SKANIKA: Sistem Komputer dan Teknik Informatika
Volume 9, Nomor 2, Juli 2026, Halaman 296-306 E-ISSN: 2721-4788

Implementasi Metode NER Statistik Berbasis Web untuk Ekstraksi Entitas pada Berita Bencana Alam TVRI

Muhammad Iqbal Shiddiq[1*], Indra[2]
[1,2]Fakultas Teknologi Informasi, Teknik Informatika, Universitas Budi Luhur, Jakarta, Indonesia
E-mail: [1*]2211501396@student.budiluhur.ac.id, [2]indra@budiluhur.ac.id
(* : corresponding author)

Abstrak

Berita bencana alam pada portal TVRI umumnya tersaji dalam teks tidak terstruktur, memuat jenis, lokasi, waktu, dan organisasi terkait sehingga ekstraksi informasi secara manual menjadi kurang efisien. Penelitian ini menerapkan *Named Entity Recognition* (NER) Statistik berbasis web menggunakan algoritma Naive Bayes untuk mengekstraksi entitas bencana, lokasi, waktu, tanggal, dan organisasi dari 200 berita bencana TVRI. Data latih dibuat melalui pelabelan manual dengan skema BIO (*Begin-Inside-Outside*) tervalidasi pakar, menghasilkan 40.667 token berlabel yang dikonversi menjadi fitur CurrentWord, Tipe Token, CurrentTag, Bef1Tag, dan Class, dengan Laplace Smoothing untuk mengatasi *zero probability*. Pengujian dilakukan sebelum dan sesudah penerapan *Random Undersampling* untuk mengetahui pengaruhnya terhadap keseimbangan kelas. Sebelum *Random Undersampling*, model mencapai accuracy 87,91%, precision 55,73%, recall 58,66%, dan F1-score 56,63%, setelah diterapkan *Random Undersampling*, accuracy turun menjadi 83,17% dan recall menjadi 55,50%, sementara precision naik menjadi 62,74% dan F1-score menjadi 57,45%. Hasil ini menunjukkan *Random Undersampling* memperbaiki keseimbangan precision-recall meski accuracy menurun, membuktikan NER Statistik berbasis Naive Bayes mampu mengekstraksi entitas secara otomatis, meski peningkatan F1-score relatif kecil (sekitar 0,82 poin persentase) sehingga diperlukan pengujian lanjutan untuk memastikan konsistensi hasil.

Kata kunci: Berita Bencana Alam, Naive Bayes, *Named Entity Recognition*, *Random Undersampling*, Pelabelan Manual

Abstract

Disaster news published on TVRI's portal is typically presented in unstructured text containing important details such as disaster type, location, time, and related organizations, making manual information extraction inefficient. This study implements a web-based Statistical Named Entity Recognition (NER) using the Naive Bayes algorithm to extract disaster, location, time, date, and organization entities from 200 TVRI disaster news articles. Training data was created through expert-validated manual labeling using the BIO (Begin-Inside-Outside) scheme, producing 40,667 labeled tokens converted into CurrentWord, Token Type, CurrentTag, Bef1Tag, and Class features, with Laplace Smoothing applied to address zero probability. Testing was conducted before and after applying Random Undersampling to assess its effect on class balance. Before Random Undersampling, the model achieved 87.91% accuracy, 55.73% precision, 58.66% recall, and 56.63% F1-score, after application, accuracy dropped to 83.17% and recall to 55.50%, while precision rose to 62.74% and F1-score to 57.45%. These results show that Random Undersampling improved the precision-recall balance despite lower accuracy, proving that Naive Bayes-based Statistical NER can automatically extract entities, though the relatively small F1-score improvement (about 0.82 percentage points) indicates a need for further testing.

***Keywords**: Naive Bayes, Named Entity Recognition, Natural Disaster News, Random Undersampling, Manual Labeling*

1. PENDAHULUAN

Indonesia merupakan negara kepulauan yang berada pada jalur Cincin Api Pasifik dan dikelilingi oleh empat lempeng tektonik sehingga rawan mengalami berbagai bencana alam akibat kondisi geografis dan geologisnya [1]. Berdasarkan data Badan Nasional Penanggulangan Bencana (BNPB), tercatat 3.233 kejadian bencana di Indonesia sepanjang tahun 2025, dengan banjir sebagai jenis bencana paling dominan, diikuti cuaca ekstrem, kebakaran hutan dan lahan, serta tanah longsor [2]. Bencana-bencana tersebut menyebabkan 1.626 korban meninggal dunia dan lebih dari sepuluh juta jiwa terdampak sehingga pengelolaan informasi kebencanaan yang cepat dan akurat menjadi kebutuhan mendesak.

Perkembangan media digital mendorong penyajian berita bencana dalam jumlah besar, salah satunya melalui portal berita daring TVRI yang secara konsisten memuat informasi jenis bencana, lokasi, waktu kejadian, korban, dan instansi terkait. Namun, informasi tersebut masih tersaji dalam bentuk teks tidak terstruktur sehingga sulit dimanfaatkan secara otomatis untuk analisis dan pengambilan keputusan. *Named Entity Recognition* (NER), sebagai salah satu tugas *Natural Language Processing* (NLP), dapat digunakan untuk mengidentifikasi dan mengklasifikasikan entitas penting pada teks berita sehingga proses ekstraksi informasi dapat dilakukan lebih cepat dibandingkan cara manual [3].

Perkembangan NER saat ini banyak mengarah pada pendekatan berbasis *deep learning* dan *transformer* yang menuntut sumber daya komputasi besar serta korpus berlabel dalam jumlah masif [4]. Sebagai gambaran, penelitian NER berbasis transformer pada dokumen hukum berbahasa Indonesia menggunakan IndoBERT dan model sejenis telah menunjukkan performa yang tinggi [5], namun pendekatan tersebut membutuhkan sumber daya komputasi dan data berlabel yang jauh lebih besar dibandingkan pendekatan statistik sederhana. Pendekatan tersebut kurang sesuai diterapkan pada konteks redaksi berita atau penelitian akademik berskala kecil-menengah, yang umumnya memiliki keterbatasan jumlah data berlabel maupun infrastruktur komputasi. Pendekatan statistik berbasis Naive Bayes menjadi alternatif yang relevan karena kesederhanaan komputasi, kemudahan interpretasi, serta kinerja yang tetap kompetitif pada tugas klasifikasi teks meskipun dilatih dengan dataset berukuran terbatas [6]. Meskipun demikian, proses pelabelan data latih untuk NER tetap membutuhkan anotasi manual yang memakan waktu dan berpotensi menimbulkan inkonsistensi antar-anotator apabila tidak divalidasi dengan baik [7].

Penelitian terdahulu telah menerapkan NER pada domain kebencanaan dengan berbagai pendekatan. Penelitian [8] menerapkan Naive Bayes dan C4.5 untuk deteksi informasi spasial-temporal pada teks bencana, penelitian [9] mengekstraksi lokasi dan waktu kejadian banjir dari berita daring menggunakan pendekatan NER, penelitian [1] menerapkan model *deep learning* berbasis BiLSTM untuk NER bencana multi-sumber berbahasa Indonesia, sedangkan penelitian [10] membandingkan kinerja model IndoBERT dan IndoLEM dalam mengekstraksi informasi kesehatan pascabencana dari berita daring. Studi-studi tersebut umumnya berfokus pada peningkatan kinerja melalui model yang kompleks. Namun, studi-studi tersebut belum banyak mengkaji secara khusus bagaimana kondisi ketidakseimbangan kelas (*class imbalance*) pada dataset hasil pelabelan BIO memengaruhi kinerja model statistik sederhana seperti Naive Bayes. Studi-studi tersebut juga belum banyak mengkaji bagaimana teknik penyeimbangan data seperti *Random Undersampling* dapat memengaruhi *trade-off* antara precision dan recall pada tugas NER [11].

Secara lebih spesifik, penelitian ini memiliki tiga perbedaan utama dibandingkan studi-studi terdahulu tersebut. Pertama, dari sisi domain, penelitian ini berfokus pada berita bencana alam TVRI yang belum pernah dikaji pada penelitian NER kebencanaan sebelumnya. Kedua, dari sisi metode, penelitian ini secara khusus mengkaji kombinasi pendekatan statistik sederhana (Naive Bayes) dengan teknik *Random Undersampling*, sementara penelitian terdahulu banyak menggunakan model yang lebih kompleks seperti C4.5, BiLSTM, IndoBERT, dan IndoLEM tanpa mengkaji secara khusus penanganan *class imbalance* pada dataset hasil pelabelan BIO. Ketiga, dari sisi evaluasi, penelitian ini secara eksplisit menganalisis trade-off antara precision dan recall akibat penerapan teknik penyeimbangan data, alih-alih hanya melaporkan skor akhir model.

Berdasarkan uraian tersebut, penelitian ini mengimplementasikan NER Statistik, yaitu pendekatan NER yang mengandalkan model probabilistik sederhana seperti Naive Bayes alih-alih model *deep learning* atau transformer yang lebih kompleks, diimplementasikan berbasis web untuk mengekstraksi entitas jenis bencana, lokasi, waktu, tanggal, dan organisasi pada berita bencana alam TVRI, dengan data latih yang dilabeli secara manual mengacu pada skema BIO. Penelitian ini bertujuan untuk: (1) mengimplementasikan model NER Statistik berbasis Naive Bayes yang mampu mengekstraksi entitas pada berita bencana alam TVRI secara otomatis dan (2) menganalisis pengaruh penerapan *Random Undersampling* terhadap kinerja model, dengan

membandingkan hasil pengujian sebelum dan sesudah penyeimbangan data menggunakan metrik accuracy, precision, recall, dan F1-score.

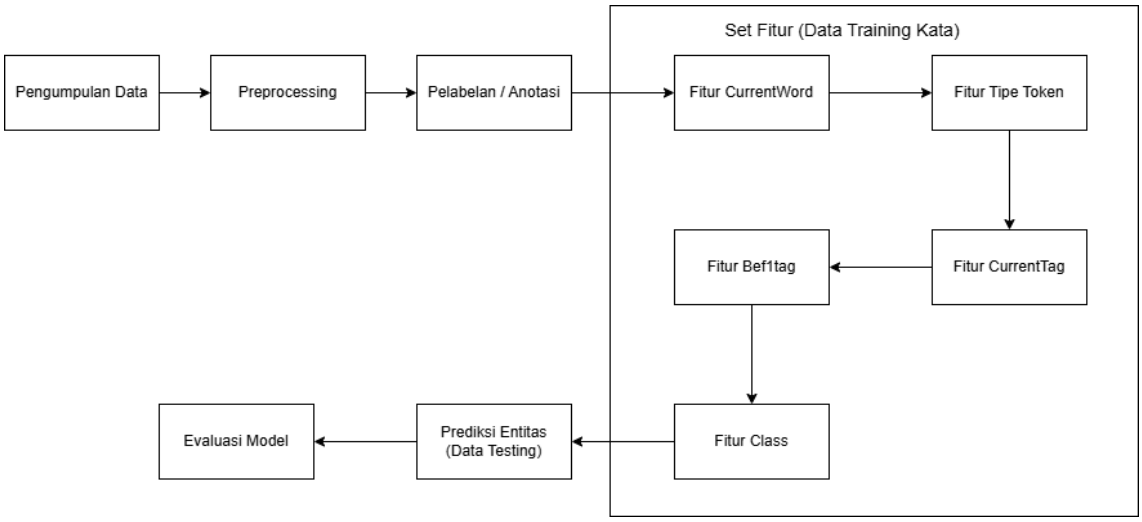
2. METODE PENELITIAN

2.1. Data Penelitian

Dataset penelitian berupa 200 berita bencana alam, terdiri dari 190 berita yang diperoleh dari portal berita daring TVRI (tvrinews.com) menggunakan kata kunci pencarian "bencana", ditambah 10 naskah berita internal yang diperoleh langsung dari TVRI bagian Program dan Berita untuk keperluan penelitian akademik. Penggunaan 10 naskah berita internal tersebut telah memperoleh izin resmi dari pihak TVRI, yang dibuktikan dengan surat keterangan riset yang ditandatangani oleh pihak terkait. Setiap berita memuat informasi tidak terstruktur mengenai jenis bencana (seperti banjir, tanah longsor, dan gempa bumi), lokasi kejadian, waktu kejadian, serta organisasi yang terlibat dalam penanganan bencana, yang menjadi lima kategori entitas target dalam proses ekstraksi, yaitu DISASTER, LOCATION, TIME, DATE, dan ORGANIZATION.

2.2. Tahapan Penelitian

Tahapan penelitian mencakup enam langkah utama sebagaimana ditunjukkan pada Gambar 1, yaitu pengumpulan data, *preprocessing*, pelabelan data latih, pembentukan fitur data training, pelatihan model Naive Bayes, serta prediksi dan evaluasi entitas pada data uji. Setiap tahapan diimplementasikan pada aplikasi berbasis web menggunakan struktur kontrol perulangan (*foreach*) untuk memproses data secara berurutan dan percabangan kondisi untuk menentukan keputusan pada setiap tahap sehingga proses pelabelan, pelatihan, dan prediksi dapat dijalankan secara konsisten dan berulang terhadap seluruh data.



Gambar 1. Diagram Alir Tahapan Penelitian NER Statistik

2.3. *Preprocessing*

Data yang telah dikumpulkan terlebih dahulu melalui lima tahap *preprocessing* yang dijalankan secara berurutan. Case Folding mengubah seluruh huruf pada teks berita menjadi huruf kecil (*lowercase*). *Cleansing* menghapus karakter yang tidak diperlukan seperti tanda baca, simbol, tagar, tautan (URL), dan spasi berlebih. Normalisasi mengubah kata tidak baku, singkatan, dan salah ketik (*typo*) menjadi bentuk baku menggunakan kamus normalisasi, misalnya "Kab" menjadi "Kabupaten", "ga" menjadi "tidak", dan "krn" menjadi "karena". Tokenizing memecah kalimat menjadi satuan kata (token), misalnya kalimat "Banjir melanda Kecamatan Kebayoran Lama" dipecah menjadi token "Banjir", "melanda", "Kecamatan", "Kebayoran", "Lama". Stopword Removal menghapus kata yang tidak informatif seperti "ke", "dari", "adalah", dan "daripada". Tahapan ini penting untuk mengurangi noise pada teks sehingga proses pemberian fitur dan pelabelan entitas dapat dilakukan lebih akurat [12], mengingat pemilihan

teknik *preprocessing* yang tepat berpengaruh signifikan terhadap kinerja model klasifikasi teks [13].

2.4. Pelabelan Data Latih

Pelabelan data latih pada metode NER Statistik dilakukan secara manual oleh penulis segera setelah tahap *preprocessing* selesai, kemudian divalidasi oleh pakar lulusan magister komputer yang memahami konteks penelitian untuk menjaga konsistensi label. Setiap token diberi label menggunakan skema BIO (*Begin, Inside, Outside*): label B (*Begin*) diberikan pada token pertama suatu entitas, label I (*Inside*) diberikan pada token lanjutan dari entitas yang sama, sedangkan label O (*Outside*) diberikan pada token yang bukan bagian dari entitas mana pun. Mekanisme label I penting diterapkan karena banyak entitas pada berita bencana alam terdiri lebih dari satu kata, seperti "Gempa Bumi", "Tanah Longsor", dan "Kebakaran Hutan" sehingga tanpa mekanisme ini sistem dapat keliru memisahkan satu entitas menjadi dua token yang berdiri sendiri. Contoh hasil BIO tagging pada kalimat "gempa bumi terjadi di bali pada senin 26 mei 2026" ditunjukkan pada Tabel 1. Berdasarkan hasil implementasi, total data yang berhasil dilabeli secara manual berjumlah 40.667 kata yang selanjutnya digunakan sebagai *ground truth* untuk melatih model Naive Bayes. Gambar 2 menggambarkan seluruh proses pelabelan dilakukan pada aplikasi berbasis web yang telah dibangun.

Tabel 1. Contoh Hasil BIO *Tagging*

Token	Label BIO
gempa	B-DISASTER
bumi	I-DISASTER
terjadi	O
di	O
bali	B-LOCATION
pada	O
senin	B-TIME
26	B-DATE
mei	I-DATE
2026	I-DATE

Gambar 2. Tampilan Layar Pelabelan Manual

2.5. Pembentukan Fitur Data *Training*

Dataset berlabel BIO selanjutnya dikonversi menjadi struktur tabular yang dapat diproses langsung oleh algoritma Naive Bayes. Setiap baris merepresentasikan satu token beserta lima atribut fitur, yaitu CurrentWord (kata hasil *preprocessing*), Tipe Token (WORD untuk kata atau

NUM untuk angka), CurrentTag (kategori *part-of-speech* token saat ini), Bef1Tag (kategori *part-of-speech* token sebelumnya, diberi penanda BOS/*Beginning of Sequence* untuk token pertama kalimat), dan Class (label BIO hasil pelabelan manual sebagai target pembelajaran). Total data latih yang terbentuk dari keseluruhan dataset berjumlah 40.667 baris, yang selanjutnya dibagi menjadi data latih dan data uji dengan rasio 80:20. Pembagian data latih dan data uji pada penelitian ini dilakukan secara acak pada level token (baris data), bukan pada level dokumen berita. Pendekatan ini perlu digarisbawahi sebagai keterbatasan penelitian karena berpotensi menimbulkan *data leakage*, yaitu token dari satu dokumen berita yang sama dapat terpisah ke dalam data latih maupun data uji sehingga model berpeluang "mengenali" konteks dokumen yang sama saat pengujian, yang dapat membuat skor evaluasi sedikit lebih optimistis dibandingkan kondisi generalisasi pada dokumen yang benar-benar baru. Penelitian selanjutnya disarankan menerapkan pembagian data pada level dokumen agar risiko ini dapat dihindari. Contoh data latih dan data uji masing-masing dapat dilihat pada Tabel 2 dan Tabel 3 berikut ini.

Tabel 2. Contoh Data Latih

CurrentWord	Tipe Token	CurrentTag	Bef1Tag	Class
gempa	WORD	NOUN	BOS	B-DISASTER
bumi	WORD	OOV	NOUN	I-DISASTER
terjadi	WORD	VPAS	OOV	O
di	WORD	PREP	VPAS	O
bali	WORD	OOV	PREP	B-LOCATION
pada	WORD	PREP	OOV	O
senin	WORD	OOV	PREP	B-TIME
26	NUM	OOV	OOV	B-DATE
mei	WORD	OOV	OOV	I-DATE
2026	NUM	OOV	OOV	I-DATE

Tabel 3. Contoh Data Uji

CurrentWord	Tipe Token	CurrentTag	Bef1Tag	Class
gempa	WORD	NOUN	BOS	?

2.6. Model Klasifikasi Naive Bayes

Model Naive Bayes digunakan untuk memprediksi kelas entitas pada data uji berdasarkan probabilitas kemunculan fitur pada data latih [6]. Dasar perhitungan model ini mengacu pada Teorema Bayes:

$$P(C|X) = \frac{P(X|C) \times P(C)}{P(X)} \quad (1)$$

dengan P(C|X) adalah probabilitas posterior kelas C, P(X|C) adalah likelihood fitur X pada kelas C, P(C) adalah probabilitas prior kelas C, dan P(X) adalah probabilitas kemunculan fitur X. Karena setiap token direpresentasikan oleh beberapa fitur yang diasumsikan saling bebas, persamaan Naive Bayes dituliskan sebagai:

$$P(C|X) = P(C) \times \prod_i P(x_i|C), i = 1..n \quad (2)$$

dengan x_i adalah fitur ke-i dan n adalah jumlah fitur yang digunakan. Kelas dengan nilai probabilitas posterior tertinggi dipilih sebagai hasil prediksi [14]. Algoritma ini telah banyak diterapkan pada berbagai tugas klasifikasi teks berbahasa Indonesia, termasuk klasifikasi opini publik pada media sosial, karena kesederhanaan komputasi dan kinerjanya yang kompetitif pada data bertipe teks [15]. Untuk mengatasi kemungkinan probabilitas nol (*zero probability*) akibat

kombinasi fitur pada data uji yang tidak muncul pada data latih, diterapkan teknik Laplace Smoothing:

$$P(x_i|C) = \frac{Count(x_i,C) + 1}{\Sigma Count(C) + |V|} \quad (3)$$

dengan Count(x_i,C) adalah frekuensi fitur x_i pada kelas C, ΣCount(C) adalah total fitur pada kelas C, dan |V| adalah jumlah nilai unik pada fitur tersebut di seluruh data latih.

a. Contoh Perhitungan Manual

Untuk memperjelas mekanisme perhitungan, berikut disajikan contoh perhitungan sederhana menggunakan subset ilustratif 10 token data latih pada Tabel 2 (bukan keseluruhan dataset penelitian yang berjumlah 40.667 token) untuk memprediksi kelas token "gempa" (Tipe Token = WORD, CurrentTag = NOUN, Bef1tag = BOS) pada data uji. Nilai probabilitas prior tiap kelas dihitung dari proporsi kemunculannya pada 10 data latih tersebut: B-DISASTER, I-DISASTER, B-LOCATION, B-TIME, dan B-DATE masing-masing muncul 1 kali (1/10), I-DATE muncul 2 kali (2/10), dan O muncul 3 kali (3/10). Selanjutnya, nilai likelihood keempat fitur terhadap masing-masing kelas dihitung menggunakan Persamaan (3), dengan jumlah nilai unik |V| berturut-turut 10 untuk CurrentWord, 2 untuk Tipe Token, 4 untuk CurrentTag, dan 5 untuk Bef1tag. Rekapitulasi nilai prior dan likelihood untuk setiap kelas disajikan pada Tabel 4.

Tabel 4. Probabilitas Prior dan Likelihood Setiap Kelas untuk Token "gempa"

Class	Prior	P(gempa|C)	P(WORD|C)	P(NOUN|C)	P(BOS|C)
B-DISASTER	1/10	2/11	2/3	2/5	2/6
I-DISASTER	1/10	1/11	2/3	1/5	1/6
B-LOCATION	1/10	1/11	2/3	1/5	1/6
B-TIME	1/10	1/11	2/3	1/5	1/6
B-DATE	1/10	1/11	1/3	1/5	1/6
I-DATE	2/10	1/12	2/4	1/6	1/7
O	3/10	1/13	4/5	1/7	1/8

Nilai posterior setiap kelas kemudian diperoleh dengan mengalikan nilai prior dengan seluruh nilai likelihood pada Tabel 4 sesuai Persamaan (2). Hasil akhir perkalian tersebut disajikan pada Tabel 5.

Tabel 5. Hasil Akhir Nilai Posterior

Class	Posterior	Hasil Desimal
B-DISASTER	1/10 × 2/11 × 2/3 × 2/5 × 2/6	0,001616
I-DISASTER	1/10 × 1/11 × 2/3 × 1/5 × 1/6	0,000202
B-LOCATION	1/10 × 1/11 × 2/3 × 1/5 × 1/6	0,000202
B-TIME	1/10 × 1/11 × 2/3 × 1/5 × 1/6	0,000202
B-DATE	1/10 × 1/11 × 1/3 × 1/5 × 1/6	0,000101
I-DATE	2/10 × 1/12 × 2/4 × 1/6 × 1/7	0,000198
O	3/10 × 1/13 × 4/5 × 1/7 × 1/8	0,000330

Berdasarkan Tabel 5, nilai posterior tertinggi diperoleh kelas B-DISASTER dengan nilai probabilitas 0,001616 sehingga token "gempa" diprediksi berlabel B-DISASTER, sesuai dengan label sebenarnya pada Tabel 1. Penerapan Laplace Smoothing memastikan tidak ada kelas yang bernilai probabilitas nol sehingga teknik ini tetap dapat digunakan pada data uji yang mengandung kombinasi fitur yang belum pernah muncul pada data latih. Mekanisme perhitungan inilah yang diterapkan pada skala penuh terhadap seluruh 40.667 data latih hasil pelabelan manual untuk menghasilkan model NER Statistik yang dievaluasi pada bagian Hasil dan Pembahasan.

2.7. Penyeimbangan Data

Dataset hasil pelabelan BIO umumnya didominasi oleh kelas "O" (bukan entitas) sehingga model cenderung bias terhadap kelas mayoritas tersebut [16]. Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan teknik *Random Undersampling* dengan rasio 1:1 antara kelas "O" dan kelas entitas, yang secara acak mengurangi jumlah sampel kelas mayoritas tanpa mengubah data kelas minoritas [11] sehingga proporsi data latih menjadi lebih seimbang dan performa model dapat dievaluasi secara lebih representatif terhadap kemampuannya mengenali seluruh kategori entitas [17].

2.8. Rancangan Pengujian

Pengujian dilakukan dengan membandingkan hasil prediksi model terhadap label sebenarnya pada data uji menggunakan Confusion Matrix, yang menghasilkan komponen True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Berdasarkan komponen tersebut, dihitung empat metrik evaluasi standar, yaitu accuracy, precision, recall, dan F1-score:

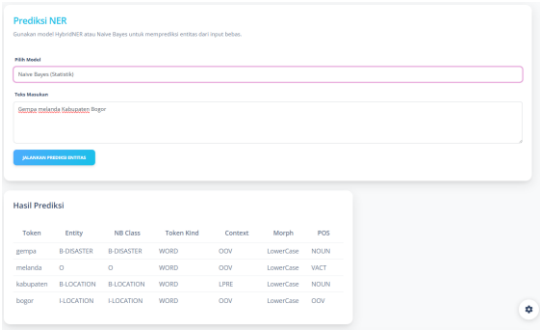
$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

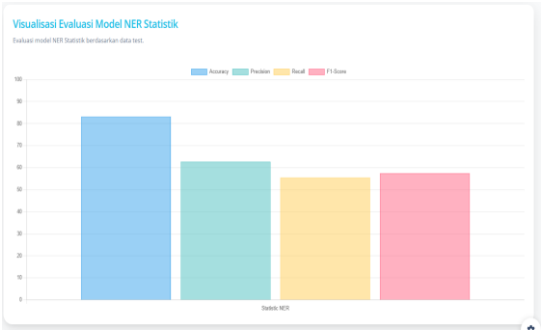
$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

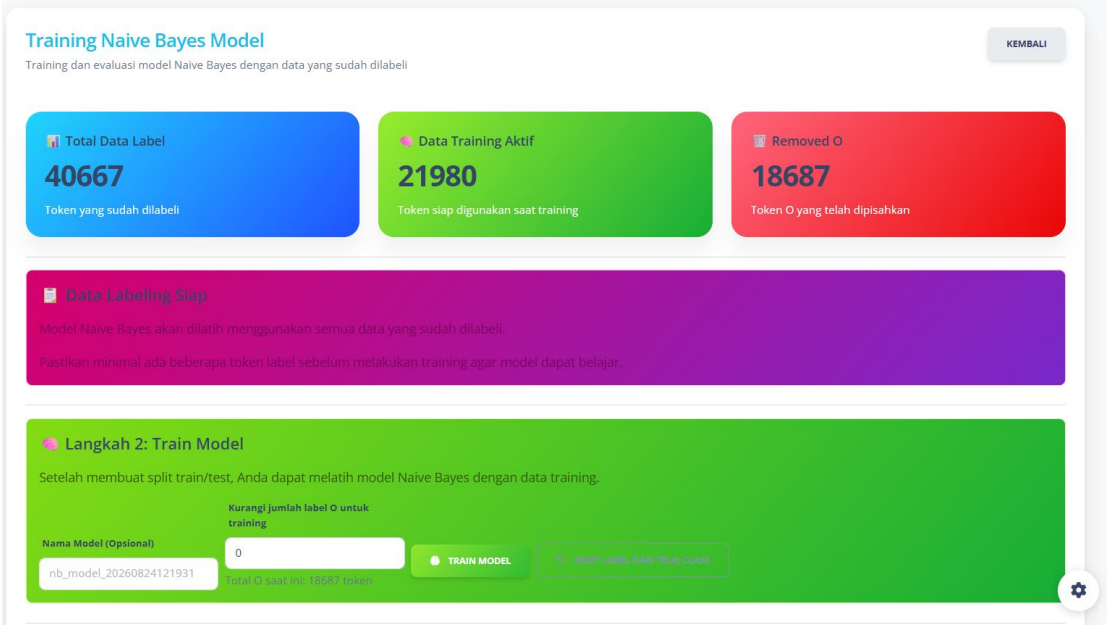
Pengujian dilakukan dua kali pada dataset uji yang sama, yaitu sebelum dan sesudah penerapan *Random Undersampling*, untuk mengetahui pengaruh penyeimbangan data terhadap performa model NER Statistik. Sistem diimplementasikan sebagai aplikasi berbasis web menggunakan bahasa pemrograman PHP dengan framework Laravel dan basis data MySQL sehingga proses pelabelan, pembentukan fitur, pelatihan model, prediksi entitas, dan visualisasi evaluasi dapat diakses melalui aplikasi yang sama. Tampilan layar prediksi entitas, visualisasi evaluasi, dan training model ditunjukkan pada pada Gambar 3, Gambar 4, dan Gambar 5.



Gambar 3. Tampilan Layar Prediksi Entitas



Gambar 4. Tampilan Layar Visualisasi Evaluasi



Gambar 5. Tampilan Layar Training Model

3. HASIL DAN PEMBAHASAN

Pengujian dilakukan terhadap model NER Statistik pada data uji yang sama, baik sebelum maupun sesudah penerapan *Random Undersampling*. Tabel 6 dan Tabel 7 menunjukkan hasil pengujian menggunakan metrik accuracy, precision, recall, dan F1-score.

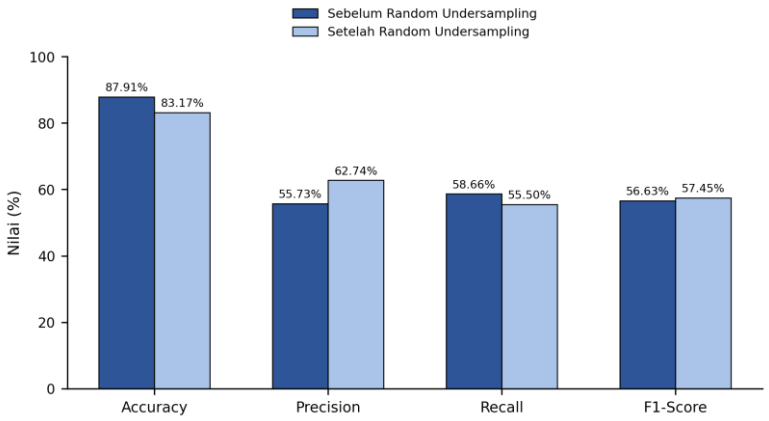
Tabel 6. Skor Pengujian Model Sebelum *Random Undersampling*

Metrik	Skor
Accuracy	87,91%
Precision	55,73%
Recall	58,66%
F1-score	56,63%

Tabel 7. Skor Pengujian Model Setelah *Random Undersampling*

Metrik	Skor
Accuracy	83,17%
Precision	62,74%
Recall	55,50%
F1-score	57,45%

Berdasarkan Tabel 6 dan Tabel 7 serta Gambar 6, penerapan *Random Undersampling* menyebabkan accuracy model menurun dari 87,91% menjadi 83,17%. Penurunan ini sejalan dengan karakteristik class imbalance, di mana nilai accuracy pada data yang timpang cenderung terinflasi karena model hanya perlu memprediksi kelas mayoritas "O" secara benar untuk memperoleh skor tinggi, tanpa benar-benar akurat mengenali kelas-kelas entitas yang jumlahnya jauh lebih sedikit [16]. Ketika proporsi kelas "O" dikurangi melalui *Random Undersampling*, accuracy yang dilaporkan menjadi representasi yang lebih jujur terhadap kemampuan model dalam mengenali seluruh kategori entitas, bukan sekadar kelas mayoritas.



Gambar 6. Grafik Perbandingan Skor Pengujian Sebelum dan Sesudah *Random Undersampling*

Sejalan dengan penurunan accuracy tersebut, precision model justru meningkat dari 55,73% menjadi 62,74%. Hal ini dapat dijelaskan melalui prinsip Teorema Bayes pada Persamaan (1), di mana nilai prior suatu kelas dihitung berdasarkan proporsi kemunculannya pada data latih, ketika jumlah data kelas "O" dikurangi, nilai prior kelas tersebut tidak lagi mendominasi secara ekstrem sehingga model memiliki peluang lebih seimbang dalam mempertimbangkan kelas-kelas entitas saat menghitung probabilitas posterior, dan pada akhirnya mengurangi kecenderungan model memberi label "O" secara berlebihan pada token yang sebenarnya merupakan entitas.

Di sisi lain, recall model menurun dari 58,66% menjadi 55,50% setelah *Random Undersampling*. Penurunan ini kemungkinan terjadi karena *Random Undersampling* turut menghilangkan sebagian token berlabel "O" yang berperan sebagai konteks di sekitar entitas, misalnya pada fitur Bef1tag yang bergantung pada tag token sebelumnya, berkurangnya keragaman konteks tersebut dapat mengurangi kemampuan model dalam mengenali variasi kombinasi fitur pada data uji, sejalan dengan temuan bahwa teknik resampling data, meskipun efektif menyeimbangkan proporsi kelas, berpotensi membuang sebagian informasi kontekstual yang sebenarnya bermanfaat bagi generalisasi model [11]. Kondisi ini menggambarkan precision-recall trade-off, di mana upaya menyeimbangkan proporsi kelas untuk meningkatkan ketepatan prediksi (precision) dapat sedikit mengorbankan cakupan pengenalan entitas (recall) [17].

Nilai F1-score meningkat dari 56,63% menjadi 57,45% setelah *Random Undersampling*, menunjukkan bahwa keseimbangan antara precision dan recall pada model membaik meskipun accuracy menurun. Hasil ini menegaskan bahwa accuracy saja tidak cukup untuk merepresentasikan kinerja model NER pada dataset dengan distribusi kelas yang timpang, dan bahwa metrik F1-score yang mempertimbangkan precision maupun recall secara bersamaan memberikan gambaran yang lebih komprehensif mengenai kinerja model [18]. Temuan ini sejalan dengan penelitian sejenis yang menunjukkan bahwa strategi pelabelan dan penyeimbangan data latih turut memengaruhi kinerja akhir model NER secara signifikan, baik pada pendekatan statistik sederhana seperti Naive Bayes maupun pada model yang lebih kompleks [19],[20].

Secara keseluruhan, hasil pengujian menunjukkan bahwa model NER Statistik berbasis Naive Bayes mampu mengekstraksi entitas jenis bencana, lokasi, waktu, tanggal, dan organisasi pada berita bencana alam TVRI, serta bahwa penerapan *Random Undersampling* memperbaiki keseimbangan antara precision dan recall meskipun berdampak pada penurunan accuracy dan recall. Perlu dicatat bahwa peningkatan F1-score setelah *Random Undersampling* tergolong kecil (56,63% menjadi 57,45%, atau sekitar 0,82 poin persentase). Karena *Random Undersampling* merupakan prosedur stokastik, selisih sekecil ini berpotensi berada dalam rentang variasi acak sehingga temuan ini perlu dimaknai sebagai indikasi awal, bukan bukti definitif, dan sebaiknya divalidasi melalui pengulangan eksperimen dengan beberapa nilai random seed berbeda pada penelitian selanjutnya. Pemilihan penerapan *Random Undersampling* pada implementasi praktis perlu mempertimbangkan prioritas evaluasi yang diinginkan, yaitu apakah menitikberatkan pada accuracy secara keseluruhan atau pada ketepatan dan keseimbangan prediksi entitas.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa model NER Statistik berbasis algoritma Naive Bayes berhasil diimplementasikan pada aplikasi berbasis web untuk mengekstraksi entitas jenis bencana, lokasi, waktu, tanggal, dan organisasi pada berita bencana alam TVRI, dengan data latih yang dilabeli secara manual mengacu pada skema BIO dan divalidasi oleh pakar. Penerapan Laplace Smoothing terbukti mampu mengatasi permasalahan *zero probability* pada kombinasi fitur yang tidak muncul pada data latih. Hasil pengujian menunjukkan bahwa sebelum *Random Undersampling*, model memperoleh accuracy 87,91%, precision 55,73%, recall 58,66%, dan F1-score 56,63%. Setelah *Random Undersampling* diterapkan, accuracy menurun menjadi 83,17% dan recall menurun menjadi 55,50%, namun precision meningkat menjadi 62,74% dan F1-score meningkat menjadi 57,45%, yang menunjukkan bahwa *Random Undersampling* mampu memperbaiki keseimbangan antara precision dan recall meskipun menurunkan accuracy, dengan besar peningkatan F1-score yang relatif kecil, yaitu sekitar 0,82 poin persentase.

Beberapa saran untuk pengembangan penelitian selanjutnya, antara lain: (1) nilai recall model, khususnya setelah *Random Undersampling*, masih tergolong belum optimal sehingga perlu dikaji teknik penyeimbangan data lain seperti SMOTE atau kombinasi *oversampling-undersampling* untuk mengurangi hilangnya informasi kontekstual pada kelas mayoritas, (2) rasio *Random Undersampling* dapat dieksplorasi lebih lanjut, misalnya rasio 2:1 atau 3:1, untuk menemukan titik optimal antara accuracy dan keseimbangan precision-recall, (3) dataset penelitian disarankan diperluas, baik dari sisi jumlah maupun sumber berita, agar proses pelatihan model lebih representatif dan hasil pengujian lebih dapat digeneralisasi, dan (4) penelitian selanjutnya dapat membandingkan kinerja model NER Statistik berbasis Naive Bayes dengan pendekatan pelabelan atau algoritma klasifikasi lain untuk memperkaya evaluasi metode ekstraksi entitas pada domain berita bencana alam. Selain itu, (5) penelitian selanjutnya disarankan menerapkan pembagian data latih dan data uji pada level dokumen berita (bukan level token) serta mendokumentasikan random seed yang digunakan sehingga potensi data leakage dapat dihindari dan proses eksperimen dapat direplikasi secara konsisten.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada dosen pembimbing atas segala arahan, masukan, dan bimbingan yang diberikan selama penelitian ini berlangsung. Terima kasih pula penulis sampaikan kepada pihak TVRI dan kampus Universitas Budi Luhur yang telah menyediakan fasilitas pendukung, serta kepada keluarga dan rekan-rekan atas dukungan moral dan motivasi yang terus mengalir. Keberhasilan penyelesaian penelitian ini tidak lepas dari kontribusi dan bantuan semua pihak baik yang telah disebutkan maupun tidak disebutkan.

DAFTAR PUSTAKA

[1] G. Fajar *et al.*, “Journal of Open Innovation : Technology , Market , and Complexity Indonesian disaster named entity recognition from multi source information using bidirectional LSTM (BiLSTM),” *J. Open Innov. Technol. Mark. Complex.*, vol. 10, no. 3, p. 100358, 2024, doi: 10.1016/j.joitmc.2024.100358.

[2] Badan Nasional Penanggulangan Bencana (BNPB), “Indonesia Tangguh Menghadapi Bencana 2025.” [Online]. Available: https://indonesiabaik.id/infografis/indonesia-tangguh-menghadapi-bencana-2025

[3] I. Budi and R. R. Suryono, “Application of named entity recognition method for Indonesian datasets: a review,” *Bull. Electr. Eng. Informatics*, vol. 12, no. 2, pp. 969–978, 2023, doi: 10.11591/eei.v12i2.4529.

[4] Z. Hu, W. Hou, and X. Liu, “Deep learning for named entity recognition: a survey,” *Neural Comput. Appl.*, vol. 36, no. 16, pp. 8995–9022, 2024, doi: 10.1007/s00521-024-09646-6.

[5] E. Yulianti, *et al*., “Named entity recognition on Indonesian legal documents: a dataset and

study using transformer-based models," *Int. J. Electr. Comput. Eng.*, vol. 14, no. 5, pp. 5489–5501, 2024, doi: 10.11591/ijece.v14i5.pp5489-5501.

[6] P. J. B. Pajila, *et al*., "A Comprehensive Survey on Naive Bayes Algorithm: Advantages, Limitations and Applications," *Proc. 4th Int. Conf. Smart Electron. Commun. ICOSEC 2023*, no. Icosec, pp. 1228–1234, 2023, doi: 10.1109/ICOSEC58147.2023.10276274.

[7] S. O. Khairunnisa, Z. Chen, and M. Komachi, "Improving Domain-Specific NE*R* in the Indonesian Language Through Domain Transfer and Data Augmentation*," J. Adv. Comput. Intell. Intell. Informatics*, vol. 28, no. 6, pp. 1299–1312, 2024, doi: 10.20965/jaciii.2024.p1299.

[8] I. Indra, *et al*., "Implementation of Named Entity Recognition (NER) for Spatio-Temporal Detection and Sentence Modeling of Natural Disasters Using Naive Bayes and C.45," in *2024 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, 2024, pp. 166–171. doi: 10.1109/ISITIA63062.2024.10668211.

[9] I. M. Karo Karo, S. Dewi, and A. Syahrin, "Ekstraksi Informasi Bencana Banjir Dari Berita Online Berbasis Named Enitity Recognition," *Multinetics*, vol. 11, no. 1, pp. 35–42, 2025, doi: 10.32722/multinetics.v11i1.7499.

[10] N. Istiqomah and F. Novika, "Perbandingan Kinerja Model NER IndoBERT dan IndoLEM dalam Ekstraksi Informasi Kesehatan Pascabencana dari Berita Daring di Indonesia," *J. Comput. Sci. Informatics Eng.*, vol. 04, no. 3, pp. 158–174, 2025, doi: 10.55537/cosie.v4i3.1174

[11] M. Carvalho, A. J. Pinho, and S. Brás, "Resampling approaches to handle class imbalance: a review from a data perspective," *J. Big Data*, vol. 12, no. 1, p. 71, Mar. 2025, doi: 10.1186/s40537-025-01119-4.

[12] C. P. Chai, "Comparison of text preprocessing methods," *Nat. Lang. Eng.*, vol. 29, no. 3, pp. 509–553, 2023, doi: 10.1017/S1351324922000213.

[13] M. Siino, I. Tinnirello, and M. La Cascia, "Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers," *Inf. Syst.*, vol. 121, no. March 2023, p. 102342, 2024, doi: 10.1016/j.is.2023.102342.

[14] N. Lestari *et al.*, "Implementation Of Text Mining And Pattern Discovery With Naive Bayes Algorithm For Classification Of Text Documents," *J. Teknol. Inf. dan Komun.*, vol. 14, no. 1, pp. 88–102, 2023.

[15] R. Hidayat and W. Windarto, "Klasifikasi Komentar Netizen X Tentang Pemecatan Shin Tae Yong Dari PSSI Dengan Algoritma Naive Bayes," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 9, no. 1, pp. 171–181, 2026, doi: 10.36080/skanika.v9i1.3605.

[16] Y. Feng, M. Zhou, and X. Tong, "Imbalanced classification: A paradigm-based review," *Stat. Anal. Data Min. ASA Data Sci. J.*, vol. 14, no. 5, pp. 383–406, Oct. 2021, doi: 10.1002/sam.11538.

[17] K. Ghosh, C. Bellinger, R. Corizzo, P. Branco, B. Krawczyk, and N. Japkowicz, "The class imbalance problem in deep learning," *Mach. Learn.*, vol. 113, no. 7, pp. 4845–4901, 2024, doi: 10.1007/s10994-022-06268-8.

[18] H. Alamro, T. Gojobori, M. Essack, and X. Gao, "BioBBC: a multi-feature model that enhances the detection of biomedical entities," *Sci. Rep.*, vol. 14, no. 1, pp. 1–14, 2024, doi: 10.1038/s41598-024-58334-x.

[19] I. M. Widi, A. Ari, and I. W. Supriana, "Dampak Penggunaan Anotasi Penamaan yang Berbeda Pada Kinerja NER," *J. Nas. Teknol. Inf. dan Apl.*, vol. 1, no. 4, pp. 1141–1148, 2023.

[20] G. Syahrani, S. Sevira, and A. Yunizar Pratama Yusuf, "Rancangan Chatbot Rekomendasi Coffee Shop Jabodetabek Dengan Menggunakan Dialogflow Natural Language Processing," *SKANIKA: Sistem Komputer dan Teknik Informormatika*, vol. 7, no. 1, pp. 74–84, 2024, doi: 10.36080/skanika.v7i1.3139.