

Penerapan Naive Bayes yang Dioptimalkan oleh SMOTE untuk Mengatasi Ketidakseimbangan Data pada Klasifikasi Komentar Bencana Alam di Jakarta

^{1*}Nugroho Bimantoro, ²Risky Aswi Ramadhani, ³Julian Sahertian

^{1,2,3} Teknik Informatika, Universitas Nusantara PGRI Kediri

E-mail: ¹nugjimbbz@gmail.com, ²riskyaswi@gmail.com, ³juliansahertian@unpkediri.ac.id

Penulis Korespondens : Risky Aswi Ramadhani

Abstrak—Analisis sentimen merupakan salah satu bidang dalam *Natural Language Processing* (NLP) yang digunakan untuk memahami opini atau tanggapan masyarakat terhadap suatu topik tertentu. Penelitian ini berfokus pada analisis sentimen terkait bencana alam di Jakarta yang diperoleh dari media sosial X. Tujuan penelitian ini adalah untuk mengetahui performa metode Naive Bayes dalam mengklasifikasikan sentimen masyarakat terhadap bencana alam. Dataset diperoleh dengan *webscrapping* menggunakan *tweet harvest* dengan rentang waktu 5 tahun terakhir dengan hasil 1151 data. Data yang digunakan terlebih dahulu melalui tahap *preprocessing* untuk membersihkan teks sebelum dilakukan pelabelan sentimen menggunakan model RoBERTa berbahasa Indonesia. Selain itu, metode SMOTE diterapkan untuk mengatasi ketidakseimbangan jumlah data pada kelas positif dan negatif. Hasil penelitian menunjukkan bahwa penerapan SMOTE mampu membantu meningkatkan kualitas distribusi data sehingga proses klasifikasi menjadi lebih optimal. Berdasarkan hasil pengujian, metode Naive Bayes memperoleh tingkat akurasi sebesar 76% dalam proses analisis sentimen. Penelitian ini menunjukkan bahwa kombinasi RoBERTa, SMOTE, dan Naive Bayes dapat memberikan hasil yang cukup baik dalam klasifikasi sentimen pada data media sosial X.

Kata Kunci— Analisis Sentimen, X, Bencana Alam, Naive Bayes, RoBERTa, SMOTE

Abstract— *Sentiment analysis is a field in Natural Language Processing (NLP) that is used to understand public opinion or responses to a particular topic. This study focuses on sentiment analysis related to natural disasters in Jakarta obtained from social media X. The purpose of this study is to determine the performance of the Naive Bayes method in classifying public sentiment towards natural disasters. The dataset was obtained through webscrapping using Harvest tweets with a time span of the last 5 years with the result of 1151 data. The data used first went through a preprocessing stage to clean the text before sentiment labeling was carried out using the Indonesian language RoBERTa model. In addition, the SMOTE method was applied to overcome the imbalance in the number of data in the positive and negative classes. The results of the study show that the application of SMOTE can help improve the quality of data distribution so that the classification process becomes more optimal. Based on the test results, the Naive Bayes method obtained an accuracy rate of 76% in the sentiment analysis process. This study shows that the combination of RoBERTa, SMOTE, and Naive Bayes can provide quite good results in sentiment classification on social media data X.*

Keywords— *Sentiment Analysis, X, Disaster, Naive Bayes, RoBERTa, SMOTE,*

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Banjir di Jakarta merupakan permasalahan lingkungan yang terus berulang setiap tahun dan berdampak luas terhadap aktivitas sosial, ekonomi, serta psikologis masyarakat. Urbanisasi yang cepat dan pengelolaan tata ruang yang kurang optimal memperparah potensi genangan di berbagai wilayah kota. Pakar tata kota dan klimatologi menyatakan bahwa meluasnya banjir di wilayah Jabodetabek tidak hanya disebabkan oleh intensitas hujan, melainkan dampak nyata dari krisis iklim global serta masalah struktural seperti penurunan muka tanah di Jakarta (Kompas.com). Situasi ini memicu munculnya keresahan publik yang kerap diekspresikan di media sosial, khususnya X. Dalam konteks ini, X tidak hanya menjadi sarana komunikasi antarwarga, tetapi juga sumber data penting yang mencerminkan persepsi dan emosi publik secara waktu nyata [1].

Pada penelitian sebelumnya, mencatat bahwa performa Naive Bayes mengalami penurunan apabila dihadapkan pada variasi kata yang terlalu beragam serta kondisi kelas data yang tidak seimbang. Selain itu, proses pelabelan datanya masih bergantung pada penanganan manual. Oleh karena itu, penelitian ini menerapkan RoBERTa untuk *automatic* labeling serta penerapan SMOTE untuk mengatasi ketimpangan kelas data [2]

Analisis sentimen merupakan salah satu bidang dalam *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi, mengolah, dan mengklasifikasikan opini atau pendapat seseorang terhadap suatu topik tertentu[3]. Analisis sentimen bertujuan untuk mengetahui kecenderungan opini yang termasuk ke dalam kategori positif, negatif, atau netral berdasarkan data teks yang dianalisis. Dalam penelitian ini, analisis sentimen digunakan untuk mengetahui respon masyarakat terhadap bencana banjir di Jakarta melalui media sosial X [4]

Data mentah dilakukan preprocessing setelahnya akan dilakukan proses labeling. Tahap preprocessing diantaranya cleansing, casefolding, tokenizing, normalisasi, stopword removal, dan stemming [5]. RoBERTa (*Robustly Optimized Bidirectional Encoder Representations from Transformers*) merupakan pengembangan dari model BERT yang dirancang untuk meningkatkan kemampuan pemrosesan bahasa alami (*Natural Language Processing*). RoBERTa mampu memahami konteks kata dalam suatu kalimat dengan lebih baik sehingga banyak digunakan dalam tugas klasifikasi teks dan analisis sentimen. Dalam penelitian ini, RoBERTa digunakan untuk proses *automatic* labeling sentimen pada dataset X terkait banjir di Jakarta [6]

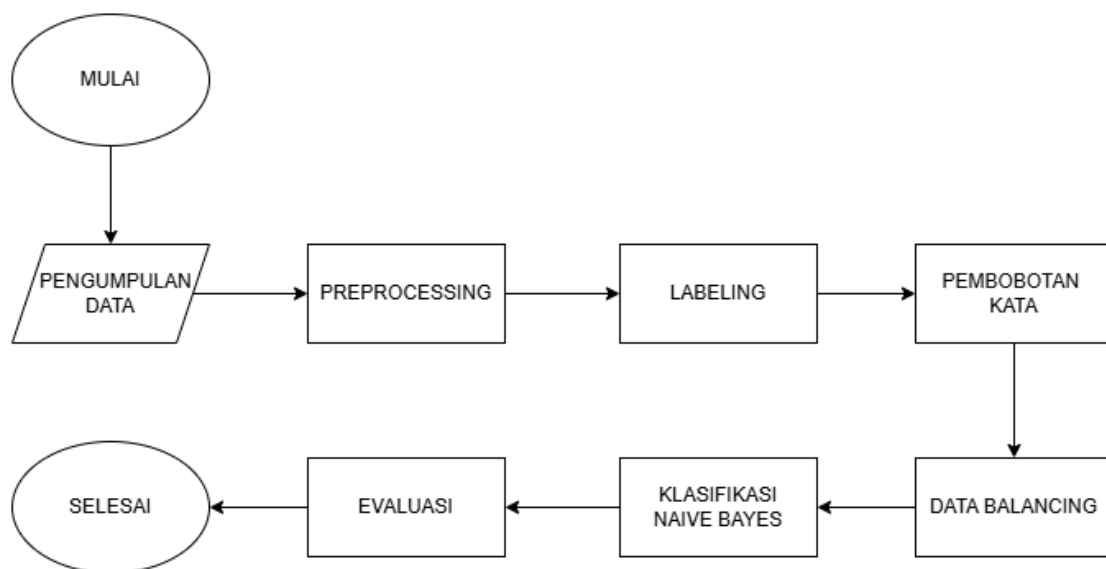
SMOTE (*Synthetic Minority Over-sampling Technique*) merupakan metode yang digunakan untuk mengatasi masalah ketidakseimbangan data pada proses klasifikasi. Ketidakseimbangan data terjadi ketika jumlah data pada salah satu kelas lebih sedikit dibandingkan kelas lainnya, sehingga model cenderung lebih mengenali kelas mayoritas. Dalam analisis sentimen, kondisi tersebut dapat menyebabkan hasil klasifikasi menjadi kurang optimal, terutama pada kelas minoritas[7]. Dalam penelitian ini menggunakan SMOTE untuk kelas minoritas positif dan negatif.

Naive Bayes merupakan algoritma klasifikasi berbasis probabilitas yang digunakan dalam text mining untuk mengklasifikasikan dokumen ke dalam kategori tertentu dengan menghitung kemungkinan masing-masing kelas berdasarkan frekuensi kemunculan fitur atau kata pada teks[8]. Algoritma ini menggunakan prinsip Teorema Bayes dengan asumsi bahwa setiap fitur bersifat independen satu sama lain, sehingga dapat memperkirakan kelas paling mungkin dari suatu teks berdasarkan nilai probabilitasnya. Keunggulan Naive Bayes terletak pada sederhana, cepat, dan relatif efektif dalam mengolah data teks dalam jumlah besar [9]

Penelitian ini terfokus pada analisis sentimen tweet mengenai bencana alam di Jakarta yang diperoleh dari media sosial X. Penelitian ini memanfaatkan model RoBERTa untuk melakukan pelabelan otomatis pada dataset sentimen berbahasa Indonesia. Hasil pelabelan tersebut kemudian digunakan dalam proses klasifikasi menggunakan algoritma Naive Bayes. Selain itu, penelitian ini menerapkan metode SMOTE (*Synthetic Minority Oversampling Technique*) untuk menangani ketidakseimbangan kelas pada dataset sentimen. Tujuan utama penelitian ini adalah untuk mengetahui performa algoritma Naive Bayes dalam mengklasifikasikan sentimen tweet terkait bencana alam di Jakarta setelah dilakukan *automatic* labeling menggunakan RoBERTa dan penanganan data tidak seimbang menggunakan SMOTE. Diharapkan hasil penelitian ini dapat memberikan wawasan mengenai kemampuan algoritma Naive Bayes dalam melakukan klasifikasi sentimen pada data media sosial berbahasa Indonesia serta pengaruh penerapan SMOTE terhadap hasil klasifikasi.

II. METODE

Tahapan-tahapan metode penelitian diantaranya.



Gambar 1. Alur Penelitian

2.1 Pengumpulan Data

Dataset diperoleh dengan cara *webscraping* menggunakan *tweet harvest* dengan kata kunci seperti “banjir jakarta”, “banjir ibukota” dalam rentang waktu 5 tahun terakhir. Dataset berjumlah 1151 tweet dengan format CSV yang memuat *conversation_id_str*, *created_at*, *id_str*, *full_text*, *image_url*, *favorite_count*, *retweet_count*, *reply_count*, dan *tweet_url*. Berikut contoh dataset.

1	conversati	created_a	favorite_c	full_text
2	1,48E+18	Sat Jan 08	930	Mari sejenak memanjatkan doa untuk saudara kita di Papua atas musibah banjir ban
3	1,36E+18	Tue Feb 09	564	Banjir Jakarta cepat surut. Alhamdulillah. Jl Bungur Gunung Sahari Selatan surut dlm
4	1,48E+18	Mon Jan 1	0	JAKARTA BANJIR AIR MATA https://t.co/AvkAk2pnDb
5	1,36E+18	Mon Feb 0	158	Jakarta selatan terendam banjir hanya 2 meter. https://t.co/O65ZNdqzox
6	1,36E+18	Mon Feb 0	3214	Genangan 10-40cm di Jl Pegangsaan Dua Kelapa Gading dalam 2 jam sudah surut. Te
7	1,48E+18	Mon Jan 1	48	Semalaman hujan di Jakarta pagi ini mengantar anak sekolah ke TEBET dr CENGKARE
8	1,48E+18	Thu Jan 06	23	Pagi ini #MensosRisma menerima bantuan senilai Rp1 miliar untuk pembangunan NT
9	1,48E+18	Sun Jan 05	14	@tvOneNews Ga bakalan banjir lg jakarta.. https://t.co/dJIPOcpTuG
10	1,48E+18	Mon Jan 1	11	Ini Jakarta? Mosok sich Keren banget Hujan koq gak banjir https://t.co/rSucnvGnVY

Gambar 2. Contoh Dataset

2.2 Preprocessing

Dalam preprocessing ada beberapa tahap, diantaranya Cleansing, Casefolding, Tokenizing, Normalisasi, Stopword, Stemming [10]. Berikut contoh dalam tabel.

Tabel 1. Cleansing

menghapus elemen yang tidak relevan dari teks, seperti URL, tag HTML, atau simbol-simbol yang tidak bermakna

Sebelum	Sesudah
Banjir di wilayah Jakarta Pusat sudah sepenuhnya surut sejak Sabtu (20/2) https://t.co/muPkpPtQLG	Banjir di wilayah Jakarta Pusat sudah sepenuhnya surut sejak Sabtu
Eneg bgt liat warga ibukota banjir dikit berisik banget. Gw disini banjir warga udah terisolir 6 hari ga da tuh gaungnya. Norak !!!	Eneg bgt liat warga ibukota banjir dikit berisik banget Gw disini banjir warga udah terisolir hari ga da tuh gaungnya Norak

Tabel 2. Casefolding

menyeragamkan huruf dengan cara mengubah seluruh teks menjadi huruf kecil

Sebelum	Sesudah
Banjir di wilayah Jakarta Pusat sudah sepenuhnya surut sejak Sabtu	banjir di wilayah jakarta pusat sudah sepenuhnya surut sejak sabtu
Eneg bgt liat warga ibukota banjir dikit berisik banget Gw disini banjir warga udah terisolir hari ga da tuh gaungnya Norak	eneg bgt liat warga ibukota banjir dikit berisik banget gw disini banjir warga udah terisolir hari ga da tuh gaungnya norak

Tabel 3. Tokenizing

memecah teks menjadi unit-unit kecil yang disebut token

Sebelum	Sesudah
banjir di wilayah jakarta pusat sudah sepenuhnya surut sejak sabtu	'banjir', 'di', 'wilayah', 'jakarta', 'pusat', 'sudah', 'sepenuhnya', 'surut', 'sejak', 'sabtu'
eneg bgt liat warga ibukota banjir dikit berisik banget gw disini banjir warga udah terisolir hari ga da tuh gaungnya norak	'eneg', 'bgt', 'liat', 'warga', 'ibukota', 'banjir', 'dikit', 'berisik', 'banget', 'gw', 'disini', 'banjir', 'warga', 'sudah', 'terisolir', 'hari', 'ga', 'da', 'tuh', 'gaungnya', 'norak'

Tabel 4. Normalisasi

memperbaiki berbagai bentuk penulisan kata slang, singkatan, typo

Sebelum	Sesudah
'banjir', 'di', 'wilayah', 'jakarta', 'pusat', 'sudah', 'sepenuhnya', 'surut', 'sejak', 'sabtu'	banjir di wilayah jakarta pusat sudah sepenuhnya surut sejak sabtu
'eneg', 'bgt', 'liat', 'warga', 'ibukota', 'banjir', 'dikit', 'berisik', 'banget', 'gw', 'disini', 'banjir', 'warga', 'sudah', 'terisolir', 'hari', 'ga', 'da', 'tuh', 'gaungnya', 'norak'	eneg banget liat warga ibukota banjir dikit berisik banget saya disini banjir warga sudah terisolir hari tidak da tuh gaungnya norak

Tabel 5. Stopword Removal

menghapus kata hubung yang tidak penting

Sebelum	Sesudah
banjir di wilayah jakarta pusat sudah sepenuhnya surut sejak sabtu	banjir wilayah jakarta pusat sepenuhnya surut sabtu
eneg banget liat warga ibukota banjir dikit berisik banget saya disini banjir warga sudah terisolir hari tidak da tuh gaungnya norak	eneg banget liat warga ibukota banjir dikit berisik banget banjir warga terisolir tidak da tuh gaungnya norak

Tabel 6. Stemming

mengubah kata berimbuhan menjadi kata dasarnya

Sebelum	Sesudah
banjir wilayah jakarta pusat sepenuhnya surut sabtu	banjir wilayah jakarta pusat penuh surut sabtu
eneg banget liat warga ibukota banjir dikit berisik banget banjir warga terisolir tidak da tuh gaungnya norak	eneg banget liat warga ibukota banjir dikit berisik banget banjir warga isolir tidak da tuh gaung norak

2.3 Labeling

Setelah melewati *preprocessing*, kemudian data akan dilabeli menggunakan RoBERTa. Pelabelan dilakukan setelah *preprocessing* pada tahap normalisasi. Dari 1151 dataset, hasil labeling yang didapat adalah netral 712 data, negatif 336 data, positif 103 data. Berikut contoh nya.

Tabel 7. Contoh Kalimat Sentimen

Kalimat	Sentimen
kabar baru akibat intensitas hujan badan meteorologi geofisika bmkkg ingat status siaga banjir dki jakarta tanggal januari	Netral
besok jakarta banjir	Negatif

2.4 Pembobotan Kata

Setelah melewati proses labeling, selanjutnya masuk ke tahap TF-IDF (*Term Frequency-Inverse Document Frequency*) yaitu, teknik pembobotan kata dalam text mining yang digunakan untuk mengukur tingkat pentingnya sebuah istilah dalam suatu dokumen dibandingkan dengan keseluruhan kumpulan dokumen. Metode ini bekerja dengan menggabungkan frekuensi kemunculan kata dalam sebuah dokumen (*term frequency*) dan seberapa jarang kata tersebut muncul di semua dokumen (*inverse document frequency*), sehingga menghasilkan bobot yang lebih tinggi pada kata yang lebih informatif [11]. Berikut adalah contoh bobot kata paling tinggi.

Tabel 8. TF-IDF

Kata	Bobot
banjir	38.360226
jakarta	35.422851
ibukota	19.483017

2.5 Data Balancing

Pada penelitian ini, SMOTE diterapkan untuk menambah jumlah data pada kelas sentimen positif dan negatif sebagai kelas minoritas, sehingga distribusi data menjadi lebih seimbang dan model klasifikasi dapat mengenali pola sentimen positif dan negatif dengan lebih baik. Rasio untuk data latih dan data uji yaitu 80:20 (920:231), dengan SMOTE diterapkan pada data latih, berikut adalah contoh nya.

```
--- SEBELUM SMOTE ---  
Jumlah data latih tiap kelas: {'Positif': 81, 'Netral': 576, 'Negatif': 263}  
  
--- SESUDAH SMOTE ---  
Jumlah data latih tiap kelas: {'Positif': 576, 'Netral': 576, 'Negatif': 576}
```

Gambar 3. Penerapan SMOTE

2.6 Klasifikasi Naive Bayes

Data yang telah diubah ke dalam bentuk representasi fitur numerik selanjutnya digunakan pada proses pelatihan dan pengujian model klasifikasi. Sebelum proses klasifikasi dilakukan, metode SMOTE diterapkan untuk mengatasi ketidakseimbangan jumlah data. Pada penelitian ini, algoritma Naive Bayes digunakan karena memiliki proses klasifikasi yang sederhana dan efektif untuk data teks, Naive Bayes melakukan klasifikasi dengan menghitung probabilitas kemunculan setiap kata terhadap kelas tertentu dan menentukan kelas dengan nilai probabilitas tertinggi sebagai hasil prediksi [12].

III. HASIL DAN PEMBAHASAN

Pada bagian ini, dipaparkan hasil analisis sentimen dari data yang telah dikumpulkan, serta pembahasan mengenai akurasi dan performa dari Naive Bayes yang dikombinasikan dengan RoBERTa, dan SMOTE.

	precision	recall	f1-score	support
Negatif	0.84	0.42	0.56	73
Netral	0.69	0.98	0.81	136
Positif	0.00	0.00	0.00	22
accuracy			0.71	231
macro avg	0.51	0.47	0.46	231
weighted avg	0.67	0.71	0.65	231

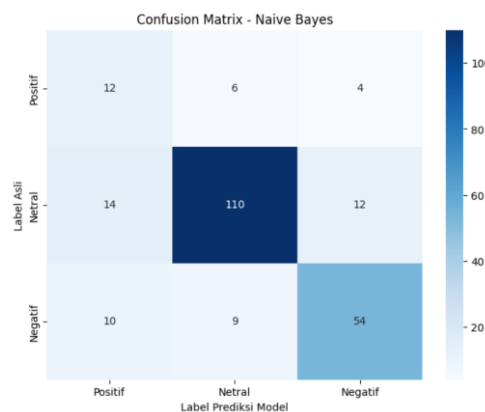
Gambar 4. Hasil Tanpa SMOTE

Hasil pengujian model Naive Bayes tanpa SMOTE menunjukkan akurasi sebesar 71% dengan total data uji sebanyak 231 sampel. Model ini bekerja sangat baik pada kelas “Netral” dengan f1-score mencapai 0.81, diikuti kelas “Negatif” sebesar 0.56. Namun, akibat adanya ketidakseimbangan data, model sama sekali gagal mengenali sentimen “Positif” dengan nilai *precision*, *recall*, dan *f1-score* sebesar 0.00.

	precision	recall	f1-score	support
Negatif	0.77	0.74	0.76	73
Netral	0.88	0.81	0.84	136
Positif	0.33	0.55	0.41	22
accuracy			0.76	231
macro avg	0.66	0.70	0.67	231
weighted avg	0.79	0.76	0.77	231

Gambar 5. Hasil Dengan SMOTE

Setelah menerapkan teknik SMOTE, performa klasifikasi Naive Bayes mengalami peningkatan akurasi menjadi 76%. Dampak paling signifikan terlihat pada kelas minoritas (“Positif”) yang kini berhasil dikenali dengan nilai *recall* sebesar 0.55 dan *f1-score* 0.41. Sementara itu, kelas “Netral” tetap mempertahankan performa tertinggi dengan *f1-score* sebesar 0.84, diikuti kelas “Negatif” sebesar 0.76.



Gambar 6. Confusion Matrix

Berdasarkan *confusion matrix* pada Gambar 6, model Naive Bayes berhasil memprediksi data dengan benar secara signifikan, yaitu 110 data “Netral”, 54 data “Negatif”, dan 12 data “Positif”. Namun, masih terdapat beberapa kesalahan klasifikasi seperti 14 data “Netral” dan 10 data “Negatif” yang keliru diprediksi sebagai sentimen “Positif”.

Proses pengujian performa model klasifikasi pada penelitian ini diawali dengan mengatasi ketimpangan data latih menggunakan SMOTE dengan rasio pembagian data train dan test sebesar 80:20. Perubahan distribusi data latih sebelum dan sesudah penerapan SMOTE disajikan secara mendalam pada Gambar 3. Berdasarkan grafik tersebut, terlihat bahwa data awal memiliki ketidakseimbangan kelas yang sangat mencolok, di mana kelas netral menjadi mayoritas dengan 576 data, diikuti kelas negatif 263 data, dan kelas positif yang hanya berjumlah 81 data. Melalui SMOTE, data pada kelas minoritas (positif dan negatif) disintesis secara otomatis hingga jumlahnya setara dengan kelas mayoritas, sehingga menghasilkan distribusi data latih yang seimbang secara merata, yakni masing-masing sebanyak 576 data untuk setiap kategori sentimen. Setelah data seimbang, dilakukan proses pengujian menggunakan algoritma Naive Bayes yang dikombinasikan dengan pembobotan fitur TF-IDF. Hasil evaluasi performa klasifikasi model

Naive Bayes dipaparkan secara lengkap melalui perbandingan laporan kinerja pada *Gambar 4* (tanpa SMOTE) dan *Gambar 5* (dengan SMOTE), serta visualisasi kontribusi tebakan model pada *Confusion Matrix* di *Gambar 6*.

Berdasarkan hasil pengujian yang tertera pada *Gambar 5*, integrasi model Naive Bayes setelah penerapan SMOTE berhasil memperoleh tingkat akurasi keseluruhan sebesar 76%. Jika dianalisis secara komparatif dengan hasil pengujian tanpa SMOTE pada *Gambar 4*, terlihat jelas bahwa penyeimbangan data latih memiliki peran yang sangat krusial. Tanpa adanya SMOTE, model Naive Bayes cenderung mengalami bias mayoritas dan sama sekali gagal mengenali sentimen minoritas, yang dibuktikan dengan nilai *precision*, *recall*, dan *f1-score* untuk kelas positif sebesar 0.00%. Namun, efektivitas perubahan ini dapat dipahami lebih detail melalui *Confusion Matrix* pada *Gambar 6*. Dari total 22 data uji asli pada kelas positif, model berhasil memprediksi dengan benar sebanyak 12 data, sementara sisanya salah diprediksi sebagai netral (6 data) dan negatif (4 data).

IV. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa penerapan metode Naive Bayes yang dikombinasikan dengan pembobotan kata TF-IDF terbukti efektif dalam mengklasifikasikan sentimen masyarakat di media sosial X terkait bencana alam di Jakarta. Hasil pengujian menunjukkan bahwa model klasifikasi berhasil mencapai tingkat akurasi keseluruhan sebesar 76%. Peran *Synthetic Minority Over-sampling Technique* (SMOTE) terbukti sangat krusial dalam mengatasi masalah ketidakseimbangan kelas data awal. Melalui penyeimbangan data latih (*Gambar 3*) dan analisis pemetaan prediksi pada *Confusion Matrix* (*Gambar 6*), model mampu mengatasi bias mayoritas yang sebelumnya membuat kelas minoritas positif tidak dikenali sama sekali (0.00% pada *Gambar 4*), menjadi lebih sensitif dan optimal dalam mengidentifikasi seluruh kelas sentimen dengan peningkatan performa yang signifikan pada *Gambar 5*. Dengan demikian, kombinasi integratif dari pelabelan otomatis berbasis model RoBERTa, penanganan data menggunakan SMOTE, dan klasifikasi algoritma Naive Bayes dapat menjadi pendekatan yang andal untuk analisis sentimen kebencanaan berbasis media sosial.

DAFTAR PUSTAKA

- [1]. A. Sobri Hashim *et al.*, “Leveraging Social Media Sentiment Analysis for Enhanced Disaster Management: A Systematic Review and Future Research Agenda,” *Online Journal of System and Management Sciences*, vol. 15, no. 4, pp. 170–191, 2025, doi: 10.33168/JSMS.2025.0412.
- [2]. Kharisma Wiati Gusti, “Perbandingan Metode Support Vector Machine dan Logistic Regression untuk Klasifikasi Bencana Alam,” 2023, doi: 10.52958/iftk.v19i2.6355.
- [3]. M. A. Nur, N. Umar, Z. Feng, and H. Gani, “EVALUATION OF INDOBERT AND ROBERTA: PERFORMANCE OF INDONESIAN LANGUAGE TRANSFORMER MODELS IN SENTIMENT CLASSIFICATION,” *JIKO*

- (*Jurnal Informatika dan Komputer*), vol. 8, no. 2, pp. 121–127, Jul. 2025, doi: 10.33387/jiko.v8i2.9988.
- [4]. N. Fathirachman Mahing *et al.*, “KLASIFIKASI TINGKAT STRES DARI DATA BERBENTUK TEKS DENGAN MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN RANDOM FOREST”, doi: 10.25126/jtiik2024118010.
- [5]. Meitaningsing, Aribowo, and Cahyana, “TELEMATIKA, Vol. 17, No. 2, OKTOBER, 2020, Pp. 99 – 110 ISSN 1829-667X,” 2020, doi: 10.31315/telematika.v1i1.3377.
- [6]. R. A. Husaini, D. Priyanto, and G. H. Martono, “ANALISIS SENTIMEN KINERJA TENAGA MEDIS INDONESIA MENGGUNAKAN MODELING ROBERTA DAN METODE MACHINE LEARNING,” 2025. doi: 10.15294/eduel.v13i1.22163.
- [7]. M. A. Hermawan, A. Faqih, and G. Dwilestari, “IMPLEMENTASI AKURASI MODEL NAIVE BAYES MENGGUNAKAN SMOTE DALAM ANALISIS SENTIMEN PENGGUNA APLIKASI BRIMO,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5748.
- [8]. D. Astari Umbu Zaza *et al.*, “Penerapan Text Mining Dalam Klasifikasi Judul Skripsi Mahasiswa Pada Universitas Stella Maris Sumba Menggunakan Metode Naïve Bayes,” 2024. doi: 10.47233/jibs.v2i3.1670.
- [9]. F. Zamzami, R. Hidayat, and R. Fathonah, “PENERAPAN ALGORITMA NAIVE BAYES CLASSIFIER UNTUK ANALISIS SENTIMEN KOMENTAR TWITTER PROYEK PEMBAGUNAN IKN,” *Faktor Exacta*, vol. 17, no. 1, May 2024, doi: 10.30998/faktorexacta.v17i1.22265.
- [10]. Cha Cha Kirana and Nabila Rizky Oktadini, “Analisis Sentimen Naïve Bayes dengan TF-IDF dan 10-Fold pada Ulasan Aplikasi X,” *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 7, no. 2, pp. 523–535, Dec. 2025, doi: 10.30865/json.v7i2.9007.
- [11]. A. Asis and A. Syam, “Penerapan Metode Text Mining Tf-Idf Untuk Klasifikasi Sentimen Penggunaan Pay later Pada Belanja Online Di Media Sosial Twitter,” 2023. doi: <https://doi.org/10.36774/dipakomsi.v17i1.1402>.
- [12]. B. Z. Ramadhan, I. Riza, and I. Maulana, “Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Dengan Menggunakan Algoritma Naïve Bayes,” 2022. doi: <https://doi.org/10.30871/jaic.v6i2.4725>.