

IMPLEMENTATION OF THE PSO-SMOTE METHOD ON THE NAIVE BAYES ALGORITHM TO ADDRESS CLASS IMBALANCE IN LANDSLIDE DISASTER DATA

PENERAPAN METODE PSO-SMOTE PADA ALGORITMA NAIVE BAYES UNTUK MENGATASI CLASS IMBALANCE DATA BENCANA TANAH LONGSOR

Azwar Damari¹, Taghfirul Azhima Yoga Siswa², Wawan Joko Pranoto³
Universitas Muhammadiyah Kalimantan Timur, Jl. Juanda, No. 15, Samarinda, Kalimantan Timur
email: 20111102441023@umkt.ac.id¹, tay758@umkt.ac.id^{2}, wjp337@umkt.ac.id³*

Abstract - Landslides in Samarinda, which often occur after floods, pose a threat to settlements, infrastructure, and the agricultural sector. This study proposes a combination of Naïve Bayes, SMOTE (Synthetic Minority Oversampling Technique), and PSO (Particle Swarm Optimization) to address class imbalance in landslide prediction. The results show that while PSO successfully improves the accuracy of the Naïve Bayes model, the application of SMOTE led to a decrease in accuracy for some method combinations. This decrease is due to changes in data distribution caused by synthetic data, which can introduce noise and affect feature selection and model optimisation. However, the combination of Naïve Bayes with PSO optimisation resulted in a modest accuracy improvement (+0.48%). These findings suggest that SMOTE should be used cautiously, while PSO is more effective in enhancing the accuracy of the landslide prediction model. The implications for practical application are that although SMOTE and PSO can improve accuracy, the impact of synthetic data on data distribution must be considered, and further testing is needed to ensure its effectiveness in real-world conditions.

Keywords - Landslide, Class Imbalance, Naive Bayes, SMOTE, PSO, Data Mining.

Abstrak - Bencana tanah longsor di Kota Samarinda, yang sering terjadi setelah banjir, mengancam pemukiman, infrastruktur, dan sektor pertanian. Penelitian ini mengusulkan penerapan metode kombinasi Naïve Bayes, SMOTE (Synthetic Minority Oversampling Technique), dan PSO (Particle Swarm Optimization). Hasil penelitian menunjukkan bahwa meskipun PSO berhasil meningkatkan akurasi model Naïve Bayes, penerapan SMOTE justru menyebabkan penurunan akurasi pada beberapa kombinasi metode. Penurunan ini disebabkan oleh perubahan distribusi data akibat data sintetis yang dapat memperkenalkan noise, mempengaruhi seleksi fitur dan optimasi model. Meskipun demikian, kombinasi Naïve Bayes dengan PSO Optimasi menghasilkan peningkatan akurasi meskipun terbatas (+0.48%). Temuan ini menunjukkan bahwa SMOTE perlu digunakan dengan hati-hati, sementara PSO lebih efektif dalam meningkatkan akurasi model prediksi tanah longsor. Implikasinya untuk praktik nyata adalah, meskipun teknik SMOTE dan PSO dapat meningkatkan akurasi, pengaruh data sintetis terhadap distribusi data harus diperhitungkan, dan perlu pengujian lebih lanjut untuk memastikan efektivitasnya dalam kondisi dunia nyata.

Kata Kunci - Tanah Longsor, Class Imbalance, Naive Bayes, SMOTE, PSO, DATA MINING.

I. PENDAHULUAN

Bencana alam merupakan salah satu kejadian bencana yang sangat berbahaya dan dapat terjadi pada setiap titik lokasi di muka bumi. Jenis bencana alam sangat bermacam-macam, seperti gerakan tanah, gempa bumi, tsunami, tanah longsor, dan lain-lain. Setiap jenis kejadian bencana alam tersebut mempunyai bahaya dan dampak yang berbeda-beda bergantung pada penyebab terjadinya bencana alam tersebut sesuai [1]. Banyaknya bentuk bencana alam terdapat bencana alam tanah longsor, tanah longsor itu sendiri adalah perpindahan material pembentuk lereng berupa batuan, bahan rombakan, tanah, atau material campuran tersebut, bergerak ke bawah atau ke luar lereng[2]. Bencana longsor adalah satu dari beberapa jenis bencana yang terjadi di Indonesia. Akibatnya, terjadi pendangkalan, rusaknya lahan pertanian dan pemukiman serta terganggunya jalur lalu lintas [3]. Frekuensi dan intensitas kejadian tanah longsor terus meningkat akibat perubahan iklim serta aktivitas manusia yang memodifikasi lanskap alami. Penelitian menunjukkan bahwa peristiwa hidro-geomorfologis, termasuk tanah longsor, kini terjadi dengan frekuensi dan magnitudo yang lebih tinggi di berbagai wilayah dunia, termasuk kawasan perkotaan [4]. Tanah longsor merupakan bencana kedua yang paling sering terjadi di Kota Samarinda setelah banjir [5].

Upaya untuk memprediksi dan mencegah dampak tanah longsor telah dilakukan melalui berbagai teknik, salah satunya adalah penggunaan data mining. Data mining memungkinkan identifikasi pola tersembunyi dalam dataset besar, yang kemudian dapat digunakan untuk klasifikasi dan pengelolaan risiko tanah longsor [6]. Salah satu tantangan utama dalam menganalisis data bencana tanah longsor adalah masalah ketidakseimbangan kelas. Ketidakseimbangan kelas terjadi ketika jumlah data yang mewakili kejadian tanah longsor (kelas minoritas) jauh lebih kecil dibandingkan dengan data yang tidak menggambarkan kejadian tersebut[7]. Ketidakseimbangan ini dapat membuat model prediktif yang dihasilkan menjadi bias, sehingga menurunkan akurasi dan efektivitas dalam mendeteksi kejadian. Penelitian mengungkapkan bahwa model yang dilatih pada dataset tidak seimbang cenderung kesulitan mengenali pola dari kelas minoritas, yang akhirnya dapat menyebabkan tingginya jumlah prediksi false negative[8]. Untuk mengatasi ketidakseimbangan data maka algoritma yang paling umum digunakan adalah metode *oversampling*. Teknik *random oversampling* adalah metode *oversampling* yang paling sederhana. namun, dapat dengan mudah menghasilkan *overfitting*. Oleh karena itu, algoritma *synthetic minority oversampling (SMOTE)* menggunakan data asli untuk mensintesis data minoritas baru yang berbeda dari aslinya, sehingga mengurangi dampak *overfitting* [9]. SMOTE merupakan salah satu metode *oversampling* untuk mensintesis dataset minoritas hingga jumlah dataset pada kelas mayoritas seimbang. Penerapan *oversampling SMOTE* mampu mengatasi ketidakseimbangan kelas pada *dataset* [10]. Selain itu, optimalisasi parameter juga penting untuk meningkatkan akurasi model klasifikasi. Salah satu algoritma optimasi yang efektif adalah Particle Swarm Optimization (PSO) [11]. PSO (Particle Swarm Optimization) adalah teknik optimasi berbasis populasi yang terinspirasi oleh perilaku sosial kelompok burung atau ikan. PSO dapat diterapkan untuk menemukan parameter terbaik dalam algoritma klasifikasi, sehingga dapat meningkatkan akurasi prediksi [12]. Pada penelitian ini, metode PSO-SMOTE diterapkan pada algoritma Naive Bayes untuk mengatasi masalah *class imbalance* pada data bencana tanah longsor. Penerapan gabungan PSO dan SMOTE dalam algoritma Naive Bayes merupakan pendekatan yang masih relatif jarang diterapkan dalam analisis data bencana tanah longsor, khususnya dalam konteks mitigasi bencana. Naive Bayes adalah metode klasifikasi yang sederhana namun efektif, yang menggunakan Teorema Bayes untuk menghitung probabilitas dan memprediksi kemungkinan berdasarkan data sebelumnya [13]. Dengan mengoptimalkan parameter Naive Bayes menggunakan PSO, penelitian ini memberikan kontribusi dalam akurasi deteksi tanah longsor yang lebih tepat dan efisien, meskipun data yang digunakan tidak seimbang.

II. SIGNIFIKASI STUDI

No	Penulis	Penelitian Terdahulu
1	Siti Mutmainah [14]	Hasil penelitian didapatkan pada penggunaan teknik <i>random oversampling</i> mendapat performa yang lebih tinggi yaitu 95% daripada teknik <i>random undersampling</i> yang mendapat performa 76%
2	Dhifa Zhafirah[15]	Penelitian ini menunjukkan bahwa penerapan ROS dapat menangani <i>imbalance</i> data, menghasilkan akurasi 96,43% (klasifikasi <i>excellent</i>) dengan model <i>C-Classification</i> , <i>kernel Radial Basis Function</i> (RBF), serta parameter sigma 1 dan <i>cost</i> 1 pada skema data latih 90% dan data uji 10%. Akurasi ini meningkat signifikan dibandingkan tanpa ROS, yang hanya mencapai 76%.
3	Raenald Syaputra, Taghfirul Azhima Yoga Siswa, Wawan Joko Pranoto [16]	Penelitian ini menemukan bahwa fitur penting untuk klasifikasi banjir adalah temperatur maksimum, kecepatan angin maksimum, arah angin maksimum, arah angin terbanyak, lamanya penyinaran matahari, dan kecepatan angin rata-rata. Dengan kombinasi SVM, SMOTE, GA, dan PSO, akurasi mencapai 82,28%. Namun, terdapat tantangan terkait kontradiksi hasil dengan penelitian lain dan variasi akibat karakteristik dataset. Hasil ini dapat digunakan oleh pemerintah dan badan penanggulangan bencana Kota Samarinda untuk memprediksi dan mencegah banjir secara lebih efektif.

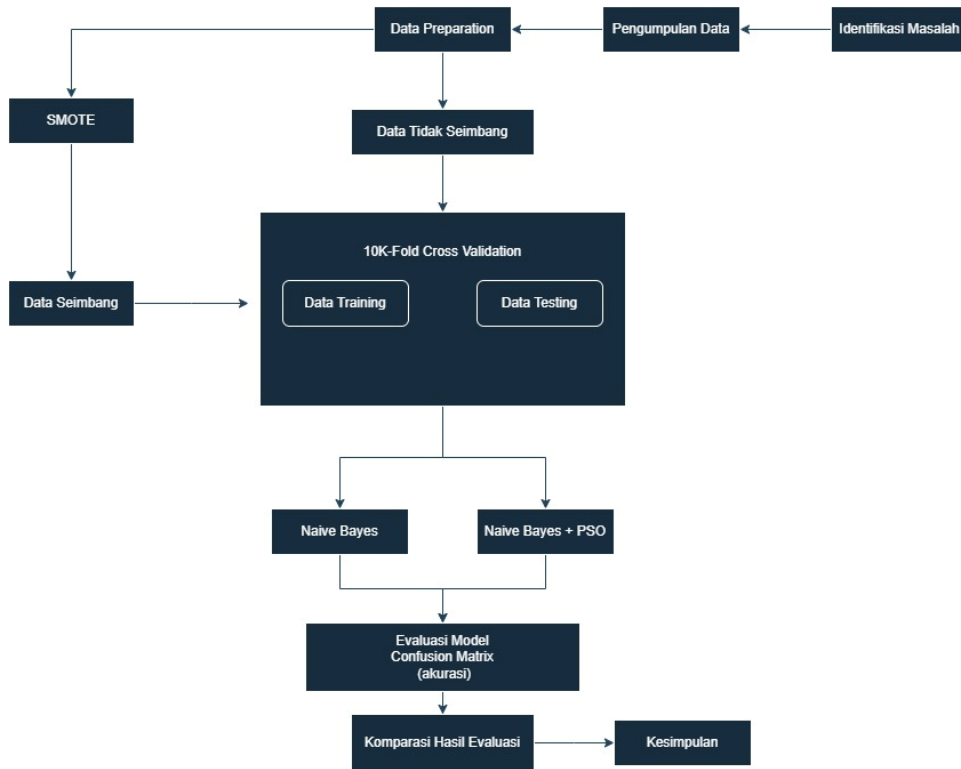
Penelitian-penelitian terdahulu menunjukkan bahwa SMOTE efektif mengatasi masalah class imbalance, terutama pada dataset bencana alam seperti tanah longsor, di mana kejadian bencana lebih jarang dibandingkan kejadian non-bencana. Siti Mutmainah dan Dhifa Zhafirah menemukan bahwa teknik oversampling seperti ROS dapat meningkatkan performa model secara signifikan, dengan akurasi 95% dan 96.43%, jauh lebih tinggi dibandingkan teknik undersampling. Selain itu, penelitian Raenald Syaputra et al. menunjukkan bahwa SMOTE, ketika digabungkan dengan algoritma optimasi seperti GA dan PSO (Particle Swarm Optimization), dapat meningkatkan akurasi klasifikasi meski menghadapi tantangan variasi hasil akibat karakteristik dataset yang berbeda. Meskipun SMOTE efektif, penelitian ini juga menunjukkan bahwa penyesuaian parameter sangat penting untuk mencapai hasil yang optimal. Penelitian ini mengatasi kekurangan dari penelitian terdahulu yang menggunakan teknik oversampling dan undersampling, yang tidak selalu menghasilkan performa optimal. Siti Mutmainah menemukan bahwa oversampling lebih efektif daripada undersampling. Dengan menggabungkan metode PSO-SMOTE pada Naive Bayes, penelitian ini berfokus pada pencarian parameter optimal melalui PSO dan memanfaatkan keuntungan SMOTE untuk menangani imbalance data. Meskipun penelitian terdahulu lebih sering menggunakan model seperti SVM, penelitian ini memilih Naive Bayes yang lebih sederhana namun tetap efektif untuk klasifikasi bencana tanah longsor. Ekspektasi dari penelitian ini adalah peningkatan akurasi klasifikasi tanah longsor yang lebih baik, dengan kemampuan model menangani imbalance data secara efektif dan menghasilkan hasil yang lebih robust berkat optimasi parameter menggunakan PSO.

1. Objek Penelitian

Penelitian ini menggunakan data kasus tanah longsor periode 2022-2023 dari BMKG dan BPBD Samarinda sebagai objek penelitian. Lokasi data diperoleh dari BPBD (Jl. Pulau Flores, Samarinda Kota) dan BMKG (Jl. Pipit No. 150, Sungai Pinang, Samarinda).

2. Prosedur Penelitian

Prosedur penelitian ini adalah serangkaian Langkah-langkah yang dilakukan untuk menyelesaikan masalah dalam penelitian.



A. Mengidentifikasi Masalah

Masalah yang ditemukan dalam penelitian ini berkaitan ketidakseimbangan data tanah longsor, di mana kelas minoritas sering terabaikan, sehingga prediksi kurang akurat. Algoritma Naive Bayes dioptimalkan dengan metode SMOTE dan PSO.

B. Pengumpulan Data

Data penelitian ini berasal dari BPBD dan BMKG Kota Samarinda (2022-2024), dengan 10 fitur dari BPBD dan 11 fitur dari BMKG yang digunakan untuk mengklasifikasikan bencana tanah longsor.

C. Data Preparation

Kemudian data yang telah diperoleh dari BPBD dan BMKG diolah melalui tahapan *Data Integration*, *Data Selection*, *Data Cleaning*, dan *data balancing*.

D. Pembagian Data

Dataset dibagi menjadi data train dan test untuk pelatihan dan pengujian model. Teknik K-Fold Cross-Validation (K=10) digunakan untuk meningkatkan akurasi evaluasi model.

E. Permodelan

Model dibangun dengan menerapkan SMOTE untuk mengatasi ketidakseimbangan data dan PSO untuk optimasi fitur, meningkatkan performa Naive Bayes.

F. Evaluasi

Di tahap ini, performa model diukur untuk mengecek akurasi dan kualitas data latih yang digunakan. Pengujian dilakukan dengan teknik *Confusion Matrix*. *Confusion Matrix* adalah sebuah teknik yang digunakan untuk melakukan perhitungan akurasi pada data *mining*.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Keterangan:

True Positive (TP), *True Negative (TN)*, *False Positive (FP)* dan *False Negative (FN)*

III. HASIL DAN PEMBAHASAN

1. Pengumpulan Data

Penelitian ini menggunakan data tanah longsor dari BPBD dan BMKG periode 2022-2024. Data mencakup 11 fitur dari BMKG (Tabel 1) dan 10 fitur dari BPBD (Tabel 2). Adapun bentuk data diperoleh adalah sebagai berikut:

TABEL I
DATA BMKG

No	Tanggal	Tn	Tx	Tavg	RH avg	RR	ss	ff x	ddd x	ff avg	Ddd car
0	2022-01-01	26	33.6	29	74	0	-	4	100	3	NE
1	2022-01-02	23.6	33.6	28.5	76.0	0	6.2	4	100	2	E
...
729	2023-12-31	24.6	32.4	28.3	84	0	6.9	4	90	2	E

TABEL II
DATA BPBD

No	Tanggal	Jam Kejadian	Jenis bencana	Lokasi/Wilayah Kelurahan Kesamatan	Luas Area M ²	Jumlah Obyek Yang Terkena Bencana	korban					Jumlah Kerugian Jiwa (RP)	Keterangan
							K L	K S	K H	K M	K K		
1	12/01/22	-	Kebakaran pemukiman	Jln. Pakis Merah 9 Rt.44, Kel. Sempaja Timur, Kec. Samarinda	-	Area Terbakar: Total 1 Bangunan Rincian 1 Bangunan Posyandu Tunas Bangsa	-	-	-	-	-	-	Masih dalam penyidikan pihak kepolisian, Dugaan sementara ada sekumpulan anak-anak bermain korek api di seputaran posyandu
...
601	01/12/24	Pukul 04.30 WITA	Pohon Tumbang	Jl. Ir. Sutami, Karang Asam Ulu, Kec. Sungai Kunjang	-	Jl. Ir. Sutami, Karang Asam Ulu, Kec. Sungai Kunjang	-	-	-	-	-	-	Penyebab Hujan intensitas Sedang – Lebat disertai angin kencang pada pukul 04.00 Wita

2. Hasil Data Preparation

Setelah pengumpulan data selesai, langkah berikutnya adalah melakukan pra-pemrosesan data, seperti pembersihan data dan seterusnya agar data tersebut siap untuk di proses pada tahap permodelan.

A. Hasil Data Integration

Pada Tabel 3 tahap integrasi data, dilakukan penggabungan dua dataset berbeda, yaitu data tanah longsor dari BPBD dan BMKG Kota Samarinda, untuk menghasilkan data sekunder yang lebih komprehensif. Proses ini menghasilkan satu dataset dengan 20 fitur, setelah menghapus fitur tanggal yang duplikat, guna memastikan kualitas data sebelum dilakukan klasifikasi.

TABEL III
HASIL DATA INTEGRASI

No	Fitur	Tipe Data
1	Tanggal	date
2	Jam Kejadian	string
3	Jenis Bencana	string
4	Lokasi Wilayah	string

5	Luas Area M ²	string
6	Objek Terkena Bencana	string
7	Korban	numeric
8	Jumlah Jiwa	numeric
9	Kerugian	string
10	Keterangan	string
11	Temperatur-maksimum (Tn) (°C)	numeric
12	Temperatur-minimum (Tx) (°C)	numeric
13	Temperatur-rata-rata (Tavg) (°C)	numeric
14	Kelembaban-rata-rata (RH_avg) (%)	numeric
15	Curah-hujan (RR) (mm)	numeric
16	Lamanya-penyinaran-matahari (ss) (hrs)	numeric
17	Kecepatan-angin-maksimum (ff_x) (m/s)	numeric
18	Arah-angin-maksimum (ddd_x) (°)	numeric
19	Kecepatan-angin-rata-rata (ff_avg) (m/s)	numeric
20	Arah-angin-terbanyak (ddd_car) (°)	numeric

B. Hasil Data Selection

TABEL IV
HASIL DATA SELECTION

No	Tanggal	Tn	Tx	Tavg	RH_avg	RR	ss	ff_x	ddd_x	ff_avg	Ddd_car	Terjadi longsor
0	2022-01-01	26	33.6	29	74	0	-	4	100	3	NE	0
1	2022-01-02	23.6	33.6	28.5	76.0	0	6.2	4	100	2	E	0
...
729	2023-12-31	24.6	32.4	28.3	84	0	6.9	4	90	2	E	0

Pada Tabel 4 fitur di pilih berdasarkan relevansinya terhadap tanah longsor. Fitur yang tidak relevan dihilangkan dan fitur yang tersisa salah satunya akan di tentukan sebagai kelas atau label.

C. Hasil Data Cleaning

TABEL V
HASIL DATA CLEANING

No	Tanggal	Tn	Tx	Tavg	RH_avg	RR	ss	ff_x	ddd_x	ff_avg	Ddd_car	Terjadi longsor
0	2022-01-02	23.6	33.6	28.5	76.0	0	6.2	4	100	2	E	0
1	2022-01-03	25.1	30.4	27.2	86	3.9	6.9	4	80	2	E	0
...
528	2023-12-31	24.6	32.4	28.3	84	0	6.9	4	90	2	E	0

Pada Tabel 5 hasil dari data Cleaning yaitu dataset awal memiliki total 729 baris data dengan indeks dari 0 hingga 729. Namun, setelah proses cleaning (seperti penghapusan duplikasi atau data kosong), hanya 528 baris data yang memenuhi kriteria untuk dianalisis lebih lanjut.

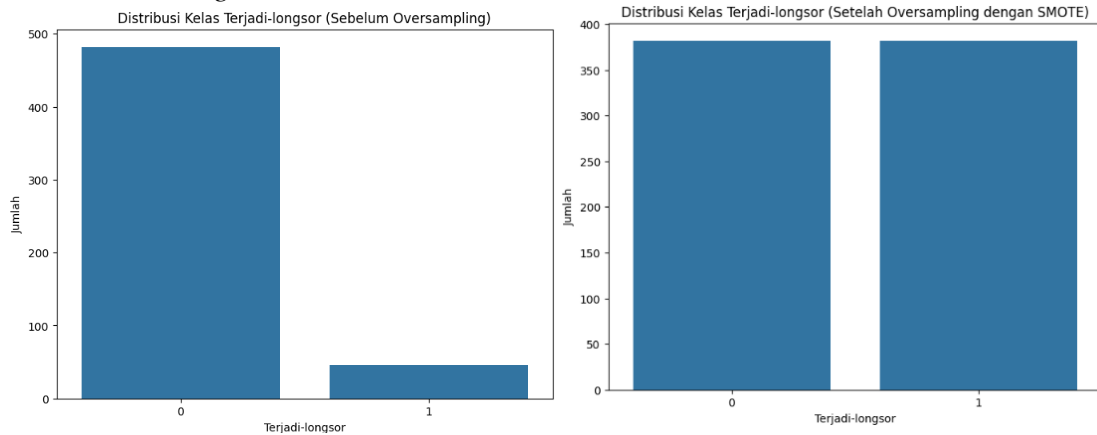
D. Hasil Data Transformasi

TABEL VI
SEBELUM DAN SESUDAH DATA TRANSFORMASI

No	(Sebelum) ddd_car	(Sesudah) ddd_car
0	E	1
1	E	1
...		
528	E	1

Tabel 6 menunjukkan transformasi kolom Arah-angin-terbanyak dari data kategorikal seperti E (Timur) dan NE (Timur Laut) menjadi data numerik menggunakan metode encoding, sehingga lebih mudah dianalisis oleh algoritma pembelajaran mesin.

E. Hasil Data *Balancing*



Gambar 1 Distribusi Kelas Sebelum Dan Sesudah *Balancing*

Pada Gambar 1 distribusi kelas sebelum dan sesudah *balancing*. Grafik sebelah kiri menunjukkan ketidakseimbangan kelas terjadi-longsor sebelum *balancing*, di mana jumlah kelas minoritas jauh lebih sedikit dibandingkan kelas mayoritas. Grafik sebelah kanan menunjukkan hasil *balancing* menggunakan metode SMOTE, di mana distribusi kelas menjadi seimbang.

TABEL VII
HASIL DISTRIBUSI KELAS SEBELUM DAN SESUDAH *BALANCING*

Kelas	Sebelum SMOTE	Sesudah SMOTE
0	382	382
1	40	382

Pada Tabel 7 sebelum SMOTE, data tidak seimbang dengan 382 sampel kelas 0 (tidak longsor) dan 40 sampel kelas 1 (longsor). Setelah SMOTE, distribusi menjadi seimbang dengan masing-masing 382 sampel. Teknik ini menciptakan sampel sintetis untuk kelas minoritas (1).

3. Hasil Permodelan Dan Evaluasi

Tahapan ini bertujuan menampilkan akurasi model Naive Bayes yang dioptimalkan dengan PSO dan SMOTE untuk mengatasi ketidakseimbangan data longsor. Evaluasi dilakukan menggunakan 10-Fold Cross Validation dan Confusion Matrix untuk menilai akurasi serta performa model secara keseluruhan.

A. Implementasi *Naïve Bayes*

TABEL VIII
HASIL RATA-RATA AKURASI *NAÏVE BAYES* TANPA SMOTE

Total Value Each Fold	TP	FP	TN	FN	Average Accuracy
	4	18	364	36	87.18%

$$Accuracy = \frac{4 + 364}{4 + 18 + 364 + 36} = \frac{368}{422} = 0.8718$$

$$Accuracy = 0.8718 \times 100\% = 87.18\%$$

TABEL IX
HASIL RATA-RATA AKURASI *NAÏVE BAYES* MENGGUNAKAN SMOTE

<i>Total Value Each Fold</i>	TP	FP	TN	FN	<i>Average Accuracy</i>
	295	138	244	87	70.55%

$$Accuracy = \frac{295 + 244}{295 + 138 + 244 + 87} = \frac{539}{764} = 0.7055$$

$$Accuracy = 0.7055 \times 100\% = 70.55\%$$

Tabel 8 menunjukkan akurasi rata-rata *Naïve Bayes* tanpa SMOTE sebesar 87.18%. Tabel 9 menunjukkan akurasi setelah menerapkan SMOTE sebesar 70.55%, lebih rendah dari tanpa SMOTE. Penurunan ini disebabkan oleh penambahan data sintetis yang memengaruhi distribusi data dan kinerja model.

B. Implementasi *Naïve Bayes* + PSO Seleksi Fitur

TABEL X
HASIL SELEKSI FITUR *NAÏVE BAYES* + PSO SELEKSI FITUR TANPA SMOTE

Fitur yang dipilih (tanpa SMOTE)	Fitur yang dipilih (dengan SMOTE)
0	0
1	1
2	3
3	4
5	5
6	6
7	7
8	8
9	

Tabel 12 menunjukkan hasil seleksi fitur menggunakan *Naïve Bayes* yang dipadukan dengan PSO. Tanpa SMOTE, fitur yang dipilih adalah indeks 0, 1, 2, 3, 5, 6, 7, dan 8. Dengan SMOTE, fitur indeks 2 tidak lagi terpilih, menunjukkan bahwa penambahan data sintetis dapat memengaruhi relevansi fitur dalam seleksi. Seleksi fitur ini membantu menentukan atribut penting untuk model prediksi optimal.

TABEL XI
HASIL RATA-RATA AKURASI *NAÏVE BAYES* + PSO SELEKSI FITUR TANPA SMOTE

<i>Total Value Each Fold</i>	TP	FP	TN	FN	<i>Average Accuracy</i>
	0	1	380	40	90.21%

$$Accuracy = \frac{0 + 380}{0 + 0 + 380 + 40} = \frac{380}{421} = 0.9021$$

$$Accuracy = 0.9021 \times 100\% = 90.21\%$$

TABEL XII
HASIL RATA-RATA AKURASI *NAÏVE BAYES* + PSO SELEKSI FITUR DENGAN SMOTE

<i>Total Value Each Fold</i>	TP	FP	TN	FN	<i>Average Accuracy</i>
	341	169	196	80	68.25%

$$Accuracy = \frac{341 + 196}{341 + 169 + 196 + 80} = \frac{399}{786} = 0.6825$$

$$Accuracy = 0.6825 \times 100\% = 68.25\%$$

Tabel 11 menunjukkan akurasi tanpa SMOTE sebesar 90.21%, menunjukkan kinerja model yang sangat baik. Di sisi lain, Tabel 12 menunjukkan penurunan akurasi menjadi 68.25% setelah penerapan SMOTE. Penurunan ini menunjukkan bahwa penggunaan SMOTE dapat mengubah distribusi data, yang berdampak pada efektivitas seleksi fitur dan kinerja model secara keseluruhan.

C. Implementasi *Naïve Bayes* + PSO Optimasi

TABEL XIII

HASIL RATA-RATA AKURASI *NAÏVE BAYES* + PSO TANPA SMOTE

<i>Total Value</i>	TP	FP	TN	FN	<i>Average Accuracy</i>
<i>Each Fold</i>	3	6	384	37	89.9%

$$Accuracy = \frac{3 + 384}{3 + 6 + 384 + 37} = \frac{387}{430} = 0.899$$

$$Accuracy = 0.899 \times 100\% = 89.9\%$$

TABEL XIV

HASIL RATA-RATA AKURASI *NAÏVE BAYES* + PSO MENGGUNAKAN SMOTE

<i>Total Value</i>	TP	FP	TN	FN	<i>Average Accuracy</i>
<i>Each Fold</i>	299	100	272	133	71,03%

$$Accuracy = \frac{299 + 272}{299 + 100 + 272 + 133} = \frac{571}{804} = 0.7103$$

$$Accuracy = 0.7103 \times 100\% = 71,03\%$$

Tabel 13 menunjukkan akurasi tanpa SMOTE sebesar 89.9%, mencerminkan kinerja model yang sangat baik tanpa data sintesis. Sebaliknya, Tabel 14 menunjukkan penurunan akurasi menjadi 71.03% dengan penggunaan SMOTE. Penurunan ini kemungkinan disebabkan oleh perubahan distribusi data yang dihasilkan oleh SMOTE, yang memengaruhi efektivitas kombinasi algoritma *Naïve Bayes* dengan PSO.

D. Implementasi *Naïve Bayes* + PSO Seleksi Fitur + PSO Optimasi

TABEL XV

HASIL RATA-RATA AKURASI *NAÏVE BAYES* + PSO SELEKSI FITUR + PSO OPTIMASI TANPA SMOTE

<i>Total Value</i>	TP	FP	TN	FN	<i>Average Accuracy</i>
<i>Each Fold</i>	0	0	382	40	90,57%

$$Accuracy = \frac{0 + 382}{0 + 382 + 0 + 40} = \frac{382}{422} = 0.9057$$

$$Accuracy = 0.9057 \times 100\% = 90,57\%$$

TABEL XVI

HASIL RATA-RATA AKURASI *NAÏVE BAYES* + PSO SELEKSI FITUR + PSO OPTIMASI DENGAN SMOTE

<i>Total Value</i>	TP	FP	TN	FN	<i>Average Accuracy</i>
<i>Each Fold</i>	311	171	205	71	68.08%

$$Accuracy = \frac{311 + 205}{311 + 205 + 171 + 71} = \frac{516}{758} = 0.6808$$

$$Accuracy = 0.6808 \times 100\% = 68.08\%$$

Tabel 15 menunjukkan akurasi tanpa SMOTE sebesar 90,57%, mencerminkan kinerja model yang sangat baik tanpa data sintesis. Sebaliknya, Tabel 16 menunjukkan penurunan akurasi menjadi 68,08% dengan penerapan SMOTE. Penurunan ini mengindikasikan bahwa penambahan data sintesis melalui SMOTE dapat memengaruhi efektivitas optimasi PSO, sehingga kinerja model menjadi kurang optimal.

E. Perbandingan Hasil Evaluasi

TABEL XVII
HASIL EVALUASI PERMODELAN RATA-RATA AKURASI TANPA SMOTE

<i>Avarage Accuracy</i>	<i>Naïve Bayes</i>	<i>Naïve Bayes + PSO Seleksi Fitur</i>	<i>Naïve Bayes + PSO Optimasi</i>	<i>Naïve Bayes + PSO Seleksi Fitur + PSO Optimasi</i>	<i>Perubahan Naïve Bayes ke Naïve Bayes + PSO Seleksi Fitur</i>	<i>Perubahan Naïve Bayes ke Random Naïve Bayes + PSO Optimasi</i>	<i>Perubahan Naïve Bayes ke Naïve Bayes + PSO Seleksi Fitur + PSO Optimasi</i>
	87.18%	90.21%	89.9%	90,57%	+3.03%	+2.72%	+3.39

Tabel 17 menunjukkan hasil evaluasi akurasi berbagai kombinasi metode Naïve Bayes tanpa SMOTE, termasuk peningkatan akurasi dengan penambahan PSO. Naïve Bayes standar memiliki akurasi 87.18%, sedangkan Naïve Bayes + PSO Seleksi Fitur meningkatkan akurasi menjadi 90.21% (+3.03%). Kombinasi Naïve Bayes + PSO Optimasi mencatat akurasi 89.9% (+2.72%). Model dengan kombinasi penuh Naïve Bayes + PSO Seleksi Fitur + PSO Optimasi mencapai akurasi tertinggi 90.57% (+3.39%). Hasil ini menunjukkan bahwa penambahan PSO, baik untuk seleksi fitur maupun optimasi, meningkatkan akurasi model, dengan kombinasi penuh memberikan hasil terbaik.

TABEL XVIII
HASIL EVALUASI PERMODELAN RATA-RATA AKURASI DENGAN SMOTE

<i>Avarage Accuracy</i>	<i>Naïve Bayes</i>	<i>Naïve Bayes + PSO Seleksi Fitur</i>	<i>Naïve Bayes + PSO Optimasi</i>	<i>Naïve Bayes + PSO Seleksi Fitur + PSO Optimasi</i>	<i>Perubahan Naïve Bayes ke Naïve Bayes + PSO Seleksi Fitur</i>	<i>Perubahan Naïve Bayes ke Random Naïve Bayes + PSO Optimasi</i>	<i>Perubahan Naïve Bayes ke Naïve Bayes + PSO Seleksi Fitur + PSO Optimasi</i>
	70.55%	68.25%	71,03%	68.08%	-2.3%	+0.48%	-2.47%

Hasil dalam Tabel 18, dimana penerapan SMOTE justru menyebabkan penurunan akurasi pada beberapa kombinasi metode, memang berbeda dari ekspektasi awal dan juga dari beberapa penelitian terdahulu yang menunjukkan bahwa SMOTE dapat meningkatkan akurasi dengan mengatasi masalah imbalance data. Penurunan akurasi ini dapat dijelaskan dengan beberapa faktor. Pertama, SMOTE bekerja dengan menghasilkan data sintesis untuk meningkatkan jumlah data minoritas, namun penambahan data sintesis ini dapat menyebabkan perubahan dalam distribusi data yang berdampak pada efektivitas model. Dalam penelitian ini, akurasi menurun saat SMOTE diterapkan, terutama ketika dikombinasikan dengan PSO Seleksi Fitur dan PSO Optimasi, kemungkinan disebabkan oleh data sintesis yang mungkin tidak sepenuhnya mencerminkan distribusi data asli dengan baik, sehingga mempengaruhi kinerja model. Selain itu, penggunaan PSO Seleksi Fitur dan PSO Optimasi dapat memperkenalkan tantangan tambahan. Teknik seleksi fitur bertujuan untuk memilih atribut yang paling relevan, namun penambahan data sintesis mungkin membawa variabel yang kurang

informatif atau bahkan noise, yang dapat mempengaruhi proses seleksi fitur dan optimasi model. Meskipun demikian, kombinasi Naïve Bayes dengan PSO Optimasi menunjukkan peningkatan akurasi meskipun hanya sedikit (+0.48%), mungkin karena PSO Optimasi lebih efektif dalam menyesuaikan parameter model dengan data yang ada, meskipun data sintetis telah mengubah distribusinya.

Perbandingan dengan penelitian terdahulu yaitu Siti Mutmainah[14] menunjukkan bahwa teknik *oversampling* seperti ROS menghasilkan performa yang lebih tinggi, tetapi penerapan SMOTE dalam penelitian Anda memberikan hasil yang berbeda. Hal ini menunjukkan bahwa meskipun SMOTE adalah teknik *oversampling*, dampaknya terhadap model bisa bervariasi tergantung pada jenis data dan metode yang digunakan. Penelitian Dhifa Zhafirah[15] melaporkan bahwa ROS dapat meningkatkan akurasi secara signifikan hingga 96.43%, namun dalam penelitian ini, penerapan SMOTE justru tidak menghasilkan peningkatan signifikan, dan bahkan menurunkan akurasi pada beberapa kombinasi. Ini mungkin disebabkan oleh cara SMOTE diterapkan yang berbeda atau perbedaan metode dalam menangani data imbalance. Selain itu, penelitian Raenald Syaputra et al[16] yang menggunakan kombinasi SMOTE dan PSO juga menunjukkan akurasi yang lebih tinggi, namun penurunan akurasi dalam penelitian Anda dapat disebabkan oleh karakteristik dataset yang berbeda atau hasil optimasi yang kurang efektif pada data sintetis. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa meskipun SMOTE dapat meningkatkan jumlah data minoritas, dampaknya terhadap akurasi model sangat tergantung pada bagaimana data sintetis tersebut berinteraksi dengan algoritma dan teknik optimasi yang digunakan.

IV. KESIMPULAN

Berdasarkan hasil penelitian, penerapan SMOTE dalam kombinasi dengan Naïve Bayes dan PSO menunjukkan dampak yang bervariasi terhadap akurasi model, dengan beberapa kombinasi mengalami penurunan akurasi. Hal ini menunjukkan bahwa SMOTE, meskipun dirancang untuk mengatasi ketidakseimbangan data, dapat mengubah distribusi data dan mempengaruhi efektivitas model. Kombinasi Naïve Bayes dengan PSO Optimasi menunjukkan peningkatan akurasi meskipun kecil. Rekomendasi dari penelitian ini adalah untuk menggunakan SMOTE dengan hati-hati, mempertimbangkan karakteristik data agar tidak memperkenalkan noise, dan lebih fokus pada optimasi model menggunakan PSO daripada seleksi fitur pada data sintetis. Dalam konteks manajemen bencana, penting untuk memperhatikan interaksi antara teknik optimasi dan data sintetis untuk meningkatkan akurasi sistem prediksi.

REFERENSI

- [1] I. N. Setiawan, D. Krismawati, S. Pramana, and E. Tanur, "Klasterisasi Wilayah Rentan Bencana Alam Berupa Gerakan Tanah Dan Gempa Bumi Di Indonesia," pp. 669–676, 2022.
- [2] D. Susanti and T. Wahyuni, "ANALISIS POTENSI BENCANA ALAM TANAH LONGSOR KABUPATEN MAJALENGKA MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER," *INFOTECH J.*, vol. 9, no. 2, pp. 299–306, Jul. 2023, doi: 10.31949/infotech.v9i2.5645.
- [3] M. Herviany, S. Putri Delima, T. Nurhidayah, and Kasini, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Daerah Rawan Tanah Longsor di Provinsi Jawa Barat," *MALCOM*

- Indones. J. Mach. Learn. Comput. Sci. J.*, vol. 1, pp. 34–40, 2021.
- [4] A. Gatto, S. Clò, F. Martellozzo, and S. Segoni, “Tracking a Decade of Hydrogeological Emergencies in Italian Municipalities,” *Data*, vol. 8, no. 10, pp. 1–11, 2023, doi: 10.3390/data8100151.
- [5] S. Wahyuni, S. Karim, and D. Arifin, “Sistem Informasi Geografis Pemetaan Daerah Rawan Longsor Kota Samarinda Berbasis Web Menggunakan Metode Skor Dan Pembobotan,” *Antivirus J. Ilm. Tek. Inform.*, vol. 15, no. 2, pp. 209–227, 2021, doi: 10.35457/antivirus.v15i2.1723.
- [6] R. Amaliah, E. Tohidi, E. Wahyudin, A. Rizki Rinaldi, and I. Iin, “Pengelompokan Data Bencana Alam Berdasarkan Wilayah Menggunakan Algoritma K-Means,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 6, pp. 3572–3579, 2024, doi: 10.36040/jati.v7i6.8253.
- [7] K. Sarawagi, K. Viridi, and N. S. Talwandi, “Predicting Landslides with Machine Learning: A Data-Driven Approach,” *Int. Conf. Comput. Intell. Comput. Appl.*, pp. 103–108, 2024.
- [8] H. Hairani and D. Priyanto, “A New Approach of Hybrid Sampling SMOTE and ENN to the Accuracy of Machine Learning Methods on Unbalanced Diabetes Disease Data,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 8, pp. 585–590, 2023, doi: 10.14569/IJACSA.2023.0140864.
- [9] G. Gumelar, Q. Ain, R. Marsuciati, S. Agustanti Bambang, A. Sunyoto, and M. Syukri Mustafa, “Kombinasi Algoritma Sampling dengan Algoritma Klasifikasi untuk Meningkatkan Performa Klasifikasi Dataset Imbalance,” 2021.
- [10] M. Dewi, T. H. Saragih, H. Rudy, R. A. Nugroho, and D. T. Nugrahadi, “PENERAPAN SMOTE-NCL UNTUK MENGATASI KETIDAKSEIMBANGAN KELAS PADA KLASIFIKASI PENYAKIT JANTUNG KORONER,” *JIP (Jurnal Inform. Polinema)*, vol. 10, no. 1, pp. 27–34, 2023.
- [11] J. B. E. Putry, A. T. Sasongko, and W. Hadikristanto, “Optimasi Decision Tree Menggunakan Particle Swarm Optimization (PSO) pada Risiko Kredit KMG Bank DKI,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 4, pp. 1403–1410, 2024, doi: 10.57152/malcom.v4i4.1521.
- [12] M. S. Hasibuan, D. Fransisca, J. Magister, T. Informatika, and F. I. Komputer, “Penggunaan Algoritma Naive Bayes dan Particle Swarm Optimization (PSO) untuk Mendeteksi Stroke,” pp. 109–118, 2022.
- [13] E. Suharyanto and A. Zein, “Analisis Data Minat Calon Mahasiswa Universitas Pamulang Dengan Menggunakan Algoritma Naive Bayes Classifier,” *Sainstech J. Penelit. Dan ...*, pp. 70–76, 2022, [Online]. Available: <https://ejournal.istn.ac.id/index.php/sainstech/article/view/1434%0Ahttps://ejournal.istn.ac.id/index.php/sainstech/article/download/1434/943>
- [14] S. Mutmainah, “Penanganan Imbalance Data Pada Klasifikasi Kemungkinan Penyakit Stroke,” *J. Sains, Nalar, dan Apl. Teknol. Inf.*, vol. 1, no. 1, pp. 10–16, 2021, doi: 10.20885/snati.v1i1.2.
- [15] D. Zhafirah, “PENANGANAN IMBALANCE DATA DENGAN RANDOM OVERSAMPLING (ROS) PADA KLASIFIKASI PENDERITA DIABETES MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM),” vol. 13, no. 1, pp. 104–116, 2023.
- [16] R. Syaputra, T. A. Y. Siswa, and W. J. Pranoto, “Model Optimasi SVM Dengan PSO-GA dan SMOTE Dalam Menangani High Dimensional dan Imbalance Data Banjir,” *Teknika*, vol. 13, no. 2, pp. 273–282, Jul. 2024, doi: 10.34148/teknika.v13i2.876.