Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi
https://ojs.stmik-banjarbaru.ac.id/index.php/jutisi/index
Jl. Ahmad Yani, K.M. 33,5 - Kampus STMIK Banjarbaru Loktabat – Banjarbaru
e-mail: puslit.stmikbjb@gmail.com
e-ISSN: 2685-0893

Implementasi Algoritma *Naïve Bayes* dan *Decision Tree* pada Sentimen Dampak Bencana Hidrometeorologi

DOI: http://dx.doi.org/10.35889/jutisi.v15i4.3858

Arry Wahyu Setiawan[1*], Suprianto [2], Yunianita Rahmawati[3], Suhendro Busono[4]
Informatika, Universitas Muhammadiyah Sidoarjo, Sidoarjo, Indonesia
*e-mail *Corresponding Author:* arrywahyu0@gmail.com

Abstract

Hydrometeorological disasters often trigger public responses on social media, so it's requiring analysis public opinion toward such disaster. This study compares the performance of Decision Tree and Naïve Bayes Algorithm to analyze performance on analyzing sentiments of respond related to hydrometeorological disaster. The dataset that used contains of 11890 comments. Comments labeled into Positive, Negative, and neutral then processed by preprocessing, TF-IDF weighting, and classification with 5-fold cross validation. Validation results Naïve Bayes had mean accuracy by 0.79 which higher than Decision Tree, which scored 0,65. Naïve Bayes also got 0.79 score on accuracy, precision, recall, and F1-Score compares to Decision tree that scored scored 0.63, 0.70, 0.63, and 0.64 respectively. Because the gap of the mean accuracy and test accuracy wasn't far off, concludes that Naïve Bayes have better generalization than Decision Tree. This result shows that Naïve Bayes is better than Decision Tree on classification of Hydrometeorological Disaster respond.

Key Words: *Sentiment Analysis; Hydrometeorology Disaster; Naïve Bayes; Decision tree; TF-IDF*

Abstrak

Bencana hidrometeorologi memicu berbagai respons masyarakat di media sosial. maka diperlukan analisis untuk memahami opini publik terhadap bencana terkait. Tujuan penelitian ini adalah membandingkan kinerja algoritma *Naïve Bayes* dan *Decision Tree* dalam melakukan analisis sentimen terkait bencana Hidrometeorologi. Data digunakan berasal dari TikTok, Instagram Facebook yang terdiri dari 11.890 komentar dengan label kelas positif, negatif, dan netral yang kemudian diproses melalui tahapan pra-pemrosesan, pembobotan TF-IDF, dan klasifikasi dan validasi *5-fold cross validation.* Hasil validasi menunjukkan *Naïve Bayes* memperoleh *mean accuracy* 0,79 lebih tinggi daripada *Decision Tree* yang mendapat skor 0,65. Pada data uji, Naïve Bayes memperoleh *accuracy, precision, recall,* dan *F1-Score* yang seragam 0,79 lebih unggul dari *Decision Tree* dengan *accuracy* 0,63, *precision* 0,70, recall 0,63, dan *F1-score* 0,64. Dikarenakan jarak *mean acmecuracy* dan *test accuracy* berdekatan disimpulkan *Naïve Bayes* mempunyai kemampuan generalisasi lebih baik daripada *Decision Tree*. Hasil ini menunjukkan *Naïve Bayes* unggul dalam mengklasifikasi sentimen terkait bencana Hidrometeorologi

Kata Kunci: *Analisis Sentimen; Bencana Hidrometeorologi; Naibe Bayes; Decision Tree; TF-IDF*

1. Pendahuluan

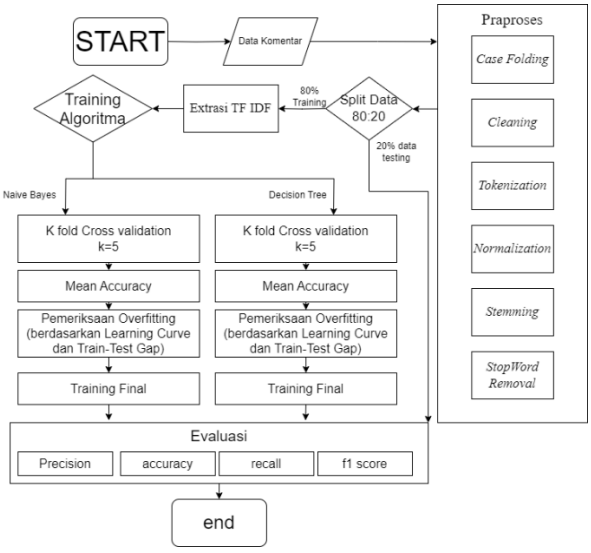
Indonesia merupakan negara yang memiliki penetrasi internet tertinggi di Asia Tenggara. Berdasarkan survey Asosiasi Penyelenggara Jasa Internet Indonesia periode 2023 hingga 2024, pengguna internet di indonesia terdapat221 juta jiwa dari 278 juta penduduk [1]. Tingginya penggunaan internet berdampak pada terbentuknya pola menyampaian opini dan interaksi masyarakat di berbagai platform media sosial. Salah satu peristiwa yang memicu banyaknya respon public di sosial media per desember 2025 adalah bencana hidrometeorologi. Dengan demikian topik ini baik untuk dikaji lebih lanjut.

Media sosial tidak hanya menjadi sarana komunikasi, Media sosial dapat meberikan informasi respons masyarakat terhadap suatu peristiwa. Berberapa penelitian terdahulu menunjukan bahwa komentar media sosial terkait bencana memuat informasi kondisi terdampak dan respons masyarakat terhadap dampak bencana [2][3]. Informasi tersebut berguna untuk sebagai bahan evaluasi bagi pemerintah dan pemangku kepentingan untuk memahami persepsi masyarakat dalam pengambilan keputusan penanganan bencana. Namun tingginya volume komentar pada sosial media menyebabkan secara manual melelahkan dan tidak efisien sehingga diperlukan metode analisis sentimen berbasis algoritma klasifikasi untuk mengelola respon dengan efisienBerdasar penelitian-penelitian terdahulu, prosedur yang dikembangkan dalam klasifikasi domain teks khususnya konteks bencana secara garis besar menggunakan tahapan pengumpulan data, prapemrosesn teks, ekstrasi fitur dan klasifikasi. Berberapa penelitian terada pada penelitian Anugrah et all dan Prasetyo et all [4][5] yang menggunakan klasifikasi naïve bayes terhadap respon korban bencana. Sedangkan pada penlitian Rozi *et al* tahapan dikembangkan menggunakan K-means untuk mengelompokan data yang selanjutnya dikelompokan menggunakan Naïve Bayes [6]. Pada penelitian Das *et al* menunjukkan pendekatan TF-IDF memperkuat respentatif setiap kata pada klasifikasi domain teks [7]. Pendekatan tersebut diemplementasi oleh Ioendri *et al* pada klasifikasi menggunakan Naïve Bayes serta diadopsi oleh Yuniar,Kismiantini, Rizkina, Kustyana *et al,* Baehaqi serta Gerliandeva et al pada penlitian analisis sentimen menggunakan Klasifikasi Naïve bayes [8][9] [5][10]. Pada konteks Decision Tree, algoritma tersebut digunakan oleh Maharani dan Fathoni [11], Pawar et al. [12], Akbar et al. [13], serta Trianto et al. [14] pada domain teks menghasilkan hasil akurasi yang cukup baik. Secara keseluruhan penelitian, menunjukan algoritma Naïve Bayes dan Decission Tree merupakan algoritma yang cukup efektif dalam klasifikasi teks dengan perbedaan pada pendekatan model klasifikasinya.

Berdasar Kajian penelitan terdahulu, terdapat keterbatasan pada prosedut analisis sentimen yang dikembangkan. Pada penelitian analisis sentimen yang mengambil konteks kebencanaan penelitian sebelum nya hanya terbatas di satu platform sosial media saja. Akibatnya, pemeriksaan respon kurang meluas ke berbagai kalangan platform karena setiap sosial media memiliki karakteristik pengguna dan gaya bahasa yang berbeda-beda selain itu pada penelitan terdahulu baik *Naïve Bayes* amupun *Decision tree* tidak membandingkan performa algoritma pada konteks multiplatform. Berdasarkan gap tersebut, kebaruan (*state of art)* penelitian ini mengembangkan metode klasifikasi teks konteks bencana dengan memperluuas sumber data lebih dari satu platform media sosial dengan membandingkan performa algoritma *Naïve Bayes* dan *Decision Tree* dalam klasifikasi sehingga dapat dihasilkan prosedur klasifikasi teks bencana yang lebih baik dari penelitian-penelitian sebelumnya.

2. Metodologi

Metodologi penelitian yang digunakan pada penelitian ini mengikuti alur proses sebagaimana ditunjukkan pada diagram alir Gambar 1:



Gambar 1 Diagram Alir proses penelitian

2.1 Data Komentar

Data Komentar dikumpulkan melalui metode web scraping menggunakan python dengan library *playwright dan beatifulsoup*. Proses scraping mengambil komentar dari periode Oktober 2025 sampai dengan desember 2025. Komentar diambil dari postingan ber kata kunci bencana yang termasuk hidrometeorologi (contoh : banjir, longsor) dalam 3 platform sosial media, yaitu tiktok, instagram dan facebook. Lalu data akan dilabelkan dengan kriteria berikut:

1) Positif: mengandung dukungan, doa, harapan, atau ucapan terima kasih
2) Negatif: mengandung kritik, keluhan, kekhawatiran, atau kemarahan
3) Netral: berisi informasi faktual tanpa ekspresi emosional dan spam

2.2 Prapemrosesan data

Prapemrosesan adalah proses dalam memperbaiki data agar ai lebih mudah dalam memahami suatu sentiment. Prapemrosesan ini terdiri dari

1) Case Folding: Perubahan seluruh huruf menjadi *lowercase*.
2) Cleaning: proses menghapus karakter yang tidak diperlukan
3) Tokenisasi: proses memecah kalimat menjadi token
4) Normalisasi: mengubah kata tidak baku atau bahasa gaul menjadi bentuk baku
5) Stemming: proses mengubah kata menjadi kata dasar.
6) Stopword removal: proses menghapus kata yang memiliki sedikit pengaruh terhadap analisis

2.3 Pembagian Data 80:20

Pemisahan dataset menjadi 2 porsi, data uji dan latih untuk *k fold cross*. Pembagian data menggunakan perbandingan 80:20, dimana 80% data duguakan untuk pelatihan model dan 20% sisanya untuk meguji kinerja daripada model. Pembagian menggunakan metode *stratified sampling.* Data latih digunakan untuk melakukan evaluasi pada model terhadap data yang belum pernah digunakan dalam pelatihan.

2.4 Pembobotan TF-IDF

Data teks akan diubah menjadi bentuk numerik sehingga model dapat melihat dan membaca tersebut. Metode yang dipakai peneliti adalah TF-IDF. TF-IDF akan memberikan berat pada setiap kata dengan menyesuaikan frekuensi. Hasil perhitungan TF-IDf adalah matrix dari vector numerik yang akan mempresentasikan kepentingan kata dalam suatu dokumen yang ditentukan terhadap seluruh corpus. Vector ini lah yang akan dilakukan training pada algoritma naïve bayes. TF-IDF memiliki rumus sebagai berikut

$$TF(t, d)= \frac{ft,d}{\sum k\ fk,d}\ IDF(t)= \log \frac{N}{nt} \quad (1)$$

TF merupakan nilai frekuensi suatu kata dalam korpus dokumen, sedangkan idf dugunakan untuk membobotkan berdasar tingkat pentingnya data pada seluruh dokumen dalam korpus. Parameter yang digunakan dama TF-IDF adalah menggunakan unigram, bigram, dan trigram sebagai fitur. Jumlah fitur yang digunakan adalah sebanyak 20.000 kata/frasa.

2.5 Implementasi Klasifikasi Model Algoritma *Naïve Bayes* dan *Decision Tree*

Setelah data diekstraksi menjadi data TF-IDF, dilakukan metode klasifikasi menggunakan algortima *Naive Bayes* dan *Decision Tree* untuk menghitung kemungkinan ada nya suatu data ke dalam kelas tertentu.

1) Naïve Bayes

Algoritma *Naïve Bayes* mnggunakan proses dimana mengklasifikasikan data berdasar nilai probabilistik menggunakan teorema Bayes. Rumus *Naïve Bayes* adalah sebagai berikut.

$$P(Ci \mid x)= \frac{P(x \mid Ci)P(Ci)}{P(x)} \quad (2)$$

Pada persamaan (2), P(Ci | x) merupakan probablitias posterior x ke dalam kelas Ci, P(x | Ci) adalah probabilitas kemunculan x pada kelas Ci. P(Ci) menyatakan probablitias prior kelas Ci, sedangkan P(x) berfungsi sebagai konstanta normalisasi

2) Decision Tree

Decision Tree adalah metode klasifikasi dan prediksi yang dibentuk dari struktur pohon. Metode ini bekerja dengan cara membagi dataset ke dalam beberapa subset berdasar atribut tertentu melalui proses pemilihan fitur terbaik sebagai node simpul. Struktur Decision Tree terdiri dari node akar yang atribut pertama yang digunakan untuk melakukan pemisahan data. Selanjutnya node simpul untuk membagi data kepada node daun. Hasil akhir berupa kelas atau keputusan Prinsip utama *Decision Tree* adalah melakukan pemilihan atribut terbaik untuk membagi data sehingga menghasilkan kelompok yang paling homogen. Untuk menentukan atribut terbaik, digunakan ukuran seperti *Entropy* digunakan untuk mengukur tingkat ketidakpastian atau ketidakhomogenan suatu data dengan memanfaatkan proporsi kelas ke I (pi). Entropy dapat dirumuskan sebagai berikut

$$\text{Entropy (S)} = -\sum_{i=1}^{n} \text{pi} \log_2 \text{pi} \tag{3}$$

Information Gain digunakan untuk menentukan atribut terbaik dalam pemisahan data dari entropy, dapat dirumuskan sebagai berikut:

$$\textit{Gain (S, A)=Entropy(S)-}\sum_{\in \text{Values(A)}} \frac{|\text{Sv}|}{|\text{S}|} \text{Entropy (Sv)} \tag{4}$$

2.6 Validasi model dengan *5-fold Cross validation*

Validasi model dilakukan dengan metode *5-Fold Cross Validation* untuk memeroleh estimasi performa model dan mengungari bias akibat pembagian data. Pada metode ini dalah latih dibagi menjadi 5 bagian (*fold*) dengan proporsi yang sama menggunakan *Stratified K-Fold*. Pada setiap iterasi satu *fold* digunakan sebagai data dalam validasi. Sedangkan *fold* sisanya digunakan sebagai data latih. Proses tersebut diulang sebanyak 5 kali sehingga setiap *fold* mampu menjadi data training. Nilai k sama dengan 5 dipilih karena data testing memiliki besar 1/5 dari jumlah kesuluruhan data.

2.7 Akurasi rata-rata (*mean accuracy*)

Akurasi rata rata akan digunakan untuk mengukur performa model berdasar hasil validasi menggunakan *K-fold Cross Validation*. Nilai ini diperoleh dari rata-rata akurasi setiap fold. Sehingga didapatkan rumus:

$$\text{Mean Accuracy} = \frac{\sum_{i=1}^{k} \textit{Accuracy i}}{k} \tag{5}$$

Pada Persamaan (5), *accuracy* merupakan nilai akurasi pada fold ke-i, sedangkan k adalah jumlah fold yang digunakan pada proses validasi. Pada penelitian ini digunakan k=5.

2.8 Pemeriksaan overfitting dengan learning curve

Pemeriksaan overfitting dilakukan menggunakan *learning curve* untuk melakukan evaluasi kemampuan generalisasi pada model. *Learning Curve* dapat menampilkan hubungan antara akurasi data latih dan akurasi data validasi apabila terdapat kurva saling menjauhi, maka model dapat disebutkan overfitting.

Pada penelitian ini, *learning curve* berisi data menggunakan hasil *K-Fold Cross Validation* pada data latih. Selain itu, pemeriksaan overfitting juga membandingkan nilai akurasi latih dan akurasi tes. Semakin kecil selisih suatu data tersebut, maka semakin baik kemampuan model dalam mengenali data yang belum pernah dipelajari.

2.9 Evaluasi model

Evaluasi Model dilakukan menggunakan data uji yang sudah dipisahkan. Evaluasi model dilakukan untuk mengukur kemampuan model dalam mengklasifikasikan sentimen. Performa model diukut berdasarkan confusion matrix dengan kriteria sebagai *Precision, Accuracy*, *Recall,*

F1-Score. Accuracy mencerminkan banyak nya komentar yang dipilih dengan benar oleh model. Dapat dinyatakan di persamaan 6:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

*Precision*mengukur ketepatan model dalam memberikan suatu prediksi suatu kelas. Dinyatakan di persamaan 7

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

Recall menunjukan tingkat keberhasilan model dalam mengenali seluruh data yang diklasifikasikan terhadap seluruh data yang benar benar pada suatu kelas. Dapat dinyatakan di pernyataan 8

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

F1-Score digunakan dalam mengikut antara nilai precision dan recall, dapat dinyatakan di pernyataan 9

$$F1\ Score = 2 \times \frac{Precision(Recall)}{Precision+Recall} \quad (9)$$

3. Hasil dan Pembahasan

3.1 Deskripsi data terkumpul

Dataset telah diperoleh dari komentar netizen pada platform media sosial yang berkaitan dengan topik penelitian. Data dikumpulkan menggunakan teknik scraping. Setelah data terkumpul, data dilabeli menjadi 3 kelas, positif negatif dan netral. Pelabelan dilakukan berturut-turut netral, positif, negatif dengan 0, 1, dan 2. Data yang terkumpul dari platform sosial media terbatas dikarenakan penemuan pembatasan pengambilan pada setiap sosial media. data yang berhasil terkumpul terdapat sebanyak 11,890 komentar dari setiap aplikasi (tiktok, instagram, dan facebook). Dikarenakan pembatasan setiap platform berbeda maka didapatkan data komentar sebanyak 1638 dari facebook, 6120 dari tiktok, dan 4132 dari instagram. Setiap kelas positif, negatif, netral berturut turut mendapatkan jumlah 3230, 2884, dan 5776.

3.2 Hasil pra-pemrosesan data

Preprocessing data dilakukan dengan 6 fase, *Case Folding, cleaning*, Tokenisasi, Normalisasi, *Stemming*, Dan *Stopword* Removal. Sebagai contoh maka diambil satu data komentar sampel: “Dan pemerintah dengan entengnya bilang kondisi sudah membaik :(”

Tabel 1. Proses Perubahan data ketika preprocessing

Tahapan	Komentar
Komentar Asli	Dan pemerintah dengan entengnya bilang kondisi sudah membaik :(
Case Folding	dan pemerintah dengan entengnya bilang kondisi sudah membaik :(
Cleaning	dan pemerintah dengan entengnya bilang kondisi sudah membaik
Tokenisasi	['dan', 'pemerintah', 'dengan', 'entengnya', 'bilang', 'kondisi', 'sudah', 'membaik']
Normalisasi	['dan', 'pemerintah', 'dengan', 'entengnya', 'berkata', 'kondisi', 'sudah', 'membaik']
Stemming	['dan', 'pemerintah', 'dengan', 'enteng', 'bilang', 'kondisi', 'sudah', 'baik']
Stopword Removal	['pemerintah', 'enteng', 'bilang', 'kondisi', 'baik']

Data pada dataset sebelum melalui proses Preprocessing memiliki 111,955 kata. Kata yang sering muncul adalah Pemerintah. Rata-rata jumlah kata per komentar adalah 9 kata dengan komentar terpendek hanya memiliki 1 kata dan komentar terpanjang memiliki 628 kata. Setelah preprocessing, kata berkurang menjadi 90,987 kata.

3.3 *Split Data*

Dari 11.890 komentar, data akan dibagi menjadi 2 bagian dengan rasio 80:20. Sebanyak 80% digunakan dalam data latih. Jumlah data latih adalah 9.512 data. 20% dari sisa data latih digunakan dalam data training sebanyak 2.378 data. Pembagian dilakukan menggunakan metode stratified sampling sehingga proporsi setiap kelas sentimen tetap terjaga pada kedua kelompok data.

3.4 Hasil Pembobotan TF-IDF

Setelah proses pembagian data, data latih kemudian dibobotkann menggunakan metode TF-IDF. TF-IDF menghasilkan representasi numerik dari setiap dokumen berdasar korpus. Hasil ini kemudian akan digunakan dalam proses klasifikasi Naïve Bayes dan Decision Tree. Sampel dari hasil tf idf adalah sebagai berikut

Tabel 2. Lima sampel data yang telah dibobotkan TF-IDF

Fitur	Bobot TF-iDF
Sedih	0.0829
perintah	0,0206
tanya	0,0116
lihat	0,0106
doa	0,0106

3.5 Perbandingan Hasil Implementasi Naïve Bayes dan Decision tree

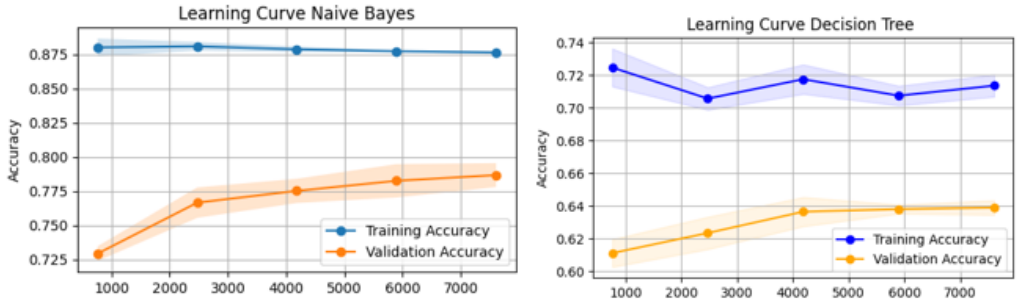
1) Parameter Model

Pada algoritma Naïve bayes pada penelitian ini menggunakan parameter alpha sebesar setengah satuan sebagai nilai laplace smoothing. Parameter ini digunakan untuk mengurangi munculnya probabilitas nol pada proses klasifikasi sehingga model dapat menghasilkan hasil prediksi dengan stabil.

Pada Algoritme *Decision Tree* digunakan parameter *entropy* sebagai dasar pemilihan atribut terbaik pada setiap proses pemisahan node. Untuk kedalaman pohon sendiri dibatasi dengan *max_depth* 20 satuan untuk mengurangi overkompleksitas model dan meminimalisir overfitting. Parameter *min_sample_split* 5 satuan digunakan untuk menentukan satuan minimal sampel yang digunakan sebelum suatu node dapat bercabang. Setelahnya *class_weight* dengan himpunan {0:1. 1:1. 2:1} digunakan untuk memberikan bobot yang sama pada seluruh kelas sentimen.

2) Hasil Training Model

Hasil dari Training model pada 9127 dengan extrasi fitur TF-IDF menghasilkan *learning curve* pada setiap algoritma klasifikasi sebagai yang disajikan pada gambar 2.



Gambar 2 Kurva Latih (a) *Decision Tree* (b) *Naïve Bayes*

Learning Curve menunjukkan akurasi latih Naïve Bayes berada pada skala 0,87 sampai dengan 0,88. Kurva validasi meningkat secara bertahap dari 0,73 menuju 0,79 seiring penambahan data. Disisi lain Decision Tree mendapatkan kurva akurasi latih sebesar 0,71 sampai dengan 0,72, dan kurva akurasi validasi pada skala 0,61 sampai dengan 0,64.

Selain kurva latih, dilakukan pemeriksaan kualitatif dengan mengambil berberapa contoh sampel komentar yang berhassil diklasifikasikan dengan benar oleh masing-masing model secara acak. Berikut berberapa sampel komentar yang berhasil di klasifikasikan oleh kedua algoritma pada tabel 3

Tabel 3 Komentar yang berhasil diklasifikasikan

Algortima	Komentar Hasil Preprocess	Kelas
Naïve Bayes	salut buat kerja keras perintah tetap hadir bantu korban bencana aceh sumatra bukti nyata peduli	Positif
	salut sama perintah indonesia terus sigap tahu prioritas bencana landa	Positif
	biar enggak cuma omong doang yuk bukti sikap peduli lewat bantu langsung	Negatif
	oi perintah mana semua janji manis mu	Negatif
	jabat manfaat rakyat tanggung mahasiswa galang donasi	Netral
Decision Tree	indonesia makin solid kerja perintah cepat tanggap bantu korban bencana aceh sumatra	Positif
	jauh perintah selalu bantu pulih pasca bencana	Positif
	perintah daerah tidak pernah keliling dengar rakyat memang paling benar viralkan biar banyak up presiden lihat	Negatif
	terus uang negara pakai buat apa tanya percuma dong bayar pajak...	Negatif
	henti saling tuding mulai saling rangkul biar pulih bareng	Netral

Selain sampel komentar yang berhasil diklasifikasikan, ditemukan juga sejumlah komentar yang gagagl dikllasifikaskan sesuai label. Berberapa contoh kesalahan tersebut disajikan pada tabel 5

Tabel 4. Komentar yang gagal di klasifikasikan

Algortima	Komentar Hasil Preprocess	Prediksi	Label Aslil
Naïve Bayes	perintah hidup enak rakyat derita sedih	Positif	Negatif
	perintah anggap anak anak tidak kaget menyebrangi jembatan sirotol mustakim	Netral	Negatif
	jangan biar korban bencana rasa sendiri yuk teman tindak	Negatif	Positif
	perintah terus upaya tangan masalah indonesia masuk jaga alam moga wilayah dampak bencana cepat pulih	Netral	Positif
	rakabuming kecewa maaf bapak tagmasih bapak kecewa	Negatif	Netral
Decision Tree	bukan konoha kalau konoha jabat hebat sedih sedih	Positif	Negatif
	innalillahi wa inna ilaihi rajiun presiden iya tenang tidak rasa salah sedih sakit hati..	Positif	Negatif
	kata sih menteri bodoh kata pohon tumbang alami	Netral	Negatif
	uang pajak bayar bukan manfaat koruptor bukti	Netral	Negatif
	my biggest flex is tidak ikut pilih kemarin bagus	Netral	Netral

3) Hasil validasi *5-Fold Cross Validation*

5-fold Cross Validation dilakukan untuk evaluasi kestabilan modle pada data latih. Proses validasi dilakukan dengan membagi data latih menjadi 5 bagian *fold* dari 9.512 data latih komentar menggunakan *Stratified K-fold* dengan tujuan menjaga proporsi setiap kelas sentimen. Hasil akurasi pada setiap fold adalah seperti pada tabel 3

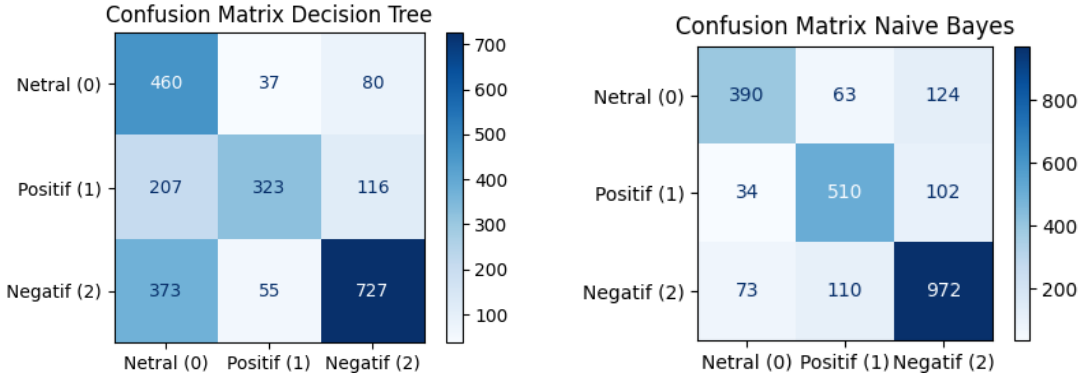
Tabel 5. Akurasi Setiap *fold* pada *K-fold cross validation*

Algortima	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Decision Tree	0,64	0,64	0.63	0.64	0,67
Naïve Bayes	0.80	0.78	0.78	0.77	0,78

Maka bisa dari tabel 3 didapat *mean accuracy* dari masing-masing algoritma *Naïve Bayes* dan *Decision Tree* Berturut-turut adalah 0,78 dan 0,63

4) Hasil Pengujian Model pada Data Tes

Pada tahap ini model diuji dengan data uji yang sudah dipisah sebelumnya. Data uji berjumlah 2378 yang merupakan 20% dari total keseluruhan set data. Hasil dari klasifikasi model setiap algoritma kepada data uji disajikan pada gambar 2.



Gambar 3 *Confusion Matrix* (a) *Decision Tree* (b) *Naïve Bayes*

Maka dapat dihitung *Accuracy, Precision, Recall,* dan *F1-Scorel* dari masing-masing model. Nilai-nilai tersebut adalah sebagai berikut

Tabel 6. Hasil Evaluasi Model terhadap data uji

Algortima	*Accuracy*	*Precision*	*Recall*	*F1-Score*
Decision Tree	0,63	0,70	0.63	0.64
Naïve Bayes	0.79	0.79	0.79	0.79

5) Pemeriksaan Overfitting

Pemeriksaan overfitting dilakukan dengan membandingkan akurasi model pada data latih, akurasi rata-rata hasil *crossvalidation* dan *akurasi* pada data uji. Hasil pemeriksaan disajikan pada tabel 7:

Tabel 7. *Gap* Akurasi Model

Algortima	*Train Accuracy*	*CV Accuracy*	*Test Accurracy*	*Train-Test gap*
Decision Tree	0,7126	0,6493	0.6350	0.0776
Naïve Bayes	0.8764	0.7869	0.7889	0.0875

Berdasar tabel 7, kedua algortima menunjukkan nilai *training accuracy* yang lebih tinggi dibandingkan testing *acccuracy.* Terdapat selisih performa antara data latih dan uji sebesar 0,0875 untuk *Naïve Bayes*, sedangkan *Decision Tree* memiliki selisih sebesar 0,0776

3.6 Pembahasan

Naïve Bayes secara konsisten lebih unggul daripada Decision dapat ditunjukan dari hasil pengujuan. Hal ini tercermin dari nilai *mean accuracy* dari K-Fold cross Validation sebesar 0,7849 lebih besar dari 0,649 angka *Decision Tree*. Akurasi data uji juga sebesar 0,79 darioada 0,64 milik *Decision Tree*. Hasil ini dari validasi ini menunjukan bahwa performa Naïve Bayes memberikikan bapabilitas model yang stabil dalam melakukan generalisasi pola sentimen dari data komentar media sosial terkair bencana hidro meteorologi. Dengan demikian, tujuan dari penelitian dalam membandingkan efektivitas kedua alforitma terjawab dengan Naïve Bayes menjadi model yang lebih andal dalam melakukan klasifikasi sentimen dampak bencana hidrometeorologi.

Keunggulan *Naïve Bayes* dapat terjadi karena kesesuaian asumsi independensi antar fitur dengan karakteristik TF-IDF yang berdimensi tinggi dan jarang (*sparse)*, sehingga model mampu mengumpulkan probabilitas seluruh fitur secara bersamaan tanpa tergandung pada pemilhan subset fitur tertentu [15]. Disisi lain, Decision Tree menggunakan mekanisme *grreedy Splitting* yang hanya memilih salah satu fitur domiann pada tiap simpul pohon pada kelas terbanyak. Pemilihan kelas dari algoritma ini menyebabkan pengabaian fitur pembeda kelas minoritas yang akhirnya menjelaskan mengapa decision tree dapat gagal mengenali kelas positif dan mendepatkan recall sebesar 0,5 meskipun parameter *class_weight* diatur pada *balanced.* Tercermin juga pada rendahnya *train accuracy* sebesar 0,712 yang menunjukan model belum mempelajari pola data secara maksimal *(underfitting)* [16].

Meski demikian, Naïve Bayes juga tidak terlepas dari kelemahan terutama pada data netral. Naïve Bayes memperoleh recall terendah pada angka 0,68 pada kelas netralll. Kesalahan ini berakar dari fenomena *class overlap*. Class overlap adalah suatu kondisi dimana frasa pada komentar netral juga sering muncul pada komentar kelas positif dan negatif. Hal ini menyebabkan probablitias kata menjadi bias akibat asumsi independen dari fitur yang mengabaikan konteks kalimat secara utuh (*bag-words limitation).* Kondisi ini diperkeruh dengan proporsi data kelas netral yang lebih sedikit dibandingkan kelas negatif, sehingga probabilitas prior turut condong ke kelas mayoritas [15] .

Temuan ini sejalan dengan prinsip dasar *text mining* yang menyebutkan performa algoritma klasifikasi sangat dipengaruhi oleh kesesuaian antara asumsi model dan karakteristik daripada struktur data. Naïve Bayes terkenall akan sebagai baseline yang kuat dalam klasifikasi teks karena kesederhanaan asumsi dari probablistik sehingga menjadi kekuatan pada data sparse dan berdimensi tinggi [17]. Sebagaimana pada penelitian di rana opini publik dan kebencanaan lainya dipaporkan bahwa naïve bayes secara konsisten memberikan performa yang kompetitif dibandingkan model berbasis pohon keputusan [18].

Penelitian ini berkontribusi memperkaya kajian analisis sentimen pada domain teks dengan konteks respond masyarakat dampak bencana hidrometeorologi. Meski demikian, penelitian ini memiliki berberapa keterbatasan seperti ketidakseimbangan proporsi antar data setiap kelas sentmen, keterbatasan TF-IDF dalam menangkap makna semantik dan sarkasme, serta ekspresi duka yang menyerupai sentimen negatif.

4. Simpulan

Berdasarkan hasil dapat dilihat bahwa algoritma *Naïve Bayes* secara konsisten menunjukkan performa yang lebih baik daripada *Decision Tree*. Pada prose *K-Fold Cross Validation*, Naïve Bayes mampu memperoleh *Mean Accuracy* sebesar 0,79 dibandingkan *Decision Tree* yang hanya memperoleh 0,65. Pada evaluasi data uji, Algoritma Naïve Bayes juga lebih unggul dalam *accuracy, precision, recall,* dan *F1-Score* yang seragam sebesar 0,79. Disisi lain *Decision Tree* memperoleh *accuracy, precision, recall,* dan *F1-Score* berturut turut sebesar 0,63, 0,70, 0,63 dan 0,64.

Analisis *confusion matrix* menunjukkan bahwa *Naïve Bayes* mengalami kesalahan klasifikasi terbanyak pada kelas netral dengan recall sebesar 0,68. Sedangkan performa terbaik pada kelas terbanyak yaitu kelas negatif. Hal ini menunjukkan kemampuan Naïve Bayes berbanding lurus dengan banyak nya data. Sedangkan *confusion matrix* milik *Decision tree*

menunjukkan kekuatan pada kelas netral dengan recall tertinggi 0,80 dan terendah pada kelas positif hanya 0,50. Perbedaan pola ini menunjukkan Naïve Bayes unggul dalam mengklasifikasikan data daripada *Decision Tree.*

Pada Pemeriksaan overfitting melalui train accuracy, cross validation accuracy dan test accuracy memperkuat bahwa naïve bayes lebih unggul daripada *Decision Tree.* Naïve Bayes menunjukkan train-test gap sebesar 0,0875 dengan kurva learning curve yang saling mendekat seiring bertambahnya data. Menunjukkan overfitting ringan seiring bertambahnya data latih sehingga mengindikasikan kemampuan generalisasi yang tetap baik. Disisi lain Decision Tree memperoleh train-test gap sebesar 0,0776, namun dengan train accuracy yang rendah yakni 0,71. menunjukkan bahwa model belum mempelajari pola data secara maksimal bahkan pada data latihnya sendiri.

Penelitian ini memiliki berberapa keterbatasan yang perlu diperhatikan. Diantaranya pengumpulan data yang dipengaruhi oleh rate limit pada masing-msing platform media sosial sehingga jumlah komentar yang diperoleh dari Tiktok, Instagram, dan Facebook tidak sepenuhnya seimbang. Distribusi data pada setiap kelas sentimen cukup signifikan yang berpotensi memengaruhi kemampuan model sehingga cenderung mengklasifikasikan kepada data mayoritas. TF-IDF juga membuat keterbatasan sehingga model hanya terbatas pada frekuensi dan belum mampu menangkap konteks kalimat seperti sentimen berbentuk sarkasme. Keterbatasan tersebut terlihat ketika adanya kesalahan klasifikasi pada komentar yang mempunyai kemiripan kosa kata antar kelas sentimen.

Berdasarkan Keterbatasan tersebut, penelitian senaljutnya disarankan menggunakan dataset yang lebih besar dan seimbang. Lebih lagi dapat menerapkan suatu metode respresesntasi teks yang mampu memahami konteks semantik seperti *word embedding*. Selain itu, Peneliti selanjutnya dapat melakukan explorasi penggunaan algoritma lain atau melakukan perubahan parameter sehingga optimal dalam klasifikasi dokumen.

Daftar Referensi

[1] A. Khairani and A. Guspa, "Pengaruh self control terhadap toxic online disinhibition effect pada generasi Z pengguna aplikasi X," CAUSALITA: Journal of Psychology, vol. 3, no. 1, pp. 243–253, 2025. doi: 10.62260/causalita.v3i1.525.

[2] D. Erokhin and N. Komendantova, "Social media data for disaster risk management and research," International Journal of Disaster Risk Reduction, vol. 114, no. June, p. 104980, 2024. doi: 10.1016/j.ijdrr.2024.104980.

[3] M. Karimiziarani and H. Moradkhani, "Social response and disaster management: Insights from Twitter data assimilation on Hurricane Ian," International Journal of Disaster Risk Reduction, vol. 95, no. July, p. 103865, 2023. doi: 10.1016/j.ijdrr.2023.103865.

[4] S. M. Anugerah, R. Wijaya, and M. A. Bijaksana, "Sentiment analysis social media for disaster using Naïve Bayes and IndoBERT," INTEK: Jurnal Penelitian, vol. 11, no. 1, pp. 51–58, 2024. doi: 10.31963/intek.v11i1.4771.

[5] V. R. Prasetyo, G. Erlangga, D. A. Prima, U. Surabaya, and P. Korespondensi, "Analisis sentimen untuk identifikasi bantuan korban bencana alam berdasarkan data di Twitter menggunakan metode K-means dan Naïve Bayes," Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), vol. 10, no. 5, pp. 1055–1062, 2023. doi: 10.25126/jtiik.2023107077.

[6] I. F. Rozi, A. T. Firdausi, and K. Islamiyah, "Analisis sentimen pada Twitter mengenai pasca bencana menggunakan metode Naïve Bayes dengan fitur N-Gram," Jurnal Informatika Polinema, vol. 6, no. 2, pp. 33–39, 2020. doi: 10.33795/jip.v6i2.316.

[7] M. Das, S. Kamalanathan, and P. Alphonse, "A comparative study on TF-IDF feature weighting method and its analysis using unstructured dataset," in Proceedings of the 5th International Conference on Computational Linguistics and Intelligent Systems (COLINS), Kharkiv, Ukraine, Apr. 22–23, 2021, pp. 0–2. arXiv:2308.04037.

[8] N. A. Ionendri, F. Candra, and A. Rizal, "News classification using natural language processing with TF-IDF and multinomial Naïve Bayes," Journal of Applied Computer Science and Technology (JACOST), vol. 6, no. 1, pp. 37–45, 2025. doi: 10.52158/jacost.v6i1.1099.

[9] P. Yuniar and Kismiantini, "Analisis sentimen ulasan pada Gojek menggunakan metode Naïve Bayes," Statistika, vol. 23, no. 2, pp. 164–175, 2023. doi: 10.29313/statistika.v23i2.2353.

[10] F. Baehaqi and N. Cahyono, "Analisis sentimen terhadap cyberbullying pada komentar di Instagram menggunakan algoritma Naïve Bayes," Indonesian Journal of Computer Science, vol. 13, no. 1, pp. 1051–1063, 2024. doi: 10.33022/ijcs.v13i1.3301.
[11] M. Maharani and F. Fathoni, "Analisis sentimen pengguna terhadap faktor penggunaan PayPal menggunakan metode decision tree," Jurnal Ilmiah Teknologi Informasi Asia, vol. 18, no. 1, pp. 71–83, 2024. doi: 10.32815/jitika.v18i1.1002.
[12] S. Pawar, M. G. Rao, and K. Pandith, "Text document categorisation using random forest and C4.5 decision tree classifier," International Journal of Computational Systems Engineering, vol. 7, no. 2–4, pp. 211–220, 2023. doi: 10.1504/IJCSYSE.2023.132924.
[13] F. M. Akbar, R. Hermansyah, S. Lusa, D. I. Sensuse, and N. Safitri, "Analisis sentimen untuk evaluasi reputasi merek motor XYZ berkaitan dengan isu rangka motor di Twitter menggunakan machine learning," Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), vol. 11, no. 3, pp. 647–654, 2024. doi: 10.25126/jtiik.938663.
[14] G. A. Trianto, M. F. Marzuki, T. Y. Sihotang, and H. Irsyad, "Klasifikasi opini terhadap resesi Indonesia 2023 pada Twitter menggunakan algoritma decision tree," in Proceedings of the 2nd Multi Data Palembang Student Conference (MSC 2023), Universitas Multi Data Palembang, pp. 1–9, 2023.
[15] A. Tounsi and M. Temimi, "A systematic review of natural language processing applications for hydrometeorological hazards assessment," Natural Hazards, vol. 116, no. 3, 2023. doi: 10.1007/s11069-023-05842-0.
[16] S. Wu, D. He, X. Wang, L. Wang, and J. Dang, "Enriching multimodal sentiment analysis through textual emotional descriptions of visual-audio content," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 39, no. 2, pp. 1601–1609, 2025. doi: 10.1609/aaai.v39i2.32152.
[17] Henderi and S. Sofiana, "A machine learning approach to Indonesian climate change sentiment analysis using Naive Bayes," International Journal of Informatics and Information Systems, vol. 8, no. 1, pp. 33–43, 2025. doi: 10.47738/ijiis.v8i1.246.
[18] M. Jamil, H. Hadiyanto, and R. Sanjaya, "Sentiment analysis: Classifying public comments on YouTube in disaster management simulation in Indonesia using Naïve Bayes and support vector machine," Ingénierie des Systèmes d'Information, vol. 29, no. 2, pp. 437–446, 2024. doi: 10.18280/isi.290205.