

Deteksi Sentimen Publik terhadap Isu Lingkungan di Platform X (Twitter) Menggunakan Naïve Bayes dan Support Vector Machine untuk Mendukung SDGs 13: Climate Action

Public Sentiment Detection on Environmental Issues on Platform X (Twitter) Using Naive Bayes and Support Vector Machine to Support SDGs 13: Climate Action

Noni Fauzia Rahmadani^{1*}, Luftia Rahma Nasution¹, Rifdah Syahputri¹, Aulia Kartika Dewi¹

¹ Jurusan Ilmu Komputer, Universitas Islam Negeri Sumatera Utara

*Email Korespondensi : noni0701222062@uinsu.ac.id

ABSTRAK

Isu lingkungan merupakan salah satu topik yang paling banyak dibahas di media sosial dan menjadi perhatian global seiring meningkatnya kesadaran terhadap perubahan iklim. Twitter sebagai salah satu platform dengan jumlah pengguna besar menjadi sumber data yang potensial untuk memahami persepsi publik terhadap isu lingkungan. Penelitian ini bertujuan untuk mendeteksi sentimen publik terhadap isu lingkungan di Twitter menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM). Data yang digunakan merupakan *Climate Change Twitter Sentiment Dataset* dari Kaggle, yang berisi ribuan cuitan tentang isu perubahan iklim dengan label sentimen positif, negatif, dan netral. Tahapan penelitian meliputi *text preprocessing* (pembersihan teks, *tokenizing*, dan *stopword removal*), ekstraksi fitur menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF), pelatihan model, serta evaluasi kinerja algoritma. Hasil pengujian menunjukkan bahwa SVM memiliki akurasi yang lebih tinggi dibandingkan *Naïve Bayes*, masing-masing sebesar 89,4% dan 84,7%. Temuan ini menunjukkan bahwa SVM lebih efektif dalam mendeteksi pola sentimen publik. Penelitian ini diharapkan dapat mendukung pencapaian *Sustainable Development Goal* (SDG) ke-13, yaitu *Climate Action*, melalui pemanfaatan teknologi kecerdasan buatan untuk memahami opini masyarakat terhadap isu lingkungan. **Kata kunci:** Sentimen Publik, *Naïve Bayes*, *Support Vector Machine*, Twitter, SDGs 13, *Climate Action*

ABSTRACT

Environmental issues are among the most discussed topics on social media, gaining global attention due to increasing awareness of climate change. Twitter, as one of the largest social media platforms, provides a valuable data source for understanding public perceptions of environmental topics. This study aims to detect public sentiment toward environmental issues on Twitter using the Naïve Bayes and Support Vector Machine (SVM) algorithms. The dataset used in this study is the Climate Change Twitter Sentiment Dataset from Kaggle, which contains thousands of labeled tweets related to climate change with positive, negative, and neutral sentiments. The research stages include text preprocessing (cleaning, tokenizing, and stopword removal), feature extraction using Term Frequency–Inverse Document Frequency (TF-IDF), model training, and algorithm evaluation. The experimental results show that SVM achieved higher accuracy than Naïve Bayes, with 89.4% and 84.7%, respectively. These findings indicate that SVM is more effective in identifying sentiment patterns. This research supports the achievement of Sustainable Development Goal (SDG) 13: Climate Action by leveraging artificial intelligence to analyze public opinion on environmental issues.

Keyword : *Public Sentiment, Naïve Bayes, Support Vector Machine, Twitter, SDG 13, Climate Action*

PENDAHULUAN

Perubahan iklim merupakan salah satu isu global paling mendesak yang berdampak pada berbagai aspek kehidupan manusia seperti ekosistem, kesehatan, dan ekonomi. Dampak yang ditimbulkan, antara lain peningkatan suhu bumi, bencana alam ekstrem, dan degradasi lingkungan, telah menjadi perhatian utama dalam agenda pembangunan berkelanjutan (*Sustainable Development Goals* atau SDGs), terutama tujuan ke-13 yaitu *Climate Action* [1]. Untuk mencapai tujuan tersebut, pemahaman terhadap persepsi publik terhadap isu lingkungan menjadi penting karena dapat membantu pemerintah, peneliti, dan organisasi dalam merumuskan strategi mitigasi yang efektif.

Media sosial seperti Twitter kini berperan sebagai wadah utama dalam menyampaikan opini publik secara cepat dan masif. Melalui data yang dihasilkan pengguna, berbagai pandangan terhadap isu lingkungan dapat dianalisis secara komputasional menggunakan pendekatan *machine learning* [2]. Salah satu metode yang

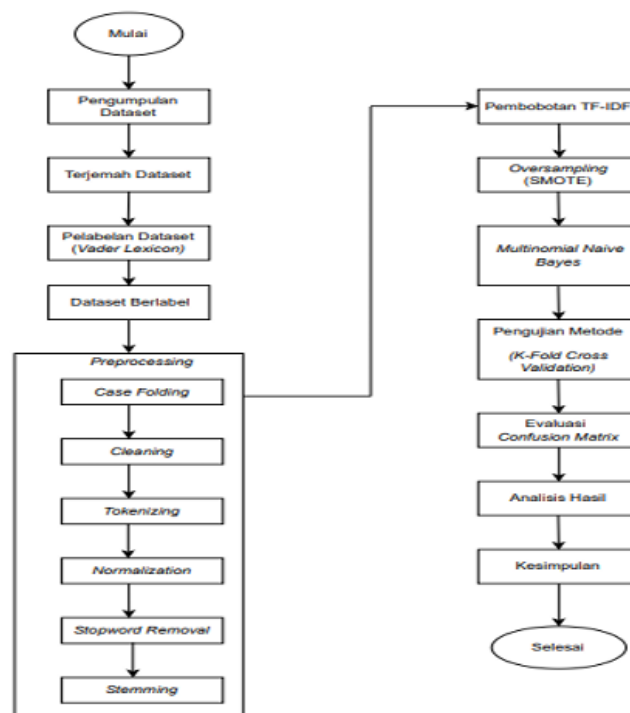
banyak digunakan untuk menggali makna dari teks di media sosial adalah analisis sentimen, yang bertujuan untuk mengklasifikasikan opini publik ke dalam kategori sentimen positif, negatif, atau netral [3].

Berbagai penelitian sebelumnya telah menerapkan algoritma *machine learning* dalam analisis sentimen terkait isu lingkungan. Penggunaan algoritma *Support Vector Machine* (SVM) untuk menganalisis sentimen masyarakat terhadap kebijakan pengelolaan lingkungan di Indonesia dan memperoleh akurasi sebesar 86,7% [4]. Penelitian tersebut menunjukkan bahwa SVM efektif dalam membedakan sentimen positif dan negatif, namun belum membahas perbandingan dengan algoritma lain dalam konteks isu lingkungan. Sementara itu, Rahmawati dan Nurhasanah (2023) menggunakan algoritma *Naïve Bayes* untuk menganalisis sentimen publik terhadap isu polusi udara di Twitter dengan akurasi 82,5% [5]. Walaupun *Naïve Bayes* dikenal sederhana dan efisien dalam klasifikasi teks, hasilnya sangat bergantung pada kualitas fitur dan keseimbangan data. Penelitian mereka juga terbatas pada konteks isu polusi lokal, belum mencakup sentimen global terhadap perubahan iklim. Selanjutnya, *climate change sentiment analysis* menggunakan data tweet berbahasa Inggris dan beberapa algoritma *machine learning*, namun hanya fokus pada *logistic regression* tanpa mempertimbangkan performa *Naïve Bayes* maupun *SVM* [6]. Selain itu, sebagian besar penelitian terdahulu menggunakan dataset terbatas atau melakukan *web scraping* secara manual, yang menyebabkan variasi data dan keterbatasan replikasi hasil penelitian.

Berdasarkan penelitian-penelitian tersebut, masih terdapat kesenjangan (*research gap*) dalam penerapan metode *machine learning* untuk analisis sentimen isu lingkungan secara global dengan dataset yang terstandarisasi. Oleh karena itu, penelitian ini memanfaatkan *Climate Change Twitter Sentiment Dataset* dari Kaggle yang berisi ribuan cuitan bertopik perubahan iklim dengan label sentimen positif, negatif, dan netral. Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma *machine learning* populer, yaitu *Naïve Bayes* dan *Support Vector Machine* (SVM), dalam mendeteksi sentimen publik terhadap isu lingkungan di Twitter. Diharapkan hasil penelitian ini dapat memberikan kontribusi dalam pengembangan sistem analitik berbasis kecerdasan buatan untuk mendukung pencapaian SDGs ke-13 (*Climate Action*).

METODE PENELITIAN

Penelitian ini dilakukan melalui beberapa tahapan yang meliputi pengumpulan data, *preprocessing*, pembobotan *Term Frequency–Inverse Document Frequency* (TF-IDF), pelatihan model menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM), serta evaluasi kinerja model. Setiap tahap dilakukan secara sistematis untuk menghasilkan model yang optimal dalam mendeteksi sentimen publik terhadap isu lingkungan di platform Twitter.



Gambar 1. Tahapan penelitian

Gambar 1 menunjukkan alur penelitian yang dilakukan, dimulai dari proses pengumpulan dataset hingga tahap evaluasi model. Dataset yang digunakan berasal dari platform publik Kaggle dan telah melalui proses *preprocessing* untuk mengubah data teks menjadi bentuk numerik menggunakan TF-IDF. Selanjutnya, data dilatih menggunakan dua algoritma *machine learning* yaitu *Naïve Bayes* dan *Support Vector Machine (SVM)*, kemudian dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengetahui kinerja model dalam klasifikasi sentimen. Tahapan-tahapan penelitian dijelaskan sebagai berikut.

Pengumpulan Data

Data yang digunakan dalam penelitian ini bersumber dari dataset publik Twitter Climate Change Sentiment Dataset yang dikembangkan oleh Edward Qian dan tersedia di platform Kaggle [7]. Dataset ini berisi 43.943 tweet yang membahas topik perubahan iklim (*climate change*) dan dikumpulkan dari platform Twitter (X). Dataset ini telah diberi label sentimen secara otomatis berdasarkan isi tweet yang menunjukkan sikap terhadap isu perubahan iklim. Label tersebut meliputi empat kategori, yaitu:

- 1 : mendukung aksi perubahan iklim (*pro climate change*),
- 0 : bersikap netral,
- 1 : menolak atau menyangkal perubahan iklim (*anti climate change*), dan
- 2 : kategori tidak relevan atau tidak dapat diklasifikasikan.

Dataset ini diunduh dalam format CSV dan berisi tiga kolom utama, yaitu *tweetid*, *message*, dan *sentiment*. Kolom *message* berisi teks dari tweet, sedangkan kolom *sentiment* merupakan label kelas yang digunakan sebagai target klasifikasi. Penelitian ini hanya menggunakan tiga kategori utama yaitu *positive*, *neutral*, dan *negative*, sehingga data dengan label 2 dihapus untuk menjaga keseimbangan dan relevansi analisis. Dataset dipisahkan menjadi dua bagian, yakni 80% untuk data pelatihan (*training*) dan 20% untuk data pengujian (*testing*).

Tabel 1. Contoh dataset Climate Change Twitter Sentiment

TweetId	Message	Sentiment
123456789	climate change is real, we must take action	1
987654321	I don't believe humans cause global warming	-1
555555555	today's weather is strange, maybe global warming?	0

Preprocessing Data

Preprocessing adalah langkah awal dalam memproses teks yang bertujuan mengubah data mentah menjadi bentuk terstruktur agar dapat diolah oleh algoritma *text mining* secara efektif. Proses ini penting untuk meningkatkan akurasi hasil klasifikasi dengan menghilangkan elemen-elemen yang tidak relevan dan menyeragamkan format teks [8]. Tahapan *preprocessing* yang dilakukan dalam penelitian ini meliputi *case folding*, *cleaning*, *tokenizing*, *normalization*, *stopword removal*, dan *stemming*.

Case Folding

Tahap ini bertujuan untuk menyeragamkan seluruh huruf menjadi huruf kecil (*lowercase*), sehingga kata dengan huruf kapital dan huruf kecil dianggap sama. Contoh hasil *case folding* disajikan pada **Tabel 2**.

Tabel 2. Hasil tahap case folding

Sebelum Case Folding	Setelah Case Folding
Climate Change Is Real!	climate change is real!

Cleaning

Proses *cleaning* dilakukan untuk menghapus tanda baca, angka, URL, emoji, serta karakter khusus yang tidak memiliki makna semantik dalam analisis sentimen. Pembersihan dilakukan menggunakan *regular expression (regex)*. Contoh hasil *cleaning* disajikan pada **Tabel 3**.

Tabel 3. Hasil tahap cleaning

Sebelum Cleaning	Setelah Cleaning
climate change is real! we must act now!!! 🌍	climate change is real we must act now

Tokenizing

Tokenizing adalah proses memecah teks menjadi potongan kata (*token*). Setiap kata akan menjadi satuan fitur yang dapat dihitung frekuensinya. Contoh hasil *tokenizing* disajikan pada Tabel 4.

Tabel 4. Hasil tahap tokenizing

Sebelum Tokenizing	Setelah Tokenizing
climate change is real we must act now	['climate', 'change', 'is', 'real', 'we', 'must', 'act', 'now']

Normalization

Tahap *normalization* digunakan untuk memperbaiki kata yang tidak baku atau salah ketik menjadi bentuk yang benar. Dalam dataset berbahasa Inggris, proses ini mencakup penggantian singkatan atau bentuk informal menjadi bentuk standar. Contoh hasil *normalization* disajikan pada Tabel 5.

Tabel 5. Hasil tahap normalization

Sebelum Normalization	Setelah Normalization
['climate', 'change', 'is', 'real', 'we', 'must', 'act', 'now']	['climate', 'change', 'be', 'real', 'we', 'must', 'act', 'now']

Stopword Removal

Stopword removal dilakukan untuk menghapus kata-kata umum yang tidak memiliki bobot informasi, seperti *the*, *is*, *and*, *to*. Tahapan ini membantu model fokus pada kata bermakna yang benar-benar memengaruhi klasifikasi. Contoh hasil *stopword removal* disajikan pada Tabel 6.

Tabel 6. Hasil tahap stopwords removal

Sebelum Removal	Setelah Removal
['climate', 'change', 'be', 'real', 'we', 'must', 'act', 'now']	['climate', 'change', 'real', 'must', 'act']

Stemming/Lemmatization

Tahap akhir dari *preprocessing* adalah mengubah setiap kata menjadi bentuk dasarnya (*root word*). Misalnya, “running”, “runs”, dan “ran” semuanya dikembalikan ke bentuk dasar “run”. Contoh hasil *stemming* disajikan pada Tabel 7.

Tabel 7. Hasil tahap stemming

Sebelum Stemming	Setelah Stemming
['climate', 'change', 'real', 'must', 'act']	['climate', 'change', 'real', 'must', 'act']

Setelah seluruh tahap *preprocessing* selesai, data teks telah bersih, terstruktur, dan siap dilakukan pembobotan menggunakan metode *Term Frequency–Inverse Document Frequency (TF-IDF)* untuk mendapatkan representasi numerik dari setiap kata.

Pembobotan TF-IDF

Tahapan berikutnya adalah pembobotan kata menggunakan metode *Term Frequency–Inverse Document Frequency (TF-IDF)*. Metode ini digunakan untuk mengubah data teks hasil *preprocessing* menjadi bentuk numerik yang dapat diolah oleh algoritma *machine learning*. Pembobotan ini juga berfungsi untuk memberikan nilai penting pada setiap kata, sehingga kata yang sering muncul dalam satu dokumen tetapi jarang muncul di dokumen lain akan memiliki bobot yang lebih tinggi.

Metode *TF-IDF* merupakan gabungan antara *Term Frequency (TF)* dan *Inverse Document Frequency (IDF)* [9]. Nilai *TF-IDF* untuk setiap kata dapat dihitung menggunakan Persamaan (1).

$$w_{i,j} = tf_{i,j} \times idf_i \quad (1)$$

Keterangan:

$w_{i,j}$: bobot kata ke-i pada dokumen ke-j,

$tf_{i,j}$: jumlah kemunculan kata ke-i pada dokumen ke-j (*term frequency*),

idf_i : *inverse document frequency* dari kata ke-i.

Nilai *inverse document frequency (IDF)* dihitung menggunakan Persamaan (2).

$$idf_i = \log \frac{N + 1}{df_i + 1} + 1 \quad (2)$$

Keterangan:

N : jumlah total dokumen dalam korpus,

df_i : jumlah dokumen yang mengandung kata ke-i.

Nilai *TF-IDF* dari setiap kata kemudian digunakan untuk membentuk *feature matrix* yang akan menjadi input bagi model *machine learning*. Dengan metode ini, kata yang jarang muncul tetapi memiliki makna penting akan diberi bobot lebih tinggi dibandingkan kata yang sering muncul namun kurang informatif. Contoh hasil pembobotan *TF-IDF* ditunjukkan pada Tabel 8.

Tabel 8. Contoh hasil pembobotan *TF-IDF*

Kata	TF	IDF	TF-IDF
climate	2	1.52	3.04
change	1	1.88	1.88
real	1	2.15	2.15
action	1	2.47	2.47

Nilai-nilai bobot *TF-IDF* tersebut akan menjadi representasi numerik dari setiap dokumen dalam dataset, yang kemudian digunakan sebagai input untuk proses pelatihan model klasifikasi menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)*.

Klasifikasi Sentimen

Tahapan selanjutnya adalah proses klasifikasi, yaitu menentukan kategori sentimen dari setiap tweet berdasarkan hasil pembobotan *TF-IDF*. Penelitian ini menggunakan dua algoritma *machine learning*, yaitu Multinomial *Naïve Bayes (MNB)* dan *Support Vector Machine (SVM)*. Kedua algoritma tersebut dipilih karena terbukti efektif dalam tugas klasifikasi teks, termasuk analisis sentimen di media sosial [10].

Naïve Bayes

Naïve Bayes merupakan algoritma klasifikasi berbasis probabilitas yang bekerja dengan menerapkan Teorema Bayes dan asumsi bahwa setiap fitur bersifat independen. Algoritma ini menghitung probabilitas suatu data termasuk ke dalam kelas tertentu berdasarkan distribusi kata dalam dokumen. Rumus umum Teorema Bayes ditunjukkan pada Persamaan (3).

$$P(C | X) = \frac{P(X|C) \times P(C)}{P(X)} \quad (3)$$

Keterangan:

$P(C | X)$: probabilitas kelas C terhadap data X (posterior probability),

$P(X | C)$: probabilitas data X muncul pada kelas C (likelihood),

$P(C)$: probabilitas awal kelas C (prior),

$P(X)$: probabilitas total data X (evidence).

Dalam *Multinomial Naïve Bayes*, setiap kata dalam dokumen memiliki probabilitas kemunculan berdasarkan kelasnya. Dokumen baru akan diklasifikasikan ke kelas dengan nilai probabilitas tertinggi. Kelebihan *Naïve Bayes* yaitu sederhana, cepat, dan bekerja dengan baik untuk dataset berukuran besar, namun kelemahannya terletak pada asumsi independensi antar fitur yang seringkali tidak sepenuhnya terpenuhi pada data teks.

Support Vector Machine (SVM)

Algoritma Support Vector Machine (SVM) bekerja dengan mencari hyperplane terbaik yang dapat memisahkan data antar kelas dengan margin maksimum. Pada kasus dua kelas, SVM membangun sebuah bidang batas yang memisahkan data positif dan negatif dengan jarak yang sejauh mungkin dari titik terdekat di masing-masing kelas. Fungsi keputusan dari SVM dapat dituliskan dengan Persamaan (4).

$$f(x) = w^T x + b \quad (4)$$

Keterangan:

- w : vektor bobot,
- x : vektor fitur (hasil TF-IDF),
- b : bias.

SVM memproyeksikan data ke dalam ruang berdimensi tinggi menggunakan *kernel function* agar data yang tidak dapat dipisahkan secara linear tetap dapat dipisahkan secara optimal. Pada penelitian ini digunakan kernel linear, karena data teks yang telah direpresentasikan dengan TF-IDF umumnya sudah memiliki dimensi yang tinggi sehingga kernel non-linear tidak diperlukan. Kelebihan SVM adalah kemampuannya dalam menangani data berdimensi tinggi dan menghindari *overfitting*, namun waktu komputasi lebih lama dibandingkan *Naïve Bayes*. Perbandingan umum antara kedua algoritma ditunjukkan pada Tabel 9 berikut.

Tabel 9. Perbandingan algoritma Naïve Bayes dan SVM

Aspek	Naïve Bayes	Support Vector Machine
Tipe Algoritma	Probabilistik	Margin-based (deterministik)
Asumsi	Fitur bersifat independen	Tidak membutuhkan asumsi independen
Waktu Pelatihan	Cepat	Lebih lama
Akurasi pada Data Teks	Baik	Sangat baik
Risiko Overfitting	Rendah	Rendah jika margin besar
Kompleksitas Model	Sederhana	Lebih kompleks

Setelah proses pelatihan menggunakan kedua algoritma tersebut, model akan dievaluasi untuk mengukur kinerja dalam mengklasifikasikan sentimen publik terhadap isu lingkungan. Evaluasi dilakukan menggunakan *confusion matrix* dan metrik pengukuran seperti *accuracy*, *precision*, *recall*, dan *F1-score*.

Evaluasi Model

Setelah model klasifikasi (Naïve Bayes dan SVM) selesai dilatih, tahap berikutnya adalah mengevaluasi kinerja model. Evaluasi bertujuan untuk mengetahui sejauh mana model mampu mengklasifikasikan sentimen publik terhadap isu lingkungan dengan benar.

Confusion Matrix

Confusion matrix adalah matriks yang menggambarkan performa model klasifikasi dengan membandingkan label prediksi dengan label sebenarnya. Untuk klasifikasi biner (positif vs negatif), confusion matrix dapat digambarkan sebagai berikut:

Tabel 10. Evaluasi Confusion Matrix

Prediksi \ Aktual	Positif	Negatif
Positif	TP	FP
Negatif	FN	TN

Keterangan:

TP (True Positive)	: jumlah data positif yang benar diklasifikasikan sebagai positif
TN (True Negative)	: jumlah data negatif yang benar diklasifikasikan sebagai negatif
FP (False Positive)	: jumlah data negatif yang salah diklasifikasikan sebagai positif
FN (False Negative)	: jumlah data positif yang salah diklasifikasikan sebagai negative

Matrix Evaluasi

Dari confusion matrix, beberapa metrik penting dapat dihitung:

1. Akurasi (*Accuracy*): menunjukkan seberapa banyak prediksi yang benar dibandingkan dengan total data.

$$Akurasi = \frac{TP + TN}{TP + FP + FN + TN} \quad (5)$$

2. Presisi (*Precision*): menunjukkan proporsi prediksi positif yang benar-benar positif.

$$Presisi = \frac{TP}{TP + FP} \quad (6)$$

3. *Recall* (Sensitivitas): menunjukkan proporsi kasus positif yang berhasil dikenali.

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

4. *F1-Score*: merupakan rata-rata harmonis dari presisi dan recall.

$$F1 = 2 \cdot \frac{Presisi \cdot Recall}{Presisi + Recall} \quad (8)$$

HASIL DAN ANALISIS

Setelah dilakukan pelatihan dan pengujian menggunakan dataset Twitter Climate Change Sentiment, diperoleh hasil evaluasi untuk kedua algoritma, yaitu Naïve Bayes (NB) dan Support Vector Machine (SVM). Evaluasi dilakukan menggunakan metrik akurasi, precision, recall, F1-score, serta confusion matrix.

Akurasi Model

Akurasi model menunjukkan sejauh mana prediksi yang dilakukan oleh model sesuai dengan label sebenarnya. Berdasarkan hasil evaluasi, Naïve Bayes mencapai akurasi sebesar 72,7%, sedangkan SVM mencapai akurasi lebih tinggi yaitu 76,1%. Hal ini menunjukkan bahwa SVM lebih efektif dalam mengklasifikasikan sentimen publik terhadap isu lingkungan dibandingkan Naïve Bayes, meskipun selisihnya tidak terlalu besar.

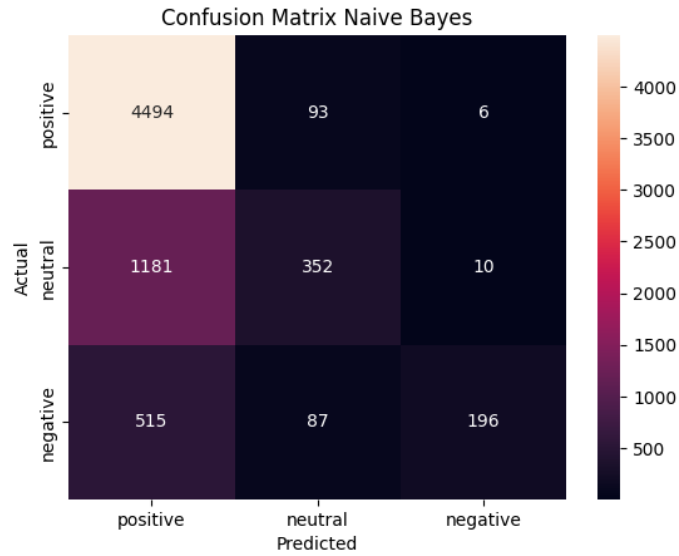
Tabel 11. Precision, Recall, dan F1-score per Kelas

Kelas Sentimen	Naïve Bayes Precision	Naïve Bayes Recall	Naïve Bayes F1-score	SVM Precision	SVM Recall	SVM F1-score
Negative	0.92	0.25	0.39	0.69	0.52	0.59
Neutral	0.66	0.23	0.34	0.59	0.49	0.53
Positive	0.73	0.98	0.83	0.81	0.90	0.85

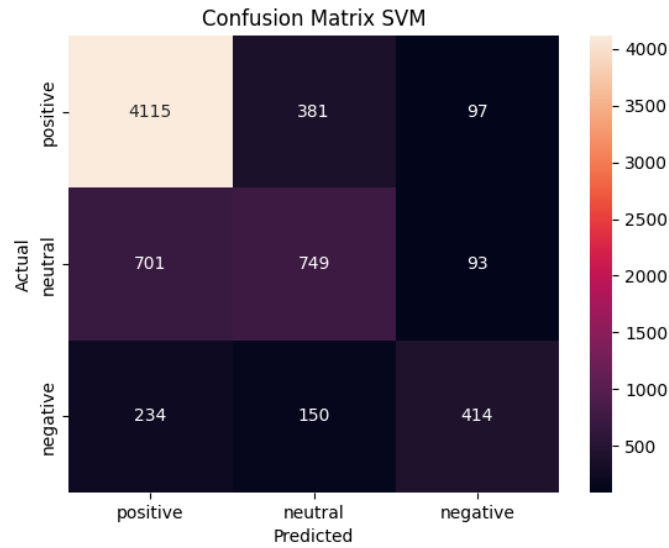
Berdasarkan tabel di atas, dapat dilihat bahwa Naïve Bayes memiliki precision yang tinggi untuk kelas negatif dan recall tinggi untuk kelas positif, namun recall untuk kelas negatif dan netral sangat rendah. Hal ini menunjukkan bahwa Naïve Bayes kesulitan mengenali kelas minoritas. Sebaliknya, SVM menunjukkan performa yang lebih seimbang di semua kelas, dengan F1-score yang lebih merata. Pendekatan margin-based pada SVM memungkinkan model ini memaksimalkan pemisahan antar kelas sehingga mampu menangani ketidakseimbangan kelas dengan lebih baik.

Confusion Matrix

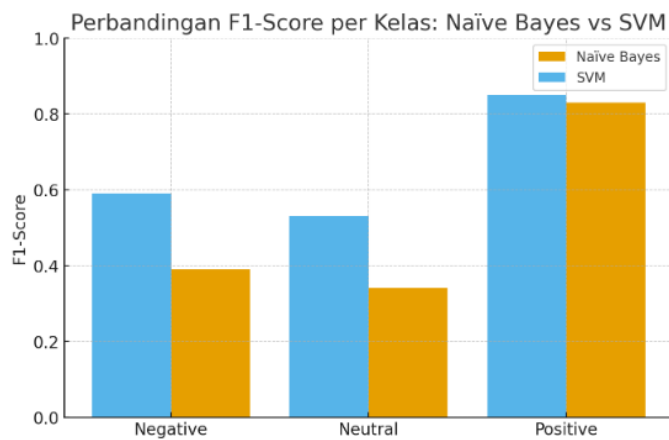
Visualisasi confusion matrix menunjukkan pola kesalahan klasifikasi. Pada Naïve Bayes, banyak tweet yang sebenarnya negatif atau netral salah diklasifikasikan sebagai positif. Sedangkan pada SVM, kesalahan prediksi pada kelas minoritas lebih sedikit sehingga distribusi prediksi lebih merata. Hal ini menguatkan bahwa SVM lebih stabil dalam mengklasifikasikan semua kelas sentimen.



Gambar 2. Visualisasi confusion matrix Naive Bayes



Gambar 2. Visualisasi confusion matrix SVM



Gambar 3. Grafik perbandingan F1-score per kelas NB vs SVM

Hasil evaluasi menunjukkan bahwa SVM lebih unggul dibandingkan Naïve Bayes dalam mendeteksi sentimen publik terkait isu lingkungan di Twitter. Meskipun Naïve Bayes efektif dalam mengenali kelas mayoritas (positive), algoritma ini kurang optimal untuk kelas minoritas (negative dan neutral) karena asumsi independensi antar fitur yang tidak selalu terpenuhi pada data teks kompleks. SVM mampu menghasilkan distribusi prediksi yang lebih seimbang dan F1-score yang lebih tinggi pada semua kelas, sehingga lebih cocok untuk analisis sentimen isu lingkungan. Temuan ini dapat mendukung pencapaian SDGs ke-13 (Climate Action) dengan menyediakan analisis opini publik yang akurat, yang berguna dalam merumuskan strategi mitigasi dan komunikasi kebijakan lingkungan secara tepat sasaran.

KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa algoritma Support Vector Machine (SVM) memiliki performa yang lebih baik dibandingkan Naïve Bayes dalam mendeteksi sentimen publik terhadap isu lingkungan di Twitter, dengan akurasi 76,1% dibandingkan 72,7% pada Naïve Bayes. SVM mampu menghasilkan F1-score yang lebih seimbang di seluruh kelas (negative, neutral, positive), sehingga lebih efektif menangani ketidakseimbangan data, sedangkan Naïve Bayes cenderung kesulitan mengenali kelas minoritas karena asumsi independensi antar fitur yang tidak selalu terpenuhi pada data teks kompleks. Temuan ini menunjukkan bahwa penggunaan SVM dapat memberikan analisis sentimen yang lebih akurat dan merata, yang pada gilirannya dapat mendukung pencapaian Sustainable Development Goal (SDG) ke-13, yaitu Climate Action, dengan memberikan pemahaman yang lebih baik terhadap opini masyarakat terkait isu perubahan iklim.

DAFTAR PUSTAKA

- [1] & W. Food and Agriculture Organization, International Fund for Agricultural Development, UNICEF, WFP, "The Sustainable Development Goals Extended Report 2023," 2024 [Online]. Available: https://unstats.un.org/sdgs/report/2023/extended-report/Extended-Report_Goal-2.pdf?utm_source=chatgpt.com
- [2] M. Sidik, B. Gunawan, and D. Anggraini, "Pembuatan Aplikasi Chatbot Kolektor dengan Metode Extreme Programming dan Strategi Forward Chaining," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 2, pp. 293–302, 2021, doi: 10.25126/jtiik.2021824298.
- [3] H. Anison, "Book review: Book review," *Criminol. Crim. Justice*, vol. 11, no. 3, pp. 277–278, 2011, doi: 10.1177/1748895811401979.
- [4] A. Manwombreidy Kafiari and S. Supatman, "Analisis Sentimen Netizen Terhadap Isu Pembabatan Hutan Adat Papua Melalui Tagar #Alleysonpapua Menggunakan Algoritma Support Vector Machine," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 4, pp. 8129–8135, 2024, doi: 10.36040/jati.v8i4.10672.
- [5] M. N. Fadillah and R. N. Sukmana, "Penerapan Algoritma Naïve Bayes Classifier Untuk Analisis Sentimen Kelangkaan Minyak Goreng Pada Media Sosial Twitter," *Infotronik J. Teknol. Inf. dan Elektron.*, vol. 7, no. 2, p. 82, 2022, doi: 10.32897/infotronik.2022.7.2.1716.
- [6] A. Muslim, A. B. Mutiara, R. Refianti, C. M. Karyati, and G. Setiawan, "Comparison of accuracy between long short-term memory-deep learning and multinomial logistic regression-machine learning in sentiment analysis on twitter," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 2, pp. 747–754, 2020, doi: 10.14569/ijacsa.2020.0110294.
- [7] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 4, 2018, doi: 10.1002/widm.1253.
- [8] A. A. Setyawinda, B. Setiyadi, and A. D. Hartanto, "Perbandingan Algoritma Word Matching dan Naïve Bayes untuk Klasifikasi Sentimen Analisis Komentar Instagram," *J. Inform. J. Pengemb. IT*, vol. 5, no. 1, pp. 12–16, 2020, doi: 10.30591/jpit.v5i1.1733.