

Analisis Opini Publik Tentang Cuaca Ekstrem Menggunakan Naïve Bayes dan K-Nearest Neighbor

Christian Richie Wijaya¹, Hansen², Hafiz Irsyad³

¹²³Program Studi Informatika, Fakultas Ilmu Komputer dan Rekayasa Universitas Multi Data Palembang

richiewijaya@mhs.mdp.ac.id , hansentan02193@mhs.mdp.ac.id , hafizirsyad@mdp.ac.id

ABSTRACT

Indonesia's location on the equator can make it vulnerable to extreme weather phenomena that can pose a threat. Extreme weather is a physical atmospheric phenomenon at a certain time within a short period of time and is also extreme. YouTube can be a medium for the public to convey opinions to the government in dealing with extreme weather. Sentiment analysis is a method used to extract, analyze and process textual data automatically to identify sentiment in an opinion. This research was conducted with the aim of analyzing the sentiment of public comments on YouTube regarding extreme weather using the method of Naïve Bayes Multinomial and KNN. The dataset was taken from comments on YouTube regarding extreme weather, totaling 1030 comments. The results show that the Naïve Bayes Multinomial method outperforms the K-Nearest Neighbor method, achieving an accuracy of 81%, a precision of 76%, a recall of 63%, and an F1-Score of 69%, compared to K-Nearest Neighbor, which only achieved an accuracy of 51%, a precision of 38%, a recall of 69%, and an F1-Score of 48%. Based on these findings, the Naïve Bayes Multinomial method is recommended for sentiment analysis of public opinion on extreme weather..

Keywords: Extreme weather, Sentiment analysis, Naïve bayes multinomial, K-Nearest Neighbor

ABSTRAK

Letak Indonesia yang berada di khatulistiwa dapat menyebabkan rentan terhadap fenomena cuaca ekstrem yang dapat memberikan ancaman. Cuaca ekstrem merupakan fenomena fisik atmosfer pada waktu tertentu dalam jangka waktu yang pendek dan juga bersifat ekstrem. Youtube dapat menjadi sebuah media untuk publik dalam menyampaikan opini kepada pemerintah dalam menghadapi cuaca ekstrem. Analisis sentimen merupakan metode yang digunakan untuk mengekstraksi, menganalisis, dan memproses data tekstual secara otomatis guna mengidentifikasi sentimen dalam suatu opini. Penelitian ini dilakukan bertujuan untuk menganalisis sentimen opini komentar publik di youtube mengenai cuaca ekstrem dengan menggunakan metode Naïve Bayes Multinomial dan K-Nearest Neighbor. Dataset diambil dari komentar di youtube mengenai cuaca ekstrem yang berjumlah 1030 komentar. Hasil penelitian menunjukkan bahwa metode Naïve Bayes Multinomial lebih unggul dengan akurasi 81%, presisi 76%, recall 63%, dan F1-Score 69% dibandingkan metode K-Nearest Neighbor yang hanya mencapai akurasi 51%, presisi 38%, recall 69%, dan F1-Score 69%. Berdasarkan hasil ini, metode Naïve Bayes Multinomial direkomendasikan untuk analisis sentimen opini publik tentang cuaca ekstrem.

Kata kunci: Cuaca ekstrem, Analisis sentimen, Naïve Bayes Multinomial, K-Nearest Neighbor

1. PENDAHULUAN

Letak Indonesia yang berada di khatulistiwa dapat menyebabkan rentan terhadap fenomena cuaca ekstrem seperti gelombang panas dan elnino [1]. Cuaca ekstrem merupakan fenomena fisik atmosfer pada waktu tertentu dalam jangka waktu yang pendek dan juga bersifat ekstrem[2]. Cuaca ekstrem dapat memberikan ancaman bencana di Indonesia. Sebagai contoh, pada tahun 2018, tercatat terjadi 2.564 bencana yang menjangkau seluruh wilayah Indonesia, tanpa ada satu pun kota atau kabupaten yang terbebas dari ancaman bencana[1].

Media sosial sekarang dapat digunakan sebagai perangkat komunikasi yang sangat populer untuk menyediakan berbagai informasi salah satunya adalah *Youtube*[3]. Menurut laman berita solopos.com pengguna *youtube* di Indonesia yang berusia 16-64 tahun sebanyak 170 juta pengguna[4]. Pengguna *youtube* juga dimudahkan dalam memberikan masukan lewat komentar, *like*, *dislike*, dan *share*[5]. *Youtube* dapat menjadi media yang dapat digunakan publik untuk menyampaikan opini dalam memberikan kebijakan yang dapat digunakan pemerintah untuk menghadapi cuaca ekstrem. Pemahaman yang akurat tentang opini publik terkait cuaca ekstrem dapat membantu pemerintah merumuskan kebijakan yang lebih responsif terhadap bencana. Analisis sentimen dapat menjadi alat yang efektif untuk mendukung tujuan ini. Namun, *platform youtube* seringkali membuat orang yang tidak bertanggung jawab menyebarkan komentar berbau negatif di dalam video - video di *youtube*.

Analisis sentimen merupakan metode yang digunakan untuk mengekstraksi, menganalisis, dan memproses data tekstual secara otomatis guna mengidentifikasi sentimen dalam suatu opini[6]. Analisis sentimen dapat kita gunakan untuk mengetahui apakah opini tersebut positif atau negatif. Metode ini membantu dalam memahami perasaan atau pandangan yang diekspresikan dalam teks. *Text Mining* merupakan proses awal pengolahan teks yang melibatkan beberapa langkah *preprocessing*, seperti *case folding*, *tokenizing*, *stopwords removal*, dan *stemming*[7]. Tahapan ini digunakan untuk mempersiapkan teks agar lebih mudah dianalisis secara mendalam.

Berdasarkan penelitian terdahulu, ada beberapa metode yang dapat digunakan dalam proses analisis sentimen yaitu metode *Naïve Bayes*, *K-Nearest Neighbor* dan *Support Vector Machine*[4][8][9], penelitian sebelumnya yang membahas analisis sentimen komentar di *Youtube* menggunakan metode *Naïve Bayes* menghasilkan akurasi tertinggi sebesar 87% [4]. Pada penelitian[8], analisis sentimen terkait kebijakan sistem zonasi sekolah berdasarkan platform *Youtube* menggunakan metode *K-Nearest Neighbor* dengan *Levenshtein Distance*. Penelitian ini menghasilkan akurasi 65,625% dengan *Classifier* = 3. Algoritma *Support Vector Machine* juga digunakan dalam penelitian analisis sentimen data Twitter pada tingkat kepercayaan masyarakat terhadap polri. Penelitian ini menghasilkan akurasi 94,79%[9]. Penelitian ini akan menggunakan metode *Naïve Bayes Multinomial* dan *K-Nearest Neighbor* dengan teknik *Random Oversampling* untuk menghasilkan analisis sentimen opini komentar *youtube* yang cenderung positif atau negatif.

Berdasarkan masalah yang sudah diuraikan, tujuan peneliti mengambil judul "Analisis Opini Publik Tentang Cuaca Ekstrem Menggunakan *Naïve Bayes* dan *K-Nearest Neighbor*" adalah untuk membandingkan hasil evaluasi dari dua algoritma klasifikasi dalam memahami pendapat masyarakat tentang cuaca ekstrem, yang diperoleh dari komentar pada video-video di kanal berita dengan kata kunci "cuaca ekstrem" di YouTube. Dengan menggunakan metode *Naïve Bayes Multinomial* dan *K-Nearest Neighbor* dengan *Euclidean Distance*, penelitian ini bertujuan mengkategorikan opini publik menjadi sentimen positif atau negatif.

2. METODE PENELITIAN

Metode yang dipakai untuk melakukan analisis sentimen adalah menggunakan model *Naïve Bayes Multinomial* dan *K-Nearest Neighbor* dengan *Euclidean Distance*. Alur dalam melakukan penelitian ini ditunjukkan pada Gambar 1 yaitu dimulai dengan pengumpulan data, *labeling* data, *preprocessing* teks, visualisasi data, data *preparation*, *training* data, evaluasi dan *testing*.



Gambar 1. Metode Penelitian

Pada gambar 1. Menunjukkan alur metode penelitian yang dirancang secara sistematis untuk mencapai tujuan penelitian. Metode ini bertujuan untuk memastikan bahwasan penelitian dilakukan secara menyeluruh.

1.1. Pengumpulan Data

Pengumpulan data dilakukan dengan mengumpulkan komentar dari platform video Youtube dengan toolkit bantuan yaitu <https://netlytic.org/> dengan kata kunci youtube “Cuaca Ekstrem” dengan banyak total data minimal 1030 komentar.

1.2. Labeling Data

Semua data yang telah dikumpulkan akan diberi label sentimen, yang dimana komentar orang tersebut merupakan komentar positif, negatif, ataupun netral. Peneliti memberi label sentimen pada setiap komentar berdasarkan kuisioner. Jumlah sentimen netral sebanyak 497 data, positif sebanyak 167 data dan negatif sebanyak 366 data seperti yang ditunjukkan pada Tabel 1.

Tabel 1. Labeling data

Sentimen	Jumlah
Netral	497
Positif	167
Negatif	366

1.3. Preprocessing Teks

Semua data yang telah dikumpulkan akan dilakukan preprocessing teks yang terdiri dari beberapa tahapan yaitu:

a. Case Folding

Yaitu mengubah semua huruf dalam komentar menjadi huruf kecil. Tabel 2 menunjukkan hasil sebelum dan sesudah dilakukannya case folding. Kalimat sebelum case folding merupakan kalimat asli dari dataset, sedangkan sesudah case folding merupakan kalimat baru yang dihasilkan.

Tabel 2. Case Folding

Sebelum	Sesudah
Yg jelas itu ya *KURSI PANAS*	yg jelas itu ya *kursi panas*
Segeeer klau ada panas	segeeer klau ada panas
Jakarta hari ini sudah normal	jakarta hari ini sudah normal

b. Normalisasi Data

Mengubah kalimat yang tidak baku atau yang disingkat menjadi kalimat yang jelas dan baku. Tabel 3 menunjukkan normalisasi dari kata-kata yang disingkat dan tidak baku menjadi sebuah kata baku yang dapat digunakan untuk tahapan selanjutnya.

Tabel 3. Normalisasi Data

Sebelum	Sesudah
lg	lagi
klau	kalau
sgl	segala

c. Stopword

Tahapan membuang kata-kata yang tidak memiliki signifikansi terhadap kalimat. Kata-kata yang sering muncul adalah kata hubung (konjungsi) dan preposisi. Tabel 4 merupakan contoh pembuangan *stopword*.

Tabel 4. Stopword

Sebelum	Sesudah
Yg jelas itu ya *KURSI PANAS*	jelas kursi panas
Segeeer klau ada panas	seger kalau ada panas
Jakarta hari ini sudah normal	jakarta hari ini normal

- d. Tokenisasi
Tahapan ini merupakan tahapan memecah kalimat menjadi kata-kata. Tabel 5 merupakan contoh hasil tokenisasi agar dapat dianalisis dengan mudah.

Tabel 5. Tokenisasi

Sebelum	Sesudah
Yg jelas itu ya *KURSI PANAS*	Jelas, kursi, panas
Segeeer klaw ada panas	Seger, kalau, ada, panas
Jakarta hari ini sudah normal	Jakarta, hari, ini, normal

- e. Stemming Data
Stemming data merupakan pencarian kata dasar dari setiap kata yang ditulis dalam kalimat.
- f. Membuang data yang memiliki sentimen netral
Data sentimen netral seringkali dibuang dari dataset. Jika data sentimen netral tidak dibuang, maka dapat dianalisis lebih lanjut dengan menggunakan ahli bahasa untuk memberikan label yang akurat. Dengan demikian, sentimen netral dapat diklasifikasikan sebagai negatif atau positif berdasarkan konteks dan nuansa bahasa yang digunakan. Karena dengan meniadakan sentiment netral dapat mengurangi noise, peningkatan akurasi, dan efisiensi komputasi. Dengan membuang data sentimen netral, model dapat lebih fokus pada pola dan struktur data yang relevan. Ini meningkatkan akurasi dan efisiensi komputasi.
- g. Mengubah menjadi numerik
Mengubah sentimen menjadi bentuk numerik 0 (negatif) dan 1 (positif) memungkinkan model pembelajaran mesin untuk memproses data dengan lebih efektif. Pendekatan ini memungkinkan peneliti untuk menganalisis sentimen dengan lebih akurat dan efisien, sehingga meningkatkan kemampuan model dalam mengklasifikasikan sentimen.

1.4. Visualisasi Data

Menampilkan visualisasi sentimen pada data sudah di *cleaning* dengan menggunakan *WordCloud* dan *plot*, serta menganalisis distribusi sentimen positif dan negatif secara grafis. Tahapan ini terdiri dari beberapa tahapan yaitu:

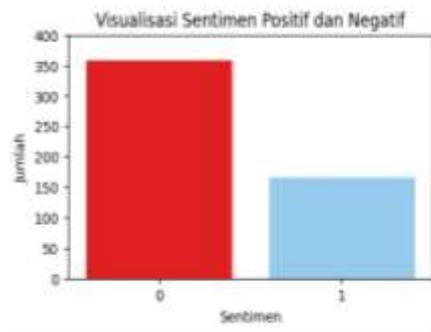
- Membuat *Data Frame* untuk memisahkan sentimen positif dan negatif.
- Membuat visualisasi sentimen positif dan negatif menggunakan *WordCloud* untuk menggambarkan kata-kata yang muncul pada sentimen positif dan negatif seperti yang ditunjukkan pada Gambar 2.



Gambar 2. Visualisasi Sentimen Negatif dan Positif

Gambar 2 menunjukkan visualisasi *WordCloud* dari opini publik yang berkategori negatif dan positif mengenai cuaca ekstrem yang terjadi di Indonesia. Dari gambar tersebut dapat dilihat kata “panas” mendominasi pada kedua visualisasi yang dapat menjadi isu utama yang terkait fenomena cuaca ekstrem yang dihadapi.

- Membuat visualisasi gabungan sentimen positif dan negatif menggunakan *plot* untuk memahami distribusi data.



Gambar 3. Plot perbandingan jumlah data negatif dan positif

Pada Gambar 3, bar berwarna merah menunjukkan jumlah data dengan sentimen negatif sebanyak 358, sementara bar berwarna biru menunjukkan jumlah data dengan sentimen positif sebanyak 165. Terlihat jelas adanya ketidakseimbangan antara jumlah data sentimen negatif dan positif.

1.5. Data Preparation

Pada tahap ini melakukan pembagian data menjadi dua bagian yaitu data untuk dilatih dan diuji. Pembagian dilakukan menggunakan fungsi `train_test_split` dengan rasio 80:20, di mana 80% data digunakan untuk pelatihan dan 20% untuk pengujian, dengan pengacakan yang dikontrol oleh “`random_state=42`”.

Selanjutnya, menggunakan *CountVectorizer* untuk mengubah data teks menjadi representasi numerik. Untuk mengatasi ketidakseimbangan data untuk pelatihan, kami memilih metode *Random Oversampling* untuk digunakan untuk menyeimbangkan data yang akan digunakan untuk dilakukan pelatihan dan pengujian dimana metode ini akan menambah sampel dari kelas minoritas dengan cara menduplikasi data dari kelas minoritas secara acak hingga jumlah sampel di setiap kelas menjadi seimbang. Proses *oversampling* ini menghasilkan data pelatihan yang telah di-resample yaitu $x_{train_resampled}$ dan $y_{train_resampled}$.

1.6. Training Data

a. Naïve Bayes Multinomial

Naïve Bayes merupakan metode klasifikasi yang sering digunakan dalam penelitian analisis sentimen karena mudah dan sederhana. Metode ini menghitung probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang tersedia[10]. *Naïve Bayes* memiliki kelebihan yaitu sangat mudah untuk diimplementasikan serta juga dapat memberikan hasil akurasi yang bagus untuk banyak kasus[11]. Ada dua tahapan klasifikasi dengan metode *Naïve Bayes*[12]. Tahapan pertama adalah tahapan dalam melatih data dan tahapan kedua adalah tahapan dalam menguji data. *Naïve Bayes* secara umum dapat dinotasikan dengan persamaan sebagai berikut:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (1)$$

Keterangan:

$P(A|B)$ = Probabilitas hipotesis A berdasarkan kondisi B.

$P(B|A)$ = Probabilitas hipotesis B berdasarkan kondisi A.

$P(A)$ = Probabilitas hipotesis A.

$P(B)$ = Probabilitas hipotesis B.

b. K-Nearest Neighbor

K-Nearest Neighbor adalah sebuah algoritma klasifikasi yang Menggunakan perhitungan *supervised learning*[13]. Algoritma ini digunakan untuk mengklasifikasi objek berdasarkan atribut dan data training. Nilai prediksi pada algoritma *K-Nearest Neighbor* menggunakan ketetanggaan[14]. Algoritma ini memiliki keunggulan pada pelatihnannya yang sederhana, efisien, dan mudah dipahami pada data yang berukuran besar[15]. Perhitungan jarak yang digunakan pada algoritma ini adalah *Euclidean Distance*. KNN dengan *Euclidean Distance* dapat dinotasikan dengan persamaan sebagai berikut:

$$D_{(a,b)} = \sqrt{\sum_{k=3}^d (a_k - b_k)^2} \quad (2)$$

Keterangan:

$D(a,b)$ = Jarak *Euclidean* antara vektor a dan vektor b.

d = Banyaknya variabel bebas.

k = Variabel data.

a_k = Nilai objek data *training* a pada variabel ke- k .

b_k = Nilai objek data *training* b pada variabel ke- k .

Langkah ini bertujuan untuk melatih data menggunakan metode *Naïve Bayes Multinomial* dan *K-Nearest Neighbor* dengan *euclidean distance* ($k = 3$) untuk menentukan prediksi akhir.

1.7. Evaluasi

Setelah data dilatih langkah selanjutnya adalah melakukan pengujian menggunakan model yang telah dilatih sebelumnya kemudian dilakukan evaluasi pada model yang sudah dibuat dengan menggunakan metode *confusion matrix*. Metode *confusion matrix* digunakan untuk mengukur akurasi, presisi, *recall*, dan *F1-Score*. Rumus-rumus pengukuran tersebut dapat dilihat pada persamaan-persamaan dibawah ini:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (6)$$

Keterangan:

TP = *True Positive*

TN = *True Negative*

FP = *False Positive*

FN = *False Negative*

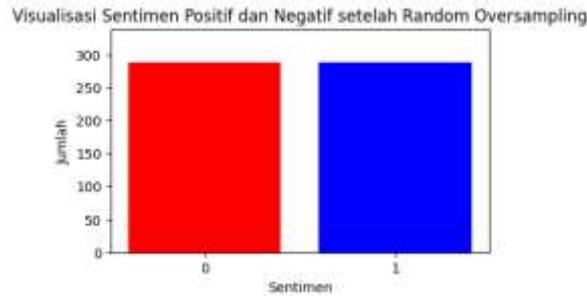
1.8. Evaluasi

Setelah melalui proses pelatihan model, tahap berikutnya adalah testing. Pada tahap ini, untuk menguji kalimat-kalimat baru untuk memprediksi apakah sentimen yang terkandung dalam kalimat tersebut bersifat positif atau negatif terhadap fenomena cuaca ekstrem di Indonesia. Testing ini penting untuk mengevaluasi akurasi masing-masing model dalam menganalisis opini publik terhadap cuaca ekstrem.

3. Hasil dan Pembahasan

3.1. Dataset

Dataset didapatkan setelah mengumpulkan komentar - komentar pada *platform Youtube* dengan kata kunci “Cuaca Ekstrem” sebanyak 1030 komentar. Setelah itu data diberi label dan dilakukan *preprocessing* teks. Dalam tahap *preprocessing teks*, sentimen diubah menjadi bentuk numerik yakni 0 (Negatif) dan 1 (Positif). Selanjutnya setelah selesai memproses teks, terdapat ketidakseimbangan pada data dengan sentimen negatif dan positif seperti yang ditampilkan pada Gambar 3. Untuk menyeimbangkan data tersebut peneliti menggunakan metode *Random Oversampling*.

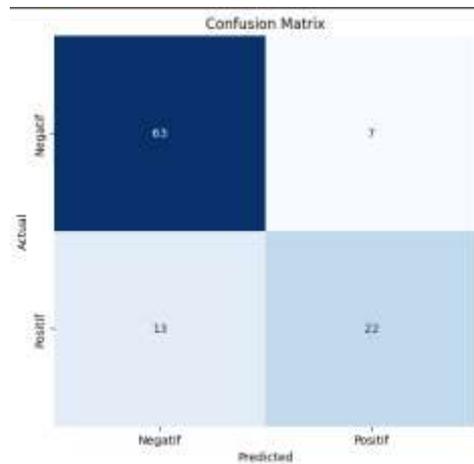


Gambar 4. Sentimen Positif dan Negatif Menggunakan *Random Oversampling*

Gambar 4 merupakan hasil perbandingan sentimen positif dan negatif setelah menggunakan metode *Random Oversampling*, di mana jumlah masing-masing sentimennya sama. Dengan menggunakan *Random Oversampling* dapat membuat jumlah data pada kedua kelasnya seimbang serta dapat menghasilkan akurasi yang lebih tinggi.

3.2. Evaluasi Model

Setelah data dipersiapkan, kemudian peneliti membagi data menjadi dua bagian yakni data latih sebesar 80% dan data uji sebesar 20%. Setelah membagi data, peneliti melakukan *training* model dengan metode *Naïve Bayes Multinomial* dan *K-Nearest Neighbor* dengan *euclidean distance* ($k = 3$). Hasil *confusion matrix* metode *Naïve Bayes Multinomial* dan *K-Nearest Neighbor* ditunjukkan pada Gambar 5 dan 6.



Gambar 5. *Confusion Matrix* Metode *Naïve Bayes Multinomial*

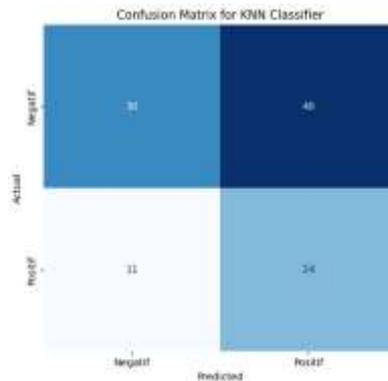
Hasil *confusion matrix* pada Gambar 5 menunjukkan *matrix* sebagai berikut:

TP (<i>True Positive</i>)	= 22
TN (<i>True Negative</i>)	= 63
FP (<i>False Positive</i>)	= 7
FN (<i>False Negative</i>)	= 13

Berdasarkan hasil *matrix* diatas dapat dihitung hasil pengukuran akurasi, presisi, *recall*, dan *F1-Score* pada model *Naïve Bayes Multinomial* yang dijelaskan pada Tabel 6.

Tabel 6. *Confusion Matrix* Metode *Naïve Bayes Multinomial*

Metode	Akurasi	Presisi	Recall	F1-Score
Naïve Bayes Multinomial	81%	76%	63%	69%



Gambar 5. *Confusion Matrix Metode K-Nearest Neighbor*

Hasil *confusion matrix* pada Gambar 6 menunjukkan *matrix* sebagai berikut:

TP (*True Positive*) = 24
 TN (*True Negative*) = 30
 FP (*False Positive*) = 40
 FN (*False Negative*) = 11

Berdasarkan hasil *matrix* diatas dapat dihitung hasil pengukuran akurasi, presisi, *recall*, dan *F1-Score* pada model *K-Nearest Neighbor* yang dijelaskan pada Tabel 7.

Tabel 7. *Confusion Matrix Metode K-Nearest Neighbor*

Metode	Akurasi	Presisi	Recall	F1-Score
<i>K-Nearest Neighbor</i>	51%	38%	69%	48%

Pada Tabel 6 dan 7 menunjukkan hasil evaluasi metode *Naïve Bayes Multinomial* yakni akurasi sebesar 81%, presisi sebesar 76%, *recall* sebesar 63%, dan *F1-Score* sebesar 69%, sedangkan metode *K-Nearest Neighbor* menghasilkan akurasi sebesar 51%, presisi sebesar 38%, *recall* sebesar 69%, dan *F1-Score* sebesar 48%. Dari hasil evaluasi tersebut menyatakan bahwa metode *Naïve Bayes Multinomial* lebih baik dibandingkan metode *K-Nearest Neighbor* dengan *euclidean distance* ($k = 3$).

3.3. Testing

Setelah melakukan evaluasi pada model yang sudah dilatih, selanjutnya peneliti akan melakukan pengujian dengan memasukkan kalimat baru sebanyak 5 kali untuk memprediksi sentimen tersebut positif atau negatif. Dalam *testing*, peneliti membandingkan hasil prediksi dari kedua metode yang digunakan yakni *Naïve Bayes Multinomial* dan *K-Nearest Neighbor* dalam memprediksi sebuah kalimat. Hasil pengujian pada 5 kalimat baru menunjukkan bahwa hasil sentimen yang dihasilkan oleh metode *Naïve Bayes Multinomial* lebih akurat dibandingkan metode *K-Nearest Neighbor*. Hasil pengujian dapat dilihat pada Tabel 8.

Tabel 8. *Confusion Matrix Metode K-Nearest Neighbor*

Pengujian	Kalimat	Naïve Baye	K-Nearest Neighbor
1	saya gak tahan sama panas di siang hari disini dingin	Negatif	Negatif
2	sampai mau mati bagus pohonnya	Negatif	Positif
3	ditebangin aja trus	Negatif	Positif

4	di sini dinginnya 100 derajat	Negatif	Positif
5	ayo main ke palembang dingin loh disini	Positif	Positif

4. KESIMPULAN

Penelitian ini menunjukkan bahwa metode Naïve Bayes Multinomial lebih unggul dibandingkan K-Nearest Neighbor (KNN) dengan Euclidean distance ($k=3$) dalam memprediksi sentimen pada komentar mengenai cuaca ekstrem di YouTube. Setelah mengumpulkan dan melabeli 1030 komentar, serta menyeimbangkan data dengan metode Random Oversampling, hasil evaluasi menunjukkan bahwa Naïve Bayes Multinomial menghasilkan akurasi sebesar 81%, sementara KNN hanya mencapai 51%. Selain itu, Naïve Bayes Multinomial memiliki F1-Score sebesar 69%, dibandingkan dengan KNN yang hanya mencapai 48%. Pengujian tambahan dengan 5 kalimat baru juga menunjukkan performa Naïve Bayes Multinomial yang lebih konsisten.

Oleh karena itu, metode Naïve Bayes Multinomial direkomendasikan sebagai pendekatan yang lebih efektif untuk analisis sentimen. Hasil penelitian ini dapat membantu pemerintah atau pengelola platform media sosial dalam memahami opini publik terkait cuaca ekstrem secara lebih akurat, yang bermanfaat untuk perencanaan kebijakan atau mitigasi bencana. Namun, penelitian ini masih memiliki keterbatasan pada jumlah data yang relatif kecil dan penggunaan teknik balancing tertentu. Penelitian di masa depan disarankan untuk menggunakan data yang lebih besar dan mengeksplorasi teknik balancing alternatif untuk hasil yang lebih robust.

Daftar Rujukan

- [1] R. W. Prabowo, "Tipologi Kerusakan Bangunan Akibat Cuaca Ekstrem Studi Kasus Cuaca Ekstrem Kabupaten Bantul 2019," *Sustain. Plan. Cult. J. Perenc. Wil. dan Kota*, vol. 2, no. 2, pp. 31–36, 2020, doi: 10.32795/space.v2i2.1123.
- [2] T. Hariyanto, F. D. Ramadhan, and C. B. Pribadi, "Studi Estimasi Curah Hujan Pada Kondisi Cuaca Ekstrem Tahun 2017 Melalui Citra Modis Level 1B Di Provinsi Jawa Timur," *Geoid*, vol. 13, no. 2, p. 187, 2018, doi: 10.12962/j24423998.v13i2.3832.
- [3] I. Kurniawan and A. Susanto, "Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019," *Eksplora Inform.*, vol. 9, no. 1, pp. 1–10, 2019, doi: 10.30864/eksplora.v9i1.237.
- [4] H. Al Rasyid Harpizon, R. Kurniawan, I. Iskandar, R. Salambue, E. Budianita, and F. Syafria, "Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 1, pp. 131–140, 2022.
- [5] R. Kurniawan, F. Lestari, A. S. Batubara, M. Z. A. Nazri, K. Rajab, and R. Munir, "Indonesian Lexicon-Based Sentiment Analysis of Online Religious Lectures Review," *2021 Int. Congr. Adv. Technol. Eng. ICOTEN 2021*, pp. 1–5, 2021, doi: 10.1109/ICOTEN52080.2021.9493530.
- [6] F. V. Sari and A. Wibowo, "Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi," *J. SIMETRIS*, vol. 10, no. 2, pp. 681–686, 2019.
- [7] T. Ridwansyah, "Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 5, pp. 178–185, 2022, doi: 10.30865/klik.v2i5.362.
- [8] N. Anggraini and M. J. Tursina, "Sentiment Analysis of School Zoning System on Youtube Social Media Using the K-Nearest Neighbor with Levenshtein Distance Algorithm," *2019 7th Int. Conf. Cyber IT Serv. Manag. CITSM 2019*, no. May, pp. 1–4, 2019, doi: 10.1109/CITSM47753.2019.8965407.
- [9] R. Marsuciati and E. Taufiq Luthfi, "Analisis Svm (Support Vector Mechine Pada Tingkat Kepercayaan

Masyarakat Terhadap Polri,” *Julyxxxx*, vol. x, No.x, no. x, pp. 1–5, 2012.

- [10] E. Nurlelah and R. Wajhillah, “Penerapan Naive Bayes Untuk Diagnosa Penyakit Diare Usia,” no. March, pp. 55–62, 2016.
- [11] D. Singgih Pamungkas, N. Ageng Setiyanto, and E. Dolphina, “Analisis Sentiment Pada Sosial Media Twitter Menggunakan Naive Bayes Classifier Terhadap Kata Kunci ‘Kurikulum 2013,’” *Techno.COM*, vol. 14, no. 4, pp. 299–314, 2015.
- [12] S. Lestari and S. Saepudin, “Analisis Sentimen Vaksin Sinovac Pada Twitter Menggunakan Algoritma Naive Bayes,” *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.)*, vol. 1, no. 1, pp. 165–170, 2021, [Online]. Available: <https://vaksin.kemkes.go.id/>
- [13] S. H. Ramadhani and M. I. Wahyudin, “Analisis Sentimen Terhadap Vaksinasi Astra Zeneca pada Twitter Menggunakan Metode Naïve Bayes dan K-NN,” *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 6, no. 4, pp. 526–534, 2022, doi: 10.35870/jtik.v6i4.530.
- [14] A. Salam, J. Zeniarja, and R. S. U. Khasanah, “Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekpress Indonesia),” *Pros. SINTAK*, pp. 480–486, 2018.
- [15] D. Cahyanti, A. Rahmayani, and S. A. Husniar, “Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara,” *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 39–43, 2020, doi: 10.33096/ijodas.v1i2.13.