

Penerapan Metode PSO-SMOTE Pada Algoritma *Random Forest* Untuk Mengatasi Class Imbalance Data Bencana Tanah Longsor

Dedy Ariyadi¹, Taghfirul Azhima Yoga Siswa^{2,*}, Rudiman³
^{1,2,3}Universitas Muhammadiyah Kalimantan Timur, Indonesia
E-mail : ¹2111102441060@umkt.ac.id, ^{2*}tay758@umkt.ac.id,
³rud959@umkt.ac.id

Abstract

Landslides are natural disasters that frequently occur in Samarinda City, with 45-80 affected areas in 2022-2023. The use of machine learning to classify landslide data faces the challenge of data imbalance, which can lead to bias towards the majority class. This study aims to address this issue by implementing the *Random Forest* algorithm combined with the *Synthetic Minority Oversampling Technique (SMOTE)* and optimization using *Particle Swarm Optimization (PSO)*. The data used comes from *BMKG* and *BPBD Samarinda City*, consisting of 11 features and 730 records. The results show that *SMOTE* successfully balanced the data, improving accuracy from 89.91% to 94.76%, an increase of 4.85%.

Keywords: *Landslide Classification, Random Forest, SMOTE, PSO*

Abstrak

Tanah longsor merupakan bencana alam yang sering terjadi di Kota Samarinda, dengan 45-80 titik terdampak pada 2022-2023. Penggunaan machine learning untuk mengklasifikasi data tanah longsor menghadapi tantangan imbalance data, yang dapat menyebabkan bias pada kelas mayoritas. Penelitian ini bertujuan mengatasi masalah tersebut dengan mengimplementasikan algoritma *Random Forest* yang dikombinasikan dengan metode *SMOTE (Synthetic Minority Oversampling Technique)* dan optimasi menggunakan *Particle Swarm Optimization (PSO)*. Data yang digunakan berasal dari *BMKG* dan *BPBD Kota Samarinda*, dengan 11 fitur dan 730 record. Hasil menunjukkan bahwa *SMOTE* berhasil menyeimbangkan data, meningkatkan akurasi dari 89,91% menjadi 94,76%, dengan peningkatan 4,85%.

Kata Kunci: *Klasifikasi Tanah Longsor, Random Forest, SMOTE, PSO*

1. Pendahuluan

Indonesia merupakan negara yang berada pada posisi geografis yang diapit oleh dua benua dan dua samudera dimana hal ini menjadikan posisi Indonesia sebagai keunggulan ekonomi yang baik akan tetapi menjadikan posisi Indonesia juga rawan akan kejadian bencana alam. Indonesia menjadi sebuah negara yang memiliki tingkat kerawanan bencana baik dari sisi geografis, iklim dan demografis [1]. Sebagai contoh Kota Samarinda yang berada di Provinsi Kalimantan Timur yang sering terjadi bencana alam tanah longsor, Badan Penanggulangan Bencana Daerah (BPBD) Samarinda Mencatat sebanyak Sembilan titik tanah longsor terdapat empat rumah warga rusak, satu rumah di kelurahan Teluk Lerong Ilir, dua rumah di jalan KS Tubun, dan satu rumah di Pasundan [2]. Tanah longsor adalah bentuk erosi dengan diiringi pergerakan tanah yang menyebabkan perpindahan material tanah maupun batuan dalam intensitas yang cukup besar. Kecepatan dari pergerakan dan perpindahan material tersebut dipengaruhi oleh jenis tanah dan bebatuan pada lokasi tersebut, selain itu hal ini juga dipengaruhi oleh intensitas hujan [3].

Terkait klasifikasi bencana alam, beberapa peneliti sebelumnya telah menerapkan beberapa pendekatan *machine learning*, seperti penggunaan algoritma *Naive Bayes*, *Support Vector Machine*, *Logistic Regression* [4], *KNN* [5], *C4.5* [6], dan lain-lain. Walaupun penelitian yang disebutkan menunjukkan akurasi rata-rata di atas 85%, namun data yang digunakan adalah jenis dataset berdimensi rendah atau *low dimension*. Berdasarkan data berdimensi yang diperoleh, maka ditemukan suatu masalah yaitu ketidakseimbangan data atau *class imbalance* data. Ketidakseimbangan data sering kali menjadi permasalahan dalam proses klasifikasi

Imbalance data kerap kali menjadi salah satu masalah dalam proses klasifikasi. Distribusi kelas yang tidak seimbang dapat memperburuk performa klasifikasi seperti *overfitting*, bias terhadap kelas mayoritas, representasi yang tidak akurat dan lain-lain [7]. Untuk mengatasi permasalahan tersebut para peneliti telah mengembangkan metode-metode khusus seperti *oversampling* atau *undersampling*. Perbedaan akurasi sesaat sebelum dan sesudah penggunaan teknik *balancing* data terkhusus *oversampling* dalam mengatasi kasus *imbalance data*, hal ini menunjukkan bahwa *imbalance data* memiliki pengaruh terhadap hasil akhir dari suatu klasifikasi [8]. Oleh karena itu, penelitian ini akan menerapkan teknik *oversampling* SMOTE dalam menangani masalah *imbalance data*.

Pada penelitian ini menggunakan algoritma klasifikasi *Random Forest* sebagai algoritma utama untuk melakukan klasifikasi data tanah longsor Kota Samarinda. *Random Forest* adalah algoritma *machine learning* yang menggabungkan hasil dari banyak *decision tree* untuk mencapai satu hasil. *Random Forest* diperkenalkan oleh Leo Breiman dan Adele Cutler, *Random Forest* merupakan salah satu algoritma *machine learning* yang populer, karena dapat menangani masalah klasifikasi maupun regresi. Salah satu keunggulan *Random Forest* adalah fleksibilitas dan kemudahan penggunaannya.

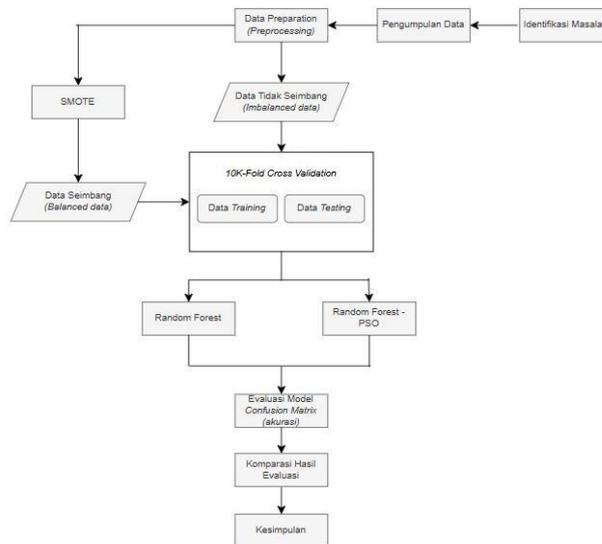
Setelah menentukan algoritma *Random Forest* penelitian ini juga menggunakan metode *Synthetic Minority Over-sampling Technique (SMOTE)*. *SMOTE* adalah sebuah metode untuk menyeimbangkan kelas yang berbeda dengan menggunakan *oversampling*. Pendekatan *SMOTE* membuat duplikasi data pada kelas minoritas agar seimbang dengan data pada kelas mayoritas. Ketidakseimbangan dataset dapat mengakibatkan hasil klasifikasi yang keliru, di mana data pada kelas minoritas sering kali salah diklasifikasikan sebagai kelas mayoritas [9]. Setelah dilakukan proses pengujian pada peneliti terdahulu diketahui bahwa penggabungan metode *SMOTE* dengan *Random Forest* memberikan hasil akurasi, sensitivitas, dan spesifisitas sebesar 99,7%. Sedangkan tanpa *SMOTE*, akurasi, sensitivitas, dan spesifisitasnya masing-masing hanya mencapai 90,2%, 90,2%, dan 89,7%. Kombinasi metode *SMOTE* dengan *Random Forest* meningkatkan performa model dengan peningkatan akurasi sebesar 9,5% dan sensitivitas sebesar 9,5% jika dibandingkan tanpa penggunaan *SMOTE* [10].

Penelitian ini juga menggunakan metode *Particle Swarm Optimization (PSO)* yang akan dikombinasikan dengan metode *SMOTE*, *PSO* adalah algoritma *optimasi* yang terinspirasi oleh perilaku kelompok partikel dalam mencari makanan di alam. Konsep dasar dari *PSO* adalah memodelkan partikel sebagai individu yang bergerak dalam ruang pencarian untuk mencari solusi terbaik. Algoritma ini pertama kali diperkenalkan oleh Dr. Eberhart dan Dr. Kennedy pada tahun 1995 [11].

Berdasarkan studi literatur yang dilakukan, belum ada peneliti yang mencoba kombinasi algoritma *Random Forest* dengan Teknik *oversampling SMOTE* dan metode *optimasi PSO* untuk meningkatkan performa dan akurasi *Random Forest* dalam menangani masalah klasifikasi *class imbalance* data tanah longsor Kota Samarinda. Oleh karena itu, penelitian ini bersifat baru diharapkan dapat memberikan kontribusi dalam meningkatkan akurasi klasifikasi suatu bencana dengan mengatasi kendala yang ada pada dataset yaitu *class imbalance* data bencana tanah longsor Kota Samarinda.

2. Metodologi Penelitian

Penelitian ini terdiri dari langkah-langkah yang dimulai dengan pengumpulan dan analisis data, hingga tahap akhir evaluasi hasil. Berikut adalah langkah-langkah penelitian yang akan dilakukan:



Gambar 1. Alur Penelitian

2.1. Identifikasi Masalah

Identifikasi masalah adalah tahap awal penelitian untuk memahami permasalahan yang menjadi fokus dan memandu seluruh proses penelitian. Masalah utama dalam penelitian ini adalah klasifikasi data bencana tanah longsor di Kota Samarinda. Selain itu, dilakukan studi pustaka untuk mengidentifikasi kesenjangan penelitian dalam klasifikasi data bencana tanah longsor.

2.2. Pengumpulan Data

Penelitian ini menggunakan data bencana tanah longsor Kota Samarinda yang diperoleh dari BPBD dan BMKG untuk periode 2022-2023. Data tersebut mencakup 10 fitur dari BPBD dan 11 fitur dari BMKG yang berkontribusi dalam proses klasifikasi bencana tanah longsor di Kota Samarinda.

2.3. Data Preparation

Data yang diperoleh dari Badan Penanggulangan Bencana Daerah (BPBD) dan Badan Meteorologi Klimatologi dan Geofisika (BMKG) memerlukan pengolahan data lebih lanjut sebelum dimasukkan ke dalam proses pemodelan, yang bertujuan untuk menghindari data yang tidak relevan. Proses pengolahan data ini meliputi beberapa tahapan penting yaitu *data integration*, *data selection*, *data cleaning*, *data transformation*, dan *data balancing* [12].

2.4. Data Integration

Data *integration* (penggabungan data) menggabungkan fitur data bencana tanah longsor dari Badan Penanggulangan Bencana Daerah (BPBD) dan Badan Meteorologi, Klimatologi dan Geofisika (BMKG) Kota Samarinda. Data ini mencakup detail tentang bencana tanah longsor di kota Samarinda dari tahun 2022 hingga 2023.

a. Data Selection

Data *selection* (pemilihan data) merupakan proses pemilihan fitur atau atribut mana saja yang relevan untuk dianalisis lebih lanjut pada klasifikasi data bencana tanah longsor, kemudian fitur-fitur tersebut yang tidak relevan akan dibuang atau dihilangkan.

b. Data Cleaning

Data *cleaning* (pembersihan data) sebelum dataset dimasukkan dalam model maka dilakukan proses data *cleaning* yang meliputi pengisian *missing value* (data yang kosong), menghaluskan noisy data, mengidentifikasi atau menghilangkan *outliner*, dan menghilangkan inkostensinya [13].

c. Data Transformation

Data *transformation* adalah proses mengubah data ke dalam format atau skala yang sesuai untuk analisis. Untuk tahap ini yang dilakukan terlebih dahulu adalah *LabelEncoder* yang bertujuan merubah data yang bertipe karakter menjadi tipe numerik [14].

d. Data Balancing dengan SMOTE

Tahap akhir pra-pemrosesan data adalah data balancing, yang bertujuan menyeimbangkan distribusi dataset untuk mencegah bias pada algoritma klasifikasi akibat ketidakseimbangan jumlah sampel antar kelas. Penelitian ini menggunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE) untuk mengatasi ketidakseimbangan data pada dataset bencana tanah longsor.

2.5. Pembagian Data

Dataset dibagi menjadi dua komponen: data latih (train) untuk melatih model dan data uji (test) untuk menguji kinerja model setelah pelatihan. Selain itu, penelitian ini menerapkan teknik *K-Fold Cross-Validation* dengan K=10 untuk mengevaluasi performa model machine learning secara lebih menyeluruh, karena metode ini dapat memberikan hasil yang lebih akurat dan andal. [15].

2.6. Permodelan

Penelitian menggunakan model klasifikasi *Random Forest* dengan *SMOTE* sebagai *Oversampling* dan PSO untuk seleksi fitur dan PSO untuk optimasi. Hasil akhir dari penelitian ini yaitu akan membandingkan kinerja model sebelum dan sesudah *SMOTE*.

2.7. Evaluasi

Tahap evaluasi merupakan langkah penting setelah pembentukan model, di mana performa model diukur untuk mengevaluasi akurasi dan kualitas data latih yang digunakan. Pengujian dilakukan menggunakan teknik *Confusion Matrix*, yaitu metode untuk menghitung akurasi dalam data mining dengan menganalisis prediksi model terhadap data aktual [16]. Adapun rumusnya adalah:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (1)$$

Keterangan:

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

3. Hasil Dan Pembahasan

3.1. Hasil Pengumpulan Data

Penelitian ini menggunakan data tanah longsor dari BPBD dan BMKG Samarinda untuk periode 2022-2023. Data mencakup 21 fitur penyebab dan dampak tanah longsor, dengan dua klasifikasi utama: terjadi tanah longsor dan tidak terjadi tanah longsor.

Tabel 1. 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 2. Data BPBD

No	Tanggal	Jam Kejadian	Jenis bencana	Lokasi/Wilayah Kelurahan Kesamatan	Luas Area M ²	Jumlah Obyek Yang Terkena Bencana	korban						Jumlah Jiwa	Kerugian (RP)	Keterangan
							KL	KS	KH	KM	KK	KK			
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 Luas Area: Terbakar 5 x 10 Mtr	-	-	-	-	-	-	-	-	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 Upaya: Melakukan Pemotongan pohon

3.2. Hasil Preparation

Pada tahapan ini akan menampilkan hasil dari *Preparation* data setelah melalui proses pengumpulan data sebelumnya agar data dapat di proses lebih lanjut layak masuk ke dalam tahap pemodelan, sehingga data bersih dari bagian-bagian yang tidak diperlukan.

a. Data Integration

Data tanah longsor yang diperoleh dari BMKG dan BPBD merupakan dua dataset berbeda yang belum terintegrasi. Untuk mempermudah pengolahan, kedua dataset ini digabung menjadi satu kesatuan, menghasilkan data sekunder yang lebih lengkap dan berkualitas sebelum memasuki proses klasifikasi. Setelah penggabungan, seluruh fitur dari kedua dataset digabung menjadi satu, menghasilkan total 20 fitur (dari 21, dikurangi fitur "tanggal" karena nama yang sama). Hasil penggabungan ini tidak dapat menampilkan keseluruhan data secara langsung, namun hanya menampilkan hasil integrasi fitur dari data tanah longsor BPBD dan BMKG.

Tabel 3. Fitur Hasil Data *Integration*

No	Fitur	Tipe Data	No	Fitur	Tipe Data
1	Tanggal	date	11	Temperatur-maksimum (Tn) (°C)	numeric
2	Jam Kejadian	string	12	Temperatur-minimum (Tx) (°C)	numeric
3	Jenis Bencana	string	13	Temperatur-rata-rata (Tavg) (°C)	numeric
4	Lokasi Wilayah	string	14	Kelembaban-rata-rata (RH_avg) (%)	numeric
5	Luas Area M ²	string	15	Curah-hujan (RR) (mm)	numeric
6	Objek Terkena Bencana	string	16	Lamanya-penyinaran-matahari (ss) (hrs)	numeric
7	Korban	numeric	17	Kecepatan-angin-maksimum (ff_x) (m/s)	numeric
8	Jumlah Jiwa	numeric	18	Arah-angin-maksimum (ddd_x) (°)	numeric
9	Kerugian	string	19	Kecepatan-angin-rata-rata (ff_avg) (m/s)	numeric
10	Keterangan	string	20	Arah-angin-terbanyak (ddd_car) (°)	numeric

b. Data Selection

Pada tahap pemilihan data, dilakukan seleksi fitur berdasarkan relevansinya terhadap faktor-faktor penyebab tanah longsor. Fitur-fitur yang tidak relevan dihilangkan sehingga hanya fitur yang relevan tersisa. Selain itu, fitur yang akan menjadi kelas atau label ditentukan pada tahap ini. Fitur "jenis bencana," yang berisi informasi tentang bencana yang terjadi, dicocokkan dengan data BMKG berdasarkan tanggal. Fitur ini kemudian digunakan sebagai kelas atau label yang digabungkan dengan data hasil integrasi sebelumnya.

Tabel 4. Hasil Data *Selection*

No	Tanggal	Tn	Tx	Tavg	RH_avg	RR	ss	ff_x	ddd_x	ff_avg	Ddd_car	Terjadi tanah longsor
1	2022-01-01	26	33.6	29.0	74.0	0	-	4.0	100	3.0	NE	0
2	2022-01-02	23.6	33.6	28.5	76.0	0	6.2	4.0	100	2.0	E	0
...
730	2023-12-31	24.6	32.4	28.3	84.0	0	6.9	4.0	90	2.0	E	0

c. Data Cleaning

Data *cleaning* dilakukan dengan cara menghapus baris yang setidaknya memiliki satu nilai yang kosong, sehingga baris-baris yang memiliki nilai tidak lengkap akan dihapus dan jumlah data yang sebelumnya berjumlah 730 data ketika melalui proses data *cleaning* akan berkurang menjadi 595 data.

Tabel 5. Hasil Data *Cleaning*

No	Tanggal	Tn	Tx	Tavg	RH_avg	RR	ss	ff_x	ddd_x	ff_avg	Ddd_car	Terjadi tanah longsor
1	2022-01-02	23.6	33.6	28.5	76.0	0	6.2	4.0	100.0	2.0	E	0
2	2022-01-03	25.1	30.4	27.2	86.0	3.9	6.9	4.0	80.0	2.0	E	0
...
730	2023-12-31	24.6	32.4	28.3	84.0	0	6.9	4.0	90.0	2.0	E	0

d. Data Transformasi

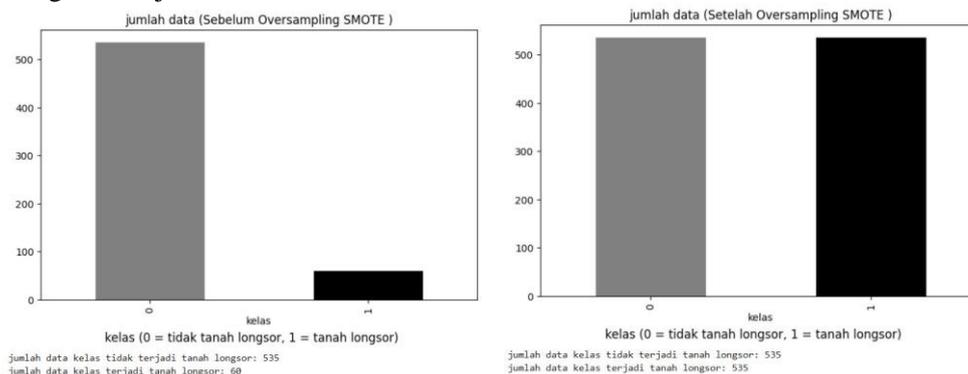
Fitur "arah angin terbanyak" (ddd_car) yang berisi simbol arah mata angin akan diubah menjadi angka menggunakan fungsi LabelEncoder dari `sklearn.preprocessing` pada Python.

Tabel 6. Sebelum dan Sesudah Data Transformation

No	(Sebelum) dd_car	(Sesudah) dd_car
1	E	1
2	E	1
...		
595	E	1

e. Data Balancing

Terdapat ketidakseimbangan antara kelas "tidak terjadi tanah longsor" (535) dan "terjadi tanah longsor" (60). Dengan menggunakan teknik oversampling SMOTE, jumlah kedua kelas diseimbangkan, sehingga keduanya memiliki 535 data, dan total data meningkat menjadi 1070 record.



Gambar 2. Barplot Sebelum Dan Sesudah *Balancing*

3.3. Hasil Pembagian Data Training dan Data Testing

Dalam penelitian ini, pembagian dataset menjadi data training dan data testing dilakukan melalui proses 10-Fold Cross-Validation, yang membagi total record secara merata untuk mengevaluasi kinerja model machine learning.

3.4. Hasil Permodalan Dan Evaluasi

a. Implementasi *Random Forest*

Tabel 7. Rata-Rata *Confusion Matrix Random Forest* Tanpa SMOTE

		<i>True Label</i>		<i>Predicted Label</i>	
		TN	TP	FN	FP
TN	533	0	60	2	
TP	0				

$$Accuracy = \frac{0 + 533}{0 + 533 + 2 + 60} = \frac{533}{595} \times 100\% = 89.57\%$$

Tabel 7 menunjukkan rata-rata nilai setiap fold dari *Confusion Matrix*, dengan TN = 533, TP = 0, FN = 60, dan FP = 2. Akurasi yang diperoleh adalah 89,57% berdasarkan perhitungan manual.

Tabel 8. Rata-Rata *Confusion Matrix Random Forest* SMOTE

		<i>True Label</i>		<i>Predicted Label</i>	
		TN	TP	FN	FP
TN	511	499	36	24	
TP	499				

$$Accuracy = \frac{499 + 511}{499 + 511 + 24 + 36} = \frac{1010}{1070} \times 100\% = 94.39\%$$

Terlihat dari Tabel 8 menunjukkan rata-rata nilai setiap fold dari Confusion Matrix, dengan TN = 511, TP = 499, FN = 36, dan FP = 24. Akurasi yang diperoleh adalah 94,39% berdasarkan perhitungan manual.

b. Implementasi *Random Forest*-PSO

Tabel 9. Rata-rata *Confusion Matrix Random Forest*-PSO Dengan SMOTE

<i>True Label</i>		<i>Predicted Label</i>	
<i>TN</i>	535	<i>FN</i>	60
<i>TP</i>	0	<i>FP</i>	0

$$Accuracy = \frac{0 + 535}{0 + 535 + 0 + 60} = \frac{535}{595} \times 100\% = 89.91\%$$

Tabel 9 menunjukkan rata-rata nilai setiap fold dari Confusion Matrix, dengan TN = 535, TP = 0, FN = 60, dan FP = 0. Akurasi yang diperoleh adalah 89,91% berdasarkan perhitungan manual.

Tabel 10. Rata-rata *Confusion Matrix Random Forest*-PSO Dengan SMOTE

<i>True Label</i>		<i>Predicted Label</i>	
<i>TN</i>	514	<i>FN</i>	35
<i>TP</i>	500	<i>FP</i>	21

$$Accuracy = \frac{500 + 514}{500 + 514 + 21 + 35} = \frac{1014}{1070} \times 100\% = 94.76\%$$

Tabel 10 menunjukkan rata-rata nilai setiap fold dari Confusion Matrix, dengan TN = 514, TP = 500, FN = 35, dan FP = 21. Akurasi yang diperoleh adalah 94,76% berdasarkan perhitungan manual.

3.5. Perbandingan Hasil Evaluasi

Tabel 11. Perbandingan Hasil Rata-Rata Akurasi Model Sebelum Penggunaan SMOTE

<i>Average Accuracy</i>	<i>Random Forest</i>	<i>Random Forest-PSO</i>	<i>Perubahan Random Forest ke Random Forest-PSO</i>
	89.57%	89.91%	0.34%

Tabel 11 menunjukkan perbandingan hasil rata-rata akurasi model sebelum penggunaan SMOTE (*Synthetic Minority Oversampling Technique*). Dalam tabel tersebut, ditampilkan rata-rata akurasi dari dua model, yaitu *Random Forest* dan *Random Forest-PSO*. Model *Random Forest* memiliki rata-rata akurasi sebesar 89,57%, sedangkan *Random Forest-PSO* menunjukkan peningkatan dengan rata-rata akurasi sebesar 89,91%. Dengan demikian, terdapat peningkatan akurasi sebesar 0,34% ketika metode PSO (*Particle Swarm Optimization*) diterapkan pada model *Random Forest*. Peningkatan ini mengindikasikan bahwa pengoptimalan menggunakan PSO mampu memperbaiki performa model *Random Forest*, meskipun peningkatan tersebut relatif kecil. Hal ini menunjukkan potensi metode PSO untuk meningkatkan hasil prediksi pada model pembelajaran mesin, khususnya pada kondisi data sebelum diterapkannya teknik SMOTE.

Tabel 12. Perbandingan Hasil Rata-Rata Akurasi Model Sesudah SMOTE

<i>Average Accuracy</i>	<i>Random Forest</i>	<i>Random Forest-PSO</i>	<i>Perubahan Random Forest ke Random Forest-PSO</i>
	94.39%	94.76%	0.37%

Tabel 12 memperlihatkan perbandingan hasil rata-rata akurasi model setelah diterapkannya SMOTE (*Synthetic Minority Oversampling Technique*). Pada tabel ini, model *Random Forest* mencatatkan rata-rata akurasi sebesar 94,39%, sementara model *Random Forest-PSO* menunjukkan akurasi yang lebih tinggi, yaitu 94,76%. Dengan demikian, terdapat peningkatan akurasi sebesar 0,37% ketika metode PSO (*Particle Swarm Optimization*) diterapkan pada *Random Forest*. Peningkatan ini lebih signifikan dibandingkan hasil sebelum penggunaan SMOTE (seperti yang ditunjukkan pada Tabel 12), yang menunjukkan bahwa kombinasi SMOTE dan PSO dapat memberikan dampak positif pada performa model. Hal ini mengindikasikan bahwa SMOTE efektif dalam menangani ketidakseimbangan data, sementara PSO mampu mengoptimalkan parameter model untuk menghasilkan akurasi yang lebih tinggi.

4. Kesimpulan

Berdasarkan hasil penerapan metode *oversampling SMOTE* sebagai teknik *balancing* pada data tanah longsor Kota Samarinda dalam mengatasi *imbalance data* bekerja dengan baik sebagaimana dapat dilihat pada Kesimpulan hasil perbandingan. Peningkatan akurasi klasifikasi data tanah longsor di Kota Samarinda dengan menggunakan kombinasi *Random Forest, PSO* dan *SMOTE* mencapai 94,76%.

Daftar Pustaka

- [1] K. Desderius, M. S. B. Arrinjani, Z. F. Sa'adia, and F. R. Lie, "Analisis tingkat risiko bencana tanah longsor di wilayah Kabupaten Blitar, Jawa Timur," *Reg. J. Pembang. Wil. dan Perenc. Partisipatif*, vol. 19, no. 1, p. 200, 2024, doi: 10.20961/region.v19i1.58889.
- [2] D. A. Zakarias Demon Daton, "Bencana Longsor Terjang 9 Titik di Samarinda, 4 Rumah Warga Rusak," *Kompas.com*, 2021. <https://regional.kompas.com/read/2021/07/02/174254478/bencana-longsor-terjang-9-titik-di-samarinda-4-rumah-warga-rusak>
- [3] M. F. Yassar *et al.*, "Penerapan Weighted Overlay Pada Pemetaan Tingkat Probabilitas Zona Rawan Longsor di Kabupaten Sumedang, Jawa Barat," *J. Geosains dan Remote Sens.*, vol. 1, no. 1, pp. 1–10, 2020, doi: 10.23960/jgrs.2020.v1i1.13.
- [4] K. W. Gusti, "Klasifikasi Bencana Alam Pada Twitter Menggunakan Naïve Bayes, Support Vector Machine Dan Logistic Regression," *Technol. J. Ilm.*, vol. 14, no. 4, p. 349, 2023, doi: 10.31602/tji.v14i4.11614.
- [5] M. Y. R. Rangkuti, M. V. Alfansyuri, and W. Gunawan, "Penerapan Algoritma K-Nearest Neighbor (Knn) Dalam Memprediksi Dan Menghitung Tingkat Akurasi Data Cuaca Di Indonesia," *Hexag. J. Tek. dan Sains*, vol. 2, no. 2, pp. 11–16, 2021, doi: 10.36761/hexagon.v2i2.1082.
- [6] A. Widiastari, S. Solikhun, and I. Irawan, "Analisa Datamining dengan Metode Klasifikasi C4.5 Sebagai Faktor Penyebab Tanah Longsor," *J. Comput. Syst. Informatics*, vol. 2, no. 3, pp. 247–255, 2021, [Online]. Available: <http://ejurnal.seminar-id.com/index.php/josyc/article/view/741>
- [7] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0192-5.
- [8] N. Razali, S. Ismail, and A. Mustapha, "Machine learning approach for flood risks prediction," *IAES Int. J. Artif. Intell.*, vol. 9, no. 1, pp. 73–80, 2020, doi: 10.11591/ijai.v9.i1.pp73-80.
- [9] N. N. Sholihah and A. Hermawan, "Implementation of Random Forest and Smote Methods for Economic Status Classification in Cirebon City," *J. Tek. Inform.*, vol. 4, no. 6, pp. 1387–1397, 2023, doi: 10.52436/1.jutif.2023.4.6.1135.
- [10] M. A. Latief, L. R. Nabila, W. Miftakhurrahman, S. Ma'rufatullah, and H.

- Tantyo, “Handling Imbalance Data using Hybrid Sampling SMOTE-ENN in Lung Cancer Classification,” *Int. J. Eng. Comput. Sci. Appl.*, vol. 3, no. 1, pp. 11–18, 2024, doi: 10.30812/ijecsa.v3i1.3758.
- [11] Kantinit, “Particle Swarm Optimization Adalah: Konsep dan Cara Kerja,” *Kantinit*, 2023. <https://kantinit.com/algorithm/particle-swarm-optimization-adalah-konsep-dan-cara-kerja/>
- [12] 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, 2024, doi: 10.34148/teknika.v13i2.876.
- [13] S. Dwiasnati and Y. Devianto, “Optimasi Prediksi Bencana Banjir menggunakan Algoritma SVM untuk penentuan Daerah Rawan Bencana Banjir,” *Pros. SISFOTEK*, pp. 202–207, 2021, [Online]. Available: <http://seminar.iaii.or.id/index.php/SISFOTEK/article/view/283>
- [14] I. R. Pratama, M. Maimunah, and E. R. Arumi, “Sistem Klasifikasi Penjualan Produk Alat Listrik Terlaris Untuk Optimasi Pengadaan Stok Menggunakan Naïve Bayes,” *J. Media Inform. Budidarma*, vol. 6, no. 4, p. 2135, 2022, doi: 10.30865/mib.v6i4.4418.
- [15] T. Ridwansyah, “Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 5, pp. 178–185, 2022, doi: 10.30865/klik.v2i5.362.
- [16] B. P. Pratiwi, A. S. Handayani, and S. Sarjana, “Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi Wsn Menggunakan Confusion Matrix,” *J. Inform. Upgris*, vol. 6, no. 2, pp. 66–75, 2021, doi: 10.26877/jiu.v6i2.6552.