%. Angka ini menunjukkan bahwa respon pengguna X terhadap hilirisasi kemenyan cenderung netral.

3.4 Visualisasi WordCloud

Visualisasi WordCloud digunakan untuk menampilkan kata yang lebih sering muncul dalam komentar X terkait isu hilirisasi kemenyan. Pada tahap pertama, sebelum dilakukan preprocessing, *WordCloud* masih terisi oleh kata umum yang tidak memiliki makna penting. Setelah melalui proses *stopword removal*, *stemming*, serta *normalization*, hasil *visualisasi* menjadi lebih terpusat pada kata-kata yang mewakili opini publik secara lebih tepat. Gambar 4 menampilkan hasil akhir pada tahapan tersebut.



Gambar 4. Visualisasi Wordcloud

Dari Gambar 4, kata-kata yang sering muncul adalah "hilir" dan "hilir kemenyan". Kata-kata tersebut merupakan fokus utama dalam membahas isu hilirisasi kemenyan yang dibicarakan di media sosial, khususnya platform X.

3.5 Feature Representation

Dalam tahap *feature representation*, penelitian ini menggunakan TF-IDF (*Term Frequency Inverse Document Frequency*) untuk mengubah komentar isu hilirisasi kemenyan menjadi bentuk numerik yang dapat dianalisis oleh *machine learning*. TF-IDF mengukur sejauh mana sebuah kata relevan dalam sebuah komentar berdasarkan frekuensi kemunculannya dalam dokumen tersebut maupun di seluruh kumpulan data. Kata-kata umum dan tidak memberikan informasi penting akan diberi bobot yang lebih rendah, sedangkan kata-kata yang jarang muncul namun memiliki makna khusus, seperti istilah terkait hilirisasi kemenyan, akan diberi bobot yang lebih tinggi. Tabel 9 menampilkan hasil dari tahap TF-IDF.

Tabel 8. Hasil TF-IDF

Term	TFIDF SUM
hilir	168.182916
kemenyan	159.374695
hilir	147.099521
kemenyan	
tidak	42.257998

Tabel 9 menampilkan hasil perhitungan TF-IDF, di mana beberapa kata seperti "hilir", "kemenyan" dan "hilir kemenyan" memiliki nilai TF-IDF yang paling tinggi, yang menunjukkan bahwa kata-kata tersebut memiliki peran penting dalam isu hilirisasi kemenyan.

3.6 Hasil Pengujian Algoritma Naive Bayes

Bagian ini menampilkan hasil evaluasi terhadap kemampuan model dalam mengklasifikasikan *sentiment* positif, negatif, dan netral pada data X yang sudah melalui tahap *preprocessing*. Uji coba model dilakukan dengan membagi dataset menjadi 80% untuk *training data* dan 20% untuk *testing data*. Model dinilai menggunakan beberapa metrik, antara lain akurasi, presisi, recal, serta *F1-score*. Hasil evaluasi ditampilkan pada Gambar 5.

Tabel 9. Hasil Pengujian Naive Bayes

Sentiment	Presisi	Recal	F1-Score
Negatif	0.90	0.13	0.23
Netral	0.63	0.92	0.75
Positif	0.66	0.42	0.51

Dari Tabel 9 dapat dilihat bahwa nilai presisi tertinggi diperoleh pada kelas negatif dengan nilai 0.90, yang menunjukkan bahwa model cukup baik dalam mengidentifikasi data yang benar-benar negatif dari seluruh prediksi negatif. Namun, nilai recall pada kelas negatif masih rendah yaitu 0.13, artinya model masih kurang mampu mengenali seluruh data negatif yang sebenarnya. Kelas netral memiliki nilai recal tertinggi yaitu 0.92 dan *F1-score* tertinggi sebesar 0.75, sehingga dapat dikatakan model paling baik dalam mengenali data netral dibandingkan kelas lainnya. Sementara itu, kelas positif memiliki nilai presisi 0.66 dan recal 0.42 dengan *F1-score* 0.51. Nilai ini menunjukkan performa model dalam mengenali *sentiment* positif berada pada tingkat sedang. Secara keseluruhan, hasil ini menggambarkan bahwa model *Naive Bayes* cenderung lebih baik dalam mengklasifikasikan *sentiment* netral, tetapi masih perlu peningkatan pada kemampuan mendeteksi *sentiment* negatif dan positif agar hasil klasifikasi menjadi lebih seimbang.

4. KESIMPULAN

Hasil dari penelitian ini model *Naive Bayes* mampu untuk klasifikasi *sentiment* dengan cukup baik, meskipun performanya masih bervariasi antar kelas. Nilai akurasi keseluruhan yang diperoleh dari hasil pengujian menunjukkan bahwa model lebih efektif dalam mengenali *sentiment* netral sebesar 55.20% dibandingkan *sentiment* positif sebesar 26.03% maupun negatif sebesar 18.77%. Hal ini dapat disebabkan oleh distribusi data yang tidak seimbang, di mana jumlah tweet dengan label netral lebih dominan dibandingkan dua kelas lainnya. Nilai recall yang rendah pada kelas negatif menunjukkan bahwa model cenderung kesulitan dalam mengenali kata-kata yang merepresentasikan *sentiment* negatif. Faktor ini bisa disebabkan oleh karakteristik bahasa di media sosial X yang cenderung informal, banyak mengandung singkatan, serta penggunaan kata tidak baku yang menyebabkan ambiguitas makna. Selain itu, data negatif yang relatif sedikit membuat model tidak memperoleh cukup variasi dalam proses pelatihan, sehingga sulit mengenali pola *sentiment* negatif secara akurat. Sebaliknya, tingginya recall pada kelas netral menunjukkan bahwa model cukup sensitif dalam mengenali kalimat yang bersifat informatif atau *deskriptif* tanpa kecenderungan emosi tertentu. Hal ini umum terjadi dalam analisis *sentiment* media sosial karena sebagian besar tweet bersifat netral atau hanya menyampaikan informasi tanpa opini. Untuk meningkatkan hasil klasifikasi, beberapa langkah perbaikan dapat dilakukan, seperti penyeimbangan data menggunakan teknik *oversampling* misalnya *SMOTE*, peningkatan jumlah data latih, serta penerapan metode *feature selection* untuk mengurangi kata-kata yang tidak relevan. Alternatif lain adalah mencoba algoritma lain seperti *Support Vector Machine (SVM)* atau *Random Forest* untuk membandingkan performa dengan *Naive Bayes*, dan pengumpulan data di dapat dari media sosial lainnya seperti *YouTube*, *TikTok*. Secara keseluruhan, hasil pengujian menunjukkan bahwa algoritma *Naive Bayes* masih dapat digunakan sebagai model dasar yang cukup baik untuk klasifikasi *sentiment* pada isu hilirisasi kemenyan di media sosial X, terutama dalam konteks penelitian *eksploratif*. Meskipun demikian, pengembangan lebih lanjut masih diperlukan agar model algoritma mampu memberikan hasil yang lebih akurat dan lebih seimbang antar kelas *sentiment* tersebut.

REFERENCES

- [1] I. Budianto and S. N. Anwar, "Analisis Sentimen Pengguna Twitter Mengenai Program Vaksinasi Covid-19 Menggunakan Algoritma Naïve Bayes," *Jurnal Teknologi Informasi*, vol. 6, no. 1, 2022, doi: 10.36294/jurti.v6i1.2551.
- [2] L. Septian, T. Aljauza, and C. Juliane, "Analisis Sentimen Putusan Mahkamah Konstitusi terhadap Batas Usia Capres dan Cawapres Menggunakan IndoBERT," *The Indonesian Journal of Computer Science*, vol. 12, no. 6, 2024, doi: 10.33022/ijcs.v12i6.3614.
- [3] I. Verawati and B. S. Audit, "Algoritma Naïve Bayes Classifier Untuk Analisis Sentimen Pengguna Twitter Terhadap Provider By.U," *Jurnal Media Informatika Budidarma*, vol. 6, no. 3, pp. 1411–1417, 2022, doi: 10.30865/mib.v6i3.4132.
- [4] M. F. Alexandi, "Ekonomi politik kebijakan hilirisasi industri," *Institut Pertanian Bogor*, Jan. 2024, [Online]. Available: <https://fem.ipb.ac.id/rubrik-iktishodia/7554/ekonomi-politik-kebijakan-hilirisasi-industri/>
- [5] E. A. Sianipar, "Potensi Resin Kemenyan (*Styrax benzoin*) dan Senyawa Aktifnya Dalam Pengobatan Penyakit," *Pharmaceutical and Biomedical Sciences Journal (PBSJ)*, vol. 5, no. 1, pp. 17–22, Jun. 2023, doi: 10.15408/pbsj.v5i1.30202.
- [6] Ridwansyah, T. C. Sunarti, K. Syamsu, F. Fahma, and E. Julianti, "Kandungan Kimia Kemenyan Sumatra Utara (*Styrax benzoin*) dan Prospek Pengembangannya," *Agrointek: Jurnal Teknologi Industri Pertanian*, vol. 19, no. 3, pp. 530–539, 2025, doi: 10.21107/agrointek.v19i3.27527.
- [7] N. Samuel and A. A. Pekuwalu, "Pengenalan Pola Tulisan Tangan Resep Dokter Menggunakan Metode Naïve Bayes Classifier Pada Puskesmas Kambaniru," *Jurnal Teknologi dan Komputer*, vol. 2, no. 1, pp. 55–61, 2022, doi: 10.57152/malcom.v2i1.174.
- [8] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naïve Bayes Untuk Analisis Sentimen Review Data Twitter BMKG Nasional," *Jurnal Tekno Kompak*, vol. 15, no. 1, pp. 131–145, 2021, doi: 10.33365/jtk.v15i1.744.
- [9] A. Wandani, Fauziah, and Andrianingsih, "Analisis Sentimen Pengguna Twitter Pada Event Flash Sale Menggunakan Algoritma K-NN, Random Forest, dan Naïve Bayes," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 5, no. 2, pp. 651–665, 2021, doi: 10.33319/jsakti.v5i2.463.
- [10] M. B. Prakoso, I. Cholissodin, and Indriati, "Analisis Sentimen Masyarakat Terhadap Sistem Pembelajaran Online Selama Pandemi Covid-19 Berdasarkan Twitter Menggunakan Metode Naïve Bayes," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 12, pp. 5376–5383, 2021, doi: 10.30865/mib.v5i12.2580.
- [11] N. Hardi, Y. Alkahfi, P. Handayani, W. Gata, and M. R. Firdaus, "Analisis Sentimen Physical Distancing Pada Twitter Menggunakan Text Mining Dengan Algoritma Naïve Bayes Classifier," *SISTEMASI: Jurnal Sistem Informasi*, vol. 10, no. 1, pp. 131–138, 2021, doi: 10.32520/stmsi.v10i1.1118.
- [12] W. Ningsih, B. Alfianda, R. Rahmaddeni, and D. Wulandari, "Perbandingan Algoritma SVM dan Naïve Bayes Dalam Analisis Sentimen Twitter Pada Penggunaan Mobil Listrik di Indonesia," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 556–562, 2024, doi: 10.57152/malcom.v4i2.1253.
- [13] A. H. Ruger, M. Suyanto, and M. P. Kurniawan, "Analisis Sentimen Pelanggan Shopee di Twitter Menggunakan Algoritma Naïve Bayes," *JIFOTECH (Journal of Information Technology)*, vol. 1, no. 2, 2021, doi: 10.33197/j-titar.vol8no1.298.
- [14] B. Ramadhani and R. R. Suryono, "Komparasi Algoritma Naïve Bayes dan Logistic Regression Untuk Analisis Sentimen *Metaverse*," *Jurnal Media Informatika Budidarma*, vol. 8, no. 2, p. 714, 2024, doi: 10.30865/mib.v8i2.7458.
- [15] R. A. Fauzi, I. Cholissodin, and B. Rahayudi, "Pemanfaatan Spark Untuk Analisis Sentimen Mengenai Netralitas Berita Dalam Membahas Pemilu Presiden 2019 Menggunakan Metode Naïve Bayes Classifier," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 3, pp. 1070–1077, 2021, doi: 10.21609/jiki.v13i2.812.
- [16] R. Singh and R. Singh, "Applications Of Sentiment Analysis and Machine Learning Techniques In Disease Outbreak Prediction – A review," *Materials Today: Proceedings*, vol. 81, pp. 1006–1011, 2023, doi: 10.1016/j.matpr.2021.04.356.



- [17] A. Prasetya, Ferdiansyah, Y. N. Kunang, E. S. Negara, and W. Chandra, “Analisis Sentimen Terhadap Cryptocurrency Berdasarkan Komentar dan Balasan Pada Platform Twitter,” *Journal of Information Systems and Informatics*, vol. 3, no. 2, 2021, doi: 10.51519/jsisfo.v3i2.195.
- [18] Y. Durachman, S. J. Putra, H. Nanang, and H. T. Sukmana, “Analysis Sentiment Of Public Opinion On Social Media Using Naïve Bayes and TF-IDF algorithms,” in *Proc. 3rd Int. Conf. Creative Communication and Innovative Technology (ICCIT)*, IEEE, ,pp. 1–6, 2024, doi: 10.1109/ICCIT62134.2024.10701191.
- [19] S. D. Rehatta, E. Sedyono, and I. Sembiring, “Analisis Penyebaran Informasi Vaksin Covid-19 Pada Twitter Menggunakan Kolaborasi SNA dan Sentiment Analysis,” *Jurnal Media Informatika Budidarma*, vol. 6, no. 2, p. 1145, 2022, doi: 10.30865/mib.v6i2.3955.
- [20] A. N. Badri, N. Noviandi, F. Anastya, and M. Roland, “Analisis Sentimen Untuk Identifikasi Kepuasan Masyarakat Terhadap Kenaikan BBM Menggunakan Algoritma Naïve Bayes,” *JIKO (Jurnal Informatika dan Komputer)*, vol. 7, no. 2, p. 287, 2023, doi: 10.26798/jiko.v7i2.873.
- [21] F. S. Utomo, “Algoritma Naïve Bayes Untuk Analisis Sentimen Review Blibli.com di Google Play Store,” *SISTEMASI*, vol. 13, no. 2, pp. 831–840, 2024, doi: 10.32520/stmsi.v13i2.
- [22] Salahuddin, I. Sabila, and Amirullah, “Analisis dan Implementasi Sistem Penilaian Ulasan Dengan Teknik Sentiment Analysis Berbasis Machine Learning Untuk Peningkatan Feedback Pemilik Cafe,” *Jurnal Teknologi dan Komputer*, vol. 2, no. 2, 2024, doi: 10.30811/jim.v9i2.5980.
- [23] W. F. Sari, R. Rahim, and F. Adrianto, “Analisis Sentimen Review Pengguna BCA Mobile Menggunakan Text Mining,” *Jurnal Teknologi Informasi dan Komunikasi*, vol. 6, no. 2, 2023, doi: 10.33365/jusiti.v10i4.1795.