

Penerapan PSO–RU Dalam Algoritma Naive Bayes Untuk Mengatasi Class Imbalance Data Bencana Tanah Longsor

Zakaria Ihza Akbar, Taghfirul Azhima Yoga Siswa*

Teknik Informatika Internasional, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Kalimantan Timur, Samarinda, Indonesia

Email: ¹zakariaihzaakbar@gmail.com, ²*tay758@umkt.ac.id

Email Penulis Korespondensi: tay758@umkt.ac.id

Submitted: 30/01/2025; Accepted: 20/03/2025; Published: 21/03/2025

Abstrak—Tanah longsor merupakan salah satu bencana alam signifikan di Indonesia yang sering menyebabkan kerugian besar baik dari segi ekonomi maupun keselamatan jiwa. Salah satu tantangan utama dalam pengolahan data tanah longsor adalah masalah ketidakseimbangan kelas (class imbalance), di mana jumlah data kejadian bencana jauh lebih sedikit dibandingkan dengan data non-bencana. Penelitian ini bertujuan untuk meningkatkan akurasi prediksi tanah longsor dengan mengintegrasikan algoritma Naive Bayes dan Particle Swarm Optimization (PSO), serta menggunakan teknik Random Undersampling (RU) untuk menangani ketidakseimbangan data. Dataset yang digunakan dalam penelitian ini mencakup data tanah longsor Kota Samarinda periode 2022-2023 yang diperoleh dari Badan Penanggulangan Bencana Daerah (BPBD) dan Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Proses penelitian meliputi tahap pra-pemrosesan data, balancing data menggunakan RU, penerapan algoritma Naive Bayes, dan optimasi menggunakan PSO. Evaluasi kinerja model dilakukan menggunakan teknik 10-Fold Cross Validation dan confusion matrix. Hasil penelitian menunjukkan bahwa penerapan algoritma Naive Bayes dengan optimasi PSO tanpa RU menghasilkan rata-rata akurasi tertinggi sebesar 89,49%, dibandingkan dengan Naive Bayes tanpa optimasi yang hanya mencapai 87,59%. Sementara itu, penerapan RU memberikan dampak yang beragam, di mana kombinasi Naive Bayes + PSO dengan RU menghasilkan rata-rata akurasi sebesar 50%, sedikit lebih baik dibandingkan Naive Bayes dengan RU yang hanya mencapai 45%. Penelitian ini membuktikan bahwa optimasi menggunakan PSO dapat meningkatkan performa model Naive Bayes dalam menghadapi dataset tanah longsor yang kompleks, meskipun teknik balancing seperti RU harus diterapkan dengan hati-hati untuk menghindari hilangnya informasi penting. Dengan hasil yang diperoleh, model ini diharapkan dapat mendukung upaya mitigasi bencana melalui prediksi yang lebih akurat, sehingga membantu pihak terkait dalam pengambilan keputusan, seperti perencanaan pengungsian dini dan pembangunan infrastruktur di area rawan longsor.

Kata Kunci: Tanah Longsor; Class Imbalance; Naive Bayes; Particle Swarm Optimization; Random Undersampling

Abstract—Landslides are one of the most significant natural disasters in Indonesia, often causing substantial economic losses and threats to human safety. A key challenge in processing landslide data is the issue of class imbalance, where the number of disaster occurrence data is significantly smaller compared to non-disaster data. This study aims to improve landslide prediction accuracy by integrating the Naive Bayes algorithm and Particle Swarm Optimization (PSO) while employing the Random Undersampling (RU) technique to address data imbalance. The dataset used in this study includes landslide data from Samarinda City for the period 2022-2023, obtained from the Regional Disaster Management Agency (BPBD) and the Meteorology, Climatology, and Geophysics Agency (BMKG). The research process involved data preprocessing, balancing data using RU, implementing the Naive Bayes algorithm, and optimizing it with PSO. Model performance was evaluated using the 10-Fold Cross Validation technique and a confusion matrix. The results show that applying the Naive Bayes algorithm with PSO optimization without RU achieved the highest average accuracy of 89.49%, compared to Naive Bayes without optimization, which only reached 87.59%. Meanwhile, the application of RU showed varied effects, with the combination of Naive Bayes + PSO with RU achieving an average accuracy of 50%, slightly better than Naive Bayes with RU, which only reached 45%. This study demonstrates that PSO optimization can improve the performance of the Naive Bayes model in handling complex landslide datasets, although balancing techniques such as RU must be applied cautiously to avoid the loss of important information. The results of this study are expected to support disaster mitigation efforts through more accurate predictions, aiding stakeholders in decision-making, such as early evacuation planning and infrastructure development in landslide-prone areas.

Keywords: Landslide; Class Imbalance; Naive Bayes; Particle Swarm Optimization; Random Undersampling

1. PENDAHULUAN

Tanah longsor merupakan fenomena geologi yang terjadi ketika massa tanah bergerak menuruni lereng akibat gravitasi, sering dipicu oleh faktor-faktor seperti curah hujan yang tinggi, aktivitas manusia, dan kondisi geologis yang tidak stabil [1]. Bencana ini tidak hanya menyebabkan kerugian material tetapi juga mengancam keselamatan jiwa, terutama di daerah yang rawan longsor. Oleh karena itu, pemahaman yang lebih baik mengenai pola kejadian tanah longsor menjadi sangat penting untuk mitigasi risiko dan pengurangan dampak bencana. Di Indonesia, tanah longsor menjadi salah satu bencana alam yang paling sering terjadi, dengan dampak yang signifikan terhadap kehidupan manusia dan lingkungan, termasuk kerugian jiwa dan harta benda [2]. Data bencana ini, yang biasanya disimpan dalam berbagai institusi dan pusat penelitian, memainkan peran penting dalam upaya mitigasi dan prediksi. Namun, salah satu tantangan utama dalam memproses data bencana adalah *class imbalance* yang terjadi karena jumlah data kejadian bencana yang jauh lebih sedikit dibandingkan dengan data non-bencana. Hal ini dapat menyebabkan model prediksi menjadi bias terhadap kelas mayoritas, sehingga kurang akurat dalam mendeteksi kejadian tanah longsor yang sebenarnya.

Data mining adalah proses dalam *data mining* yang bertujuan untuk membangun model yang dapat memprediksi kelas atau kategori dari data yang belum diketahui berdasarkan pola yang ditemukan dalam data yang

sudah ada. Metode ini semakin banyak digunakan dalam berbagai bidang, termasuk dalam analisis bencana alam, karena kemampuannya dalam mengolah data dalam jumlah besar dan menemukan pola tersembunyi. Metode seperti algoritma *Naive Bayes* telah terbukti efektif dalam klasifikasi data bencana, membantu dalam memprediksi kemungkinan terjadinya tanah longsor berdasarkan variabel yang ada [3].

Naive Bayes adalah algoritma klasifikasi yang berbasis pada *Teorema Bayes* dengan asumsi independensi antar fitur. Meskipun sederhana, *Naive Bayes* sering kali menunjukkan kinerja yang baik dalam berbagai aplikasi klasifikasi, terutama ketika data memiliki distribusi yang jelas. Selain itu, algoritma ini memiliki keunggulan dalam kecepatan komputasi dan kemudahan implementasi, menjadikannya pilihan yang menarik untuk berbagai aplikasi klasifikasi berbasis data. Algoritma ini didasarkan pada *Teorema Bayes* dan digunakan untuk memprediksi probabilitas suatu kelas berdasarkan fitur-fitur yang ada dalam data. Menurut penelitian Mardiani telah meneliti tentang klasifikasi potensi bencana tsunami di Indonesia menggunakan *machine learning*. Dalam konteks bencana alam, penggunaan *machine learning* sangat penting karena dapat membantu dalam analisis data yang kompleks dan beragam, serta memberikan rekomendasi berbasis data untuk pengambilan keputusan yang lebih baik. Penelitian ini menunjukkan bahwa meskipun *Naive Bayes* memiliki akurasi yang lebih rendah dibandingkan dengan algoritma lain, penerapan teknik optimasi dan resampling dapat meningkatkan kinerjanya[4]. Penelitian oleh Chamola menyoroiti bagaimana algoritma *machine learning*, termasuk *Naive Bayes*, dapat digunakan dalam manajemen bencana dan pandemi. Penelitian Chamola menunjukkan bahwa algoritma ini dapat membantu dalam pengenalan dan klasifikasi, yang sangat penting dalam konteks bencana alam[5]. Pada penelitian Ahmad Yani menunjukkan bagaimana algoritma *Naive Bayes* dapat diterapkan untuk klasifikasi berbasis data, termasuk seleksi fitur untuk meningkatkan efisiensi model, yang relevan dengan mengoptimalkan kinerja *Naive Bayes* menggunakan *PSO*[6]. Oleh karena itu, diperlukan pendekatan yang dapat meningkatkan akurasi klasifikasi pada data yang tidak seimbang, salah satunya dengan menggabungkan *Naive Bayes* dengan teknik optimasi seperti *PSO*.

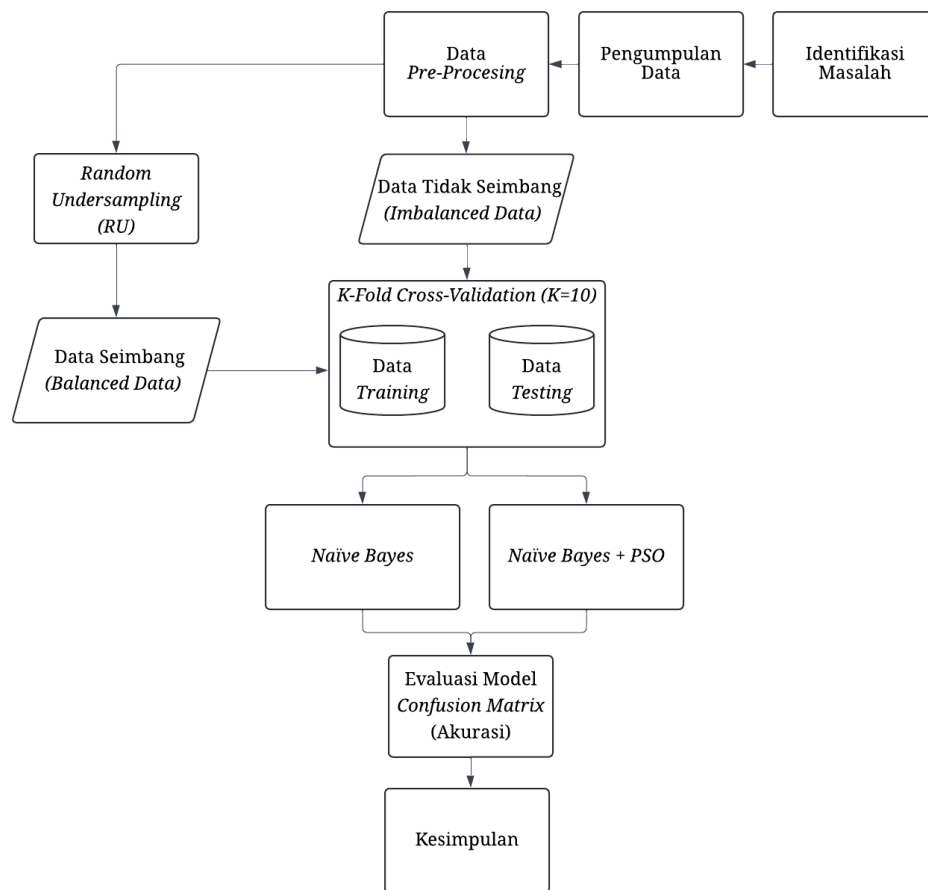
Particle Swarm Optimization (PSO) adalah algoritma optimasi yang terinspirasi oleh perilaku sosial dari kawanan burung atau ikan, di mana setiap partikel dalam *swarm* berusaha menemukan solusi optimal dengan memanfaatkan informasi dari partikel lain [7]. Dengan mengintegrasikan *PSO* ke dalam proses *undersampling*, diharapkan dapat diperoleh subset data yang tidak hanya seimbang tetapi juga representatif, sehingga meningkatkan kemampuan prediksi model *Naive Bayes* terhadap kejadian tanah longsor [8]. Optimalisasi menggunakan *PSO-RU* bertujuan mengatasi *class imbalance*, sehingga hasil prediksi lebih akurat dan tidak bias. Dengan prediksi akurat, proses mitigasi dapat lebih tepat sasaran karena masyarakat dan instansi terkait dapat mempersiapkan langkah pencegahan di area rawan longsor. Penelitian Rachman dan Handayani menunjukkan bahwa penggunaan algoritma *Naive Bayes* dalam konteks klasifikasi dapat memberikan hasil yang baik, namun perlu dioptimalkan untuk mencapai performa yang lebih baik [9]. Penelitian Kamaruddin menunjukkan bahwa penggunaan *PSO* dalam mengoptimalkan algoritma *Naive Bayes* dapat memberikan hasil yang lebih baik dalam klasifikasi data yang tidak seimbang[10].

Dalam penelitian ini, akan mengeksplorasi bagaimana penggabungan *Naive Bayes* dengan *PSO* dapat memberikan solusi yang lebih baik untuk masalah *class imbalance*. Dengan memanfaatkan kekuatan *PSO* dalam optimasi dan kemampuan *Naive Bayes* dalam klasifikasi, penelitian ini bertujuan untuk mengembangkan model yang tidak hanya lebih akurat tetapi juga lebih adaptif dalam menangani data dengan distribusi yang tidak merata dan berharap dapat mengembangkan model yang lebih akurat dan efisien untuk menangani data yang tidak seimbang. Pada penelitian ini data yang tidak seimbang akan di *balance* menggunakan teknik *Random Undersampling (RU)*. Menurut penelitian yang dilakukan oleh Qadrini Qadrini membahas penggunaan teknik *undersampling* untuk meningkatkan performa model klasifikasi. Dalam konteks bencana alam, *RU* dapat digunakan untuk mengurangi jumlah data dari kelas mayoritas, sehingga model *Naive Bayes* dapat lebih fokus pada kelas minoritas yang mungkin lebih relevan dalam konteks bencana. Penelitian ini menunjukkan bahwa dengan mengurangi jumlah data dari kelas mayoritas, model dapat lebih efektif dalam mengenali pola dari kelas minoritas[11].

Secara keseluruhan, penelitian ini bertujuan untuk memberikan kontribusi terhadap pemahaman yang lebih baik tentang bagaimana algoritma klasifikasi dapat dioptimalkan untuk menangani masalah *class imbalance*, serta mengeksplorasi potensi kombinasi antara *Naive Bayes* dan *PSO* dalam konteks ini. Oleh karena itu, kombinasi dengan *PSO* diharapkan dapat meningkatkan performa model dengan menyesuaikan parameter dan memilih fitur secara optimal. Dengan demikian, hasil dari penelitian ini diharapkan dapat memberikan wawasan baru dan aplikasi praktis dalam bidang *data mining* dan klasifikasi.

2. METODOLOGI PENELITIAN

Penelitian ini melibatkan serangkaian langkah-langkah yang akan dilaksanakan untuk mencapai tujuan penelitian. Langkah-langkah ini dimulai dari identifikasi masalah, pengumpulan data, analisis data dan berakhir pada evaluasi dan kesimpulan. Gambar 1 berikut adalah urutan langkah-langkah dari penelitian ini:



Gambar 1. Alur Penelitian

Pada Gambar 1 dibawah menunjukkan proses dimulai dengan pengumpulan data dan *pre-processing* untuk membersihkan serta menyiapkan data. Setelah itu, data dianalisis dan ditemukan bahwa jumlah sampel dalam setiap kelas tidak seimbang. Untuk mengatasinya, digunakan metode *Random Undersampling (RU)* agar jumlah data di setiap kelas menjadi seimbang. Data yang sudah seimbang kemudian dibagi menggunakan metode *K-Fold Cross-Validation (K=10)* untuk meningkatkan keakuratan model. Selanjutnya, dilakukan klasifikasi menggunakan dua metode: *Naïve Bayes* sebagai model dasar dan *Naïve Bayes* yang dioptimasi dengan *PSO (Particle Swarm Optimization)* untuk meningkatkan performa. Hasil dari kedua model dibandingkan menggunakan *Confusion Matrix*, yang mengukur tingkat akurasi masing-masing metode.

2.1 Identifikasi masalah

Identifikasi masalah merupakan tahap awal dalam penelitian untuk mengetahui masalah yang terjadi dan juga sebagai pemandu seluruh proses dalam penelitian ini. Masalah utama dalam penelitian ini bersangkutan dengan objek penelitiannya yaitu mengklasifikasi data bencana tanah longsor Kota Samarinda.

2.2 Pegumpulan Data

Pegumpulan data adalah tahap kedua dalam penelitian untuk mengumpulkan data yang diperlukan dalam penelitian ini untuk mencari data *imbalance* berasal dari data tanah longsor di Kota Samarinda yang didapatkan dari Badan Penanggulangan Bencana Daerah (BPBD) dan Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Kota Samarinda periode tahun 2022-2023.

2.3 Data Pre-prosesing

Setelah pengumpulan data BMKG dan BPBD selesai, langkah selanjutnya adalah melakukan pra-pemrosesan data, seperti *Data Integration*, *Data Selection*, *Data Cleaning* dan *Data Balancing*, agar data tersebut siap untuk diproses pada tahap permodelan.

2.4 Penerapan *Random Undersampling*

Langkah selanjutnya adalah menerapkan teknik *random undersampling*. Teknik ini bertujuan untuk mengurangi jumlah contoh dari kelas mayoritas untuk mencapai keseimbangan dengan kelas minoritas. Penerapan *random undersampling* dalam bencana tanah longsor juga dapat berkontribusi pada pengembangan komunitas yang tangguh

terhadap bencana. Dengan meningkatkan akurasi model prediksi, masyarakat dapat lebih siap menghadapi risiko tanah longsor melalui pendidikan dan pelatihan yang tepat [12].

2.5 Pembagian Data

Metode ini bertujuan untuk memberikan penilaian yang lebih akurat yang dapat difungsikan terhadap performa model dalam mengklasifikasi data bencana tanah longsor di Kota Samarinda. *Cross-validation* dapat membantu meningkatkan efektivitas *Particle Swarm Optimization (PSO)* karena teknik ini memastikan bahwa model yang dioptimalkan memiliki kemampuan generalisasi yang baik terhadap data yang tidak terlihat selama proses pelatihan.

2.6 Penerapan Naive Bayes

Setelah data dibagi, langkah selanjutnya adalah menerapkan algoritma *Naive Bayes*. Algoritma ini dipilih karena kesederhanaannya dan kemampuannya untuk menangani data dengan dimensi tinggi serta *class imbalance*. Dalam penelitian ini, model *Naive Bayes* akan diterapkan pada data pelatihan untuk membangun model klasifikasi. Hasil dari penerapan *Naive Bayes* akan dievaluasi menggunakan metrik *accuracy*.

2.7 Penerapan Particle Swarm Optimization dengan Naive Bayes

Penerapan *Particle Swarm Optimization (PSO)* dengan *Naive Bayes* pada tahap ini *naive bayes* akan kombinasi dengan *particle swarm optimization*. Setelah model dilatih, evaluasi dilakukan untuk mengukur akurasi dan efektivitas model dalam memprediksi tanah longsor. Penerapan *PSO* dan *Naive Bayes* dalam prediksi tanah longsor tidak hanya memberikan pendekatan yang inovatif dalam analisis data, tetapi juga berkontribusi pada upaya mitigasi bencana yang lebih efektif.

2.8 Evaluasi Model

Evaluasi model adalah tahap akhir setelah pembentukan model dengan menggunakan data pengujian yang telah dibagi sebelumnya. Evaluasi yang digunakan pada penelitian ini adalah performa *accuracy* dan pengujian dilakukan dengan teknik *confusion matrix*. Dalam penelitian yang dilakukan oleh Kamil, evaluasi akurasi menggunakan *confusion matrix* menunjukkan bahwa metode ekstraksi citra untuk pemetaan bencana tanah longsor memperoleh akurasi tertinggi sebesar 73,53% [13].

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

3. HASIL DAN PEMBAHASAN

3.1 Hasil Data Pre-prosesing

Penelitian ini dilakukan dengan menggunakan data tanah longsor di Kota Samarinda yang didapatkan dari Badan Penanggulangan Bencana Daerah (BPBD) dan Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Kota Samarinda periode tahun 2022-2023. Pada tahap Data *Pre-processing*, proses ini dilakukan untuk memastikan bahwa data yang digunakan telah bersih dan siap diproses lebih lanjut. *Pre-processing* terdiri dari beberapa langkah berikut:

3.1.1 Hasil Data Integration

Hasil data *integration* Proses ini menghasilkan satu dataset dengan total 20 fitur. Proses ini bertujuan untuk mendapatkan dataset berkualitas sebelum memasuki tahap klasifikasi. Tabel 1 berikut adalah hasil penggabungan Data BPBD dan BMKG:

Tabel 1. Penggabungan Data BPBD dan BMKG

No	Fitur	Tipe Data
1	Tanggal	<i>date</i>
2	Jam Kejadian	<i>string</i>
3	Jenis Bencana	<i>string</i>
4	Lokasi Wilayah	<i>string</i>
5	Luas Area M2	<i>string</i>
6	Objek Terkena Bencana	<i>string</i>
7	Korban	<i>numeric</i>
8	Jumlah Jiwa	<i>numeric</i>
9	Kerugian	<i>string</i>
10	Keterangan	<i>string</i>
11	Temperatur-maksimum (Tx)	<i>numeric</i>
12	Temperatur-minimum (Tn)	<i>numeric</i>
13	Temperatur-rata-rata (Tavg)	<i>numeric</i>

14	Kelembaban-rata-rata (RH_avg)	numeric
15	Curah-hujan (RR)	numeric
16	Lamanya-penyinaran-matahari (ss)	numeric
17	Kecepatan-angin-maksimum (ff_x)	numeric
18	Arah-angin-maksimum (ddd_x)	numeric
19	Kecepatan-angin-rata-rata (ff_avg)	numeric
20	Arah-angin-terbanyak (ddd_car)	numeric

3.1.2 Hasil Data Selection

Pada tahap data *selection*, fitur yang dipilih berdasarkan relevansinya terhadap penyebab tanah longsor, sementara fitur yang tidak relevan dihilangkan. Kemudian data BPBD tersisa dicocokkan diubah menjadi “terjadi tanah longsor” sebagai kelas atau label. Kemudian fitur lainnya yang terpilih mencakup keseluruhan fitur yang ada pada data BMKG. Fitur terkait data tanah longsor dengan angka 0 dan 1.

Tabel 2. Hasil Data Selection

No	Tanggal	Tn	Tx	Tavg	RH_avg	RR	ss	ff_x	ddd_x	ff_avg	ddd_car	Terjadi-longsor
0	01-01-2022	26.0	33.6	29.0	74.0	0.0	--	4.0	100.0	3.0	NE	0
1	02-01-2022	23.6	33.6	28.5	76.0	0.0	6.2	4.0	100.0	2.0	E	0
...
728	30-12-2023	24.2	32.6	28.3	82.0	--	10.4	4.0	60.0	1.0	NE	0
729	31-12-2023	24.6	32.4	28.3	84.0	0.0	6.9	4.0	90.0	2.0	E	0

Pada Tabel 2 menunjukkan fitur terkait data tanah longsor dengan angka 0 dan 1. Angka 0 menunjukkan tidak terjadinya longsor, sedangkan angka 1 menunjukan terjadinya longsor.

3.1.3 Hasil Data Cleaning

Pada tahap *cleaning* dilakukan proses pembersihan data dengan menghilangkan atau menghapus baris yang memiliki setidaknya satu nilai yang kosong dari data tanah longsor yang berjumlah 729 baris.

Tabel 3. Hasil Data Cleaning

No	Tanggal	Tn	Tx	Tavg	RH_avg	RR	ss	ff_x	ddd_x	ff_avg	ddd_car	Terjadi-longsor
1	02-01-2022	23.6	33.6	28.5	76.0	0.0	6.2	4.0	100.0	2.0	E	0
2	03-01-2022	25.1	30.4	27.2	86.0	3.9	6.9	4.0	80.0	2.0	E	0
...
594	29-12-2023	23.7	32.7	28.7	77.0	3.5	5.9	5.0	60.0	2.0	NE	1
595	31-12-2023	24.6	32.4	28.3	84.0	0.0	6.9	4.0	90.0	2.0	E	0

Pada Tabel 3 menunjukkan hasil proses setelah pembersihan data, baris yang mempunyai minimal satu nilai kosong dihapus. Pada proses data cleaning ini, pengurangan jumlah data yang sebelumnya dari 729 data menjadi 595 data.

3.1.4 Hasil Data Transformation

Pada tahap *transformation* bertujuan untuk mengubah format data dari *string* menjadi *numeric* agar mempermudah algoritma untuk menganalisis data. Proses ini menggunakan *LabelEncoder* fitur yang akan diubah dari *string* ke *numeric* yaitu fitur arah angin terbanyak (ddd_car), yang berisi simbol arah mata angin.

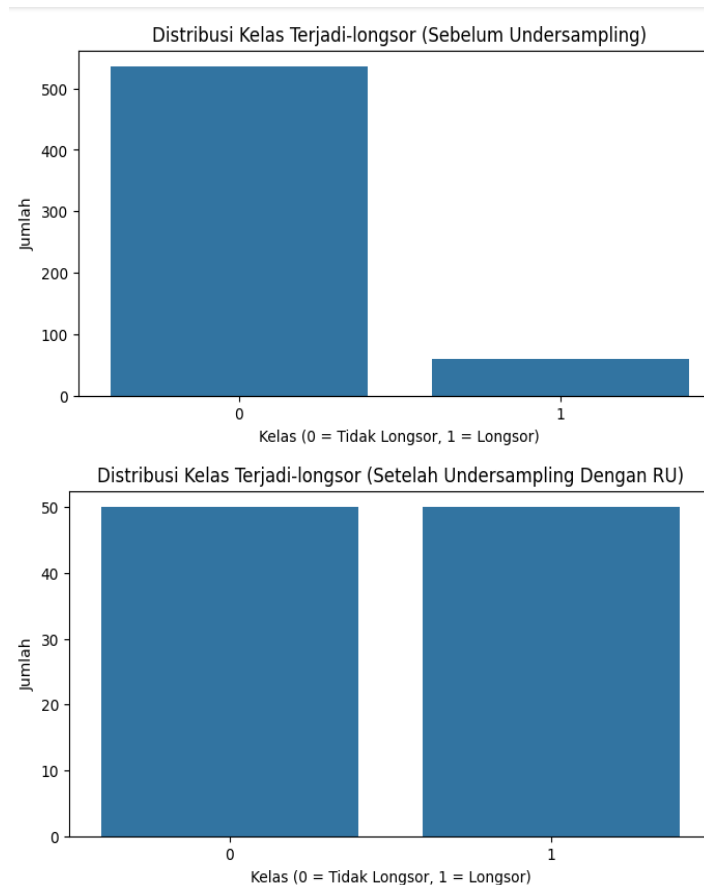
Tabel 4. Hasil Data Transformation

No	(sebelum) ddd_car	(sesudah) ddd_car
1	E	1
2	E	1
3	NE	3
4	E	1
...
593	C	0
594	NE	3
595	E	1

Pada Tabel 4 menunjukkan perubahan data sebelum dan sesudah transformasi, perubahan simbol-simbol arah angin terbanyak sebelumnya berupa kategorik (*string*) diubah menjadi bentuk angka (*numeric*).

3.1.5 Hasil Data *Balancing*

Setelah melalui tahap diatas, data siap untuk proses *balancing* agar mengatasi masalah *class imbalance*. Pada tahap *balancing* bertujuan memastikan bahwa data memiliki distribusi kelas yang seimbang atau proporsional, terutama jika digunakan untuk model klasifikasi. Pada tahap *balancing* penelitian ini menggunakan teknik *Random Undersampling*, teknik ini diterapkan untuk menyeimbangkan jumlah sampel antara kelas mayoritas (tidak terjadi longsor) dan kelas minoritas (terjadi longsor). Metode *undersampling* bertujuan mengurangi jumlah data dari kelas mayoritas dengan menghapus sebagian data untuk menyeimbangkan proporsi.



Gambar 2. Distribusi Kelas Sebelum Dan Setelah *Balancing*

Pada Gambar 2 menunjukkan perbedaan jumlah kelas yang dikategorikan angka 0 adalah (tidak terjadi longsor) dan berjumlah 426 data sampel, sedangkan kategori angka 1 (terjadi longsor) berjumlah 50 data sampel. Sedangkan setelah proses *RU* perbedaan jumlah kelas yang sudah seimbang antara kelas mayoritas dan kelas minoritas, jumlah kelas kategori tidak terjadi longsor dan terjadi longsor masing-masing berjumlah 50 data sampel.

3.2 Hasil Pemodelan Dan Evaluasi

Pada tahap ini menunjukkan hasil akurasi yang diperoleh dari pengujian model algoritma pembelajaran mesin *Naïve Bayes* yang dikombinasikan dengan metode optimasi *Particle Swarm Optimization (PSO)* dan menggunakan data yang sudah diseimbangkan oleh *Random Undersampling (RU)*. Menggunakan teknik *10-Fold Cross Validation* dan *confusion matrix* untuk mengukur akurasi. Setiap model akan diuji pada dataset sebelum dan sesudah *balancing* dan akan dievaluasi hasil akurasinya dengan menggunakan *confusion matrix* untuk memastikan keakuratan dan performa model secara menyeluruh.

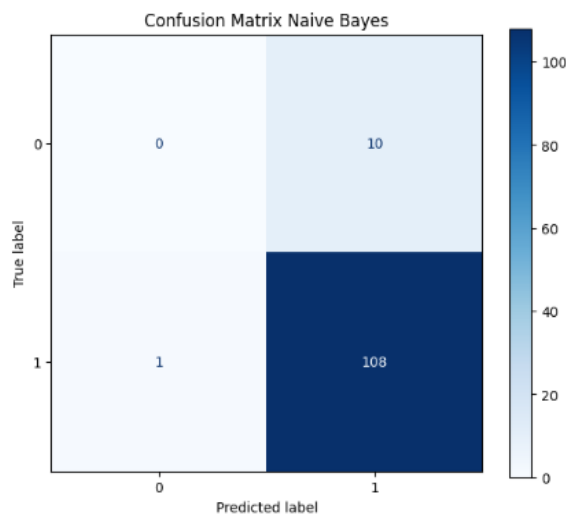
3.2.1 Pemodelan *Naive Bayes* Tanpa *RU*

Tabel 5. Hasil Pengujian *10-Fold Cross Validation Naive Bayes* Tanpa *RU*

Fold	Akurasi (%)
1	89.58%
2	89.58%
3	89.58%

4	89.58%
5	87.50%
6	87.50%
7	87.23%
8	85.11%
9	82.98%
10	87.23%
Rata-rata	87.59%

Pada Tabel 5 menunjukkan hasil evaluasi kinerja algoritma *Naive Bayes* tanpa penerapan *Random Undersampling (RU)* menggunakan metode *10-Fold Cross Validation* dan menghasilkan nilai rata-rata akurasi keseluruhan 87.59%



Gambar 3. Visual *Confusion Matrix Naive Bayes*

Tabel 6. Perhitungan Rata-rata *Confusion Matrix Naive Bayes*

True Label		Predicted Label	
TN	0	FN	1
TP	108	FP	10

$$Accuracy = \frac{108+0}{(108+0+10+1)} \times 100 = \frac{108}{119} \times 100 = 90.75\%$$

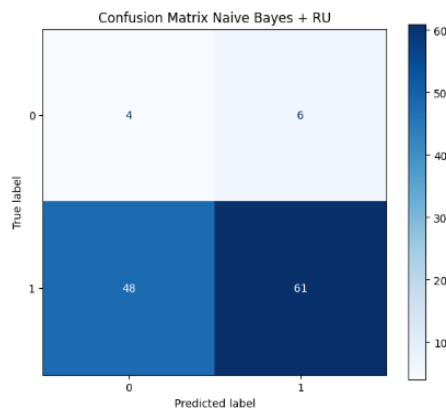
Pada Tabel 6 menunjukkan hasil evaluasi kinerja algoritma *Naive Bayes* dengan menggunakan rata-rata nilai dari *confusion matrix* dan menghasilkan nilai rata-rata akurasi keseluruhan 90.75%.

3.2.2 Pemodelan *Naive Bayes* Dengan *RU*

Tabel 7. Hasil Pengujian *10-Fold Cross Validation Naive Bayes* Dengan *RU*

Fold	Akurasi (%)
1	60%
2	20%
3	30%
4	40%
5	40%
6	60%
7	50%
8	60%
9	30%
10	60%
Rata-rata	45%

Pada Tabel 7 menunjukkan hasil evaluasi kinerja algoritma *Naive Bayes* tanpa penerapan *Random Undersampling (RU)* menggunakan metode *10-Fold Cross Validation* dan menghasilkan nilai rata-rata akurasi keseluruhan 45%.



Gambar 4. Visual *Confusion Matrix Naive Bayes Dengan RU*

Tabel 8. Perhitungan Rata-rata *Confusion Matrix Naive Bayes Dengan RU*

True Label		Predicted Label	
TN	4	FN	48
TP	61	FP	6

$$Accuracy = \frac{61+4}{(61+4+6+48)} \times 100 = \frac{65}{119} \times 100 = 54.62\%$$

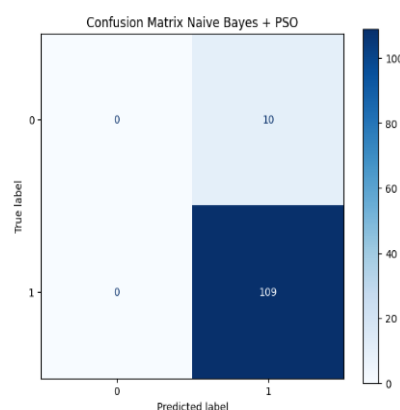
Pada Tabel 8 menunjukkan hasil evaluasi kinerja algoritma *Naive Bayes* dengan *Random Undersampling (RU)* menggunakan rata-rata nilai dari *confusion matrix* dan menghasilkan nilai rata-rata akurasi keseluruhan 54.62%

3.2.3 Pemodelan *Naive Bayes + PSO Tanpa RU*

Tabel 9. Hasil Pengujian *10-Fold Cross Validation Naive Bayes + PSO Tanpa RU*

Fold	Akurasi (%)
1	89.58%
2	89.58%
3	89.58%
4	89.58%
5	89.58%
6	89.58%
7	89.36%
8	89.36%
9	89.36%
10	89.36%
Rata-rata	89.49%

Pada Tabel 9 menunjukkan hasil evaluasi kinerja algoritma *Naive Bayes* tanpa penerapan *Random Undersampling (RU)* menggunakan metode *10-Fold Cross Validation* dan menghasilkan nilai rata-rata akurasi keseluruhan 89.49%.



Gambar 5. Visual *Confusion Matrix Naive Bayes + PSO Tanpa RU*

Tabel 10. Perhitungan Rata-rata *Confusion Matrix Naive Bayes + PSO Tanpa RU*

<i>True Label</i>		<i>Predicted Label</i>	
<i>TN</i>	0	<i>FN</i>	0
<i>TP</i>	109	<i>FP</i>	10

$$Accuracy = \frac{109+0}{(109+0+10+0)} \times 100 = \frac{109}{119} \times 100 = 91.59\%$$

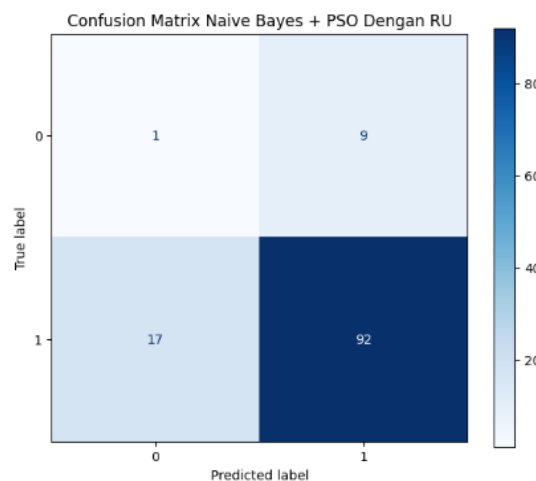
Pada Tabel 10 menunjukkan hasil evaluasi kinerja algoritma *Naive Bayes + PSO* dengan menggunakan rata-rata nilai dari *confusion matrix* dan menghasilkan nilai rata-rata akurasi keseluruhan 91.59%.

3.2.4 Pemodelan *Naive Bayes + PSO Dengan RU*

Tabel 11. Hasil Pengujian *10-Fold Cross Validation Naive Bayes + PSO Dengan RU*

Fold	Akurasi (%)
1	50%
2	40%
3	40%
4	70%
5	70%
6	30%
7	50%
8	50%
9	50%
10	50%
Rata-rata	50%

Pada Tabel 11 menunjukkan hasil evaluasi kinerja algoritma *Naive Bayes* tanpa penerapan *Random Undersampling (RU)* menggunakan metode *10-Fold Cross Validation* dan menghasilkan nilai rata-rata akurasi keseluruhan 50%.



Gambar 6. *Visual Confusion Matrix Naive Bayes + PSO Dengan RU*

Tabel 12. Perhitungan Rata-rata *Confusion Matrix Naive Bayes + PSO Dengan RU*

<i>True Label</i>		<i>Predicted Label</i>	
<i>TN</i>	1	<i>FN</i>	17
<i>TP</i>	92	<i>FP</i>	9

$$Accuracy = \frac{92+1}{(92+1+9+17)} \times 100 = \frac{93}{119} \times 100 = 78.15\%$$

Pada Tabel 12 menunjukkan hasil evaluasi kinerja algoritma *Naive Bayes + PSO* dengan *RU* menggunakan rata-rata nilai dari *confusion matrix* dan menghasilkan nilai rata-rata akurasi keseluruhan 78,15%.

3.3 Perbandingan Hasil Evaluasi

Tabel 13. Perbandingan Hasil Evaluasi *10-Fold Cross Validation* Pada Kedua Model Tanpa *RU*

<i>Fold</i>	<i>Naive Bayes</i>	<i>Naive Bayes + PSO</i>	Perubahan Hasil Model	Status
1	89.58%	89.58%	0%	Tetap
2	89.58%	89.58%	0%	Tetap
3	89.58%	89.58%	0%	Tetap
4	89.58%	89.58%	0%	Tetap
5	87.50%	89.58%	2.08%	Naik
6	87.50%	89.58%	2.08%	Naik
7	87.23%	89.36%	2.13%	Naik
8	85.11%	89.36%	4.25%	Naik
9	82.98%	89.36%	6.38%	Naik
10	87.23%	89.36%	2.13%	Naik
Rata- Rata	87.59%	89.49%	1.9%	Naik

Pada Tabel 13 menunjukkan setelah perbandingan kedua model memiliki penilaian yang tinggi dan mengalami peningkatan dengan model *Naive Bayes* memiliki nilai rata-rata 87.59%, sedangkan rata-rata akurasi dari model *Naive Bayes + PSO* adalah 89.49%, dengan rata-rata peningkatan sebesar 1.9%.

Tabel 14. Perbandingan Hasil Evaluasi *Confusion Matrix* Pada Kedua Model Tanpa *RU*

<i>Naive Bayes</i>	<i>Naive Bayes + PSO</i>	Perubahan Hasil Model	Status
90.75%	91.59%	+0.84%	Naik

Pada Tabel 14 menunjukkan perbandingan akurasi antara model *Naive Bayes* tanpa optimasi dan model *Naive Bayes* yang dioptimasi dengan *Particle Swarm Optimization (PSO)*. Model *Naive Bayes* tanpa *PSO* menghasilkan akurasi sebesar 90.75%, sementara model yang menggunakan *PSO* menghasilkan akurasi yang lebih tinggi, yaitu 91.59%.

Tabel 15. Perbandingan Hasil Evaluasi *10-Fold Cross Validation* Pada Kedua Model Dengan *RU*

<i>Fold</i>	<i>Naive Bayes</i>	<i>Naive Bayes + PSO</i>	Perubahan Hasil Model	Status
1	60%	50%	-10%	Turun
2	20%	40%	20%	Naik
3	30%	40%	10%	Naik
4	40%	70%	30%	Naik
5	40%	70%	30%	Naik
6	60%	30%	-30%	Turun
7	50%	50%	0%	Tetap
8	60%	50%	-10%	Turun
9	30%	50%	20%	Naik
10	60%	50%	-10%	Turun
Rata- Rata	45%	50%	0.5%	Naik

Pada Tabel 15 menunjukkan perbandingan hasil akurasi antara model *Naive Bayes* dan model *Naive Bayes* yang dioptimasi menggunakan *Particle Swarm Optimization (PSO)* dengan penerapan teknik *Random Undersampling (RU)* pada dataset tanah longsor. Secara rata-rata, model *Naive Bayes* dengan *RU* memiliki akurasi 45%, sementara model *Naive Bayes + PSO* dengan *RU* memiliki rata-rata akurasi sebesar 50%. Hal ini menunjukkan adanya peningkatan rata-rata akurasi sebesar 0.5%, meskipun peningkatan ini tergolong kecil.

Tabel 16. Perbandingan Hasil Evaluasi *Confusion Matrix* Pada Kedua Model Dengan *RU*

<i>Naive Bayes</i>	<i>Naive Bayes + PSO</i>	Perubahan Hasil Model	Status
54.62%	78.15%	+23.53%	Naik

Pada Tabel 16 menunjukkan perbandingan akurasi antara model *Naive Bayes* tanpa optimasi dan model *Naive Bayes* yang dioptimasi dengan *Particle Swarm Optimization (PSO)* pada data yang telah diolah. Model *Naive Bayes*

tanpa *PSO* menghasilkan akurasi sebesar 54.62%, sedangkan setelah dioptimasi dengan *PSO*, akurasinya meningkat menjadi 78.15%. Peningkatan ini menunjukkan adanya perbaikan kinerja yang signifikan sebesar +23.53%.

4. KESIMPULAN

Penelitian ini bertujuan untuk meningkatkan akurasi prediksi data bencana tanah longsor yang memiliki masalah *class imbalance*. Berdasarkan hasil penelitian, integrasi algoritma *Naive Bayes* dengan teknik *Particle Swarm Optimization (PSO)* menunjukkan peningkatan akurasi yang signifikan dalam klasifikasi data tanah longsor di Kota Samarinda. Pada dataset tanpa penerapan *Random Undersampling (RU)*, model *Naive Bayes* memiliki rata-rata akurasi sebesar 87.59%, sementara kombinasi *Naive Bayes + PSO* menghasilkan rata-rata akurasi yang lebih tinggi, yaitu 89.49%, dengan peningkatan rata-rata sebesar 1.9%. Sebaliknya, pada dataset yang menggunakan teknik *Random Undersampling (RU)* untuk menyeimbangkan data, rata-rata akurasi model *Naive Bayes* sebesar 45% meningkat menjadi 50% setelah dioptimasi dengan *PSO*, meskipun peningkatannya relatif kecil, yakni 0.5%. Peningkatan algoritma *Naive Bayes* dengan optimasi menggunakan *PSO* menunjukkan peningkatan performa *fold* dengan nilai rata-rata 89,49%, dibandingkan *Naive Bayes* saja yang memiliki nilai rata-rata 87.59% dengan peningkatan rata-rata sebesar +1,9%. Peningkatan pada akurasi *confusion matrix* dengan kombinasi *Naive bayes + PSO* tanpa *RU* menghasilkan nilai akurasi dari 90.75% meningkat jadi 91.59%, sedangkan kombinasi *Naive bayes + PSO* dengan *RU* terjadi peningkatan signifikan akurasi rata-rata yang dari 54.62% menjadi 78.15% terjadi peningkatan +23.53%. Pada *confusion matrix* dan beberapa *fold*, menunjukkan bahwa *PSO* berhasil mengoptimalkan parameter model tanpa mengurangi informasi dalam data. Hasil ini menunjukkan bahwa *PSO* mampu mengoptimalkan parameter model *Naive Bayes* secara efektif, terutama pada dataset yang tidak mengalami pengurangan variasi data akibat *undersampling*. Secara keseluruhan, kombinasi *Naive Bayes* dan *PSO* memberikan performa yang lebih baik dibandingkan *Naive Bayes* saja, terutama dalam pengolahan dataset yang kompleks dan tidak seimbang. Hal ini mendukung upaya mitigasi bencana tanah longsor melalui prediksi yang lebih akurat, sehingga memungkinkan perencanaan tindakan preventif seperti pengungsian dini dan pembangunan infrastruktur di daerah rawan longsor secara lebih efektif.

REFERENCES

- [1] M. Miftakhul Huda, M. Zainul Ikhwan, dan M. Insan Rendra, "Pendekatan Analisis Spasial untuk Deteksi Kerawanan Longsor di Kecamatan Kedugadem Kabupaten Bojonegoro," *Jurnal Informatika dan Sistem Informasi*, vol. 16, no. 1, 2024, doi: 10.37424/informasi.v16i1.297.
- [2] S. Ujang Effendi, K. Apriani, D. Syavani, N. Khairani, dan R. Aprianti, "Pengaruh Pengetahuan dan Sikap Terhadap Kesiapsiagaan Masyarakat Dalam Menghadapi Bencana Tanah Longsor," *Jurnal Sains Kesehatan*, vol. 30, no. 1, 2023, DOI: <https://doi.org/10.37638/jsk.30.1.16-25>.
- [3] L. Xia, J. Shen, T. Zhang, G. Dang, dan T. Wang, "GIS-Based Landslide Susceptibility Modeling Using Data Mining Techniques," *Frontiers in Earth Science*, vol. 11, 2023, DOI: 10.3389/feart.2023.1187384.
- [4] E. Mardiani et al., "Classification of Potential Tsunami Disaster Due to Earthquakes in Indonesia Based on Machine Learning," *International Journal of Software Engineering and Computer Science (IJSECS)*, vol. 4, no. 1, pp. 13–23, Apr. 2024, DOI: 10.35870/ijsecs.v4i1.2084.
- [5] V. Chamola, V. Hassija, S. Gupta, A. Goyal, M. Guizani, dan B. Sikdar, "Disaster and Pandemic Management Using Machine Learning: A Survey," *IEEE Internet of Things Journal*, Vol 8 No21, 2020, DOI: 10.1109/JIOT.2020.3044966.
- [6] A. Karima dan T. Azhima Yoga Siswa, "Prediksi Kinerja Mahasiswa Dalam Perkuliahan Berbasis Learning Management System Menggunakan Algoritma Naive Bayes," *Jurnal Ilmiah Komputer STMIK Banjarbaru*, Vol 18 No. 2, Jul. 2022, DOI: 10.35889/progresif.v18i2.922.
- [7] E. I. Goma, S. Sunimbar, dan I. S. Angin, "Analisis Geologi Kejadian Longsor di Desa Wolotolo Kecamatan Detusoku Kabupaten Ende," *JPG (Jurnal Pendidikan Geografi)*, vol. 9, no. 2, Sep. 2022, DOI: 10.20527/jpg.v9i2.13471.
- [8] I. Madani et al., "Analisis Spasial Tingkat Kerawanan Tanah Longsor di Kecamatan Wagir, Kabupaten Malang Melalui Sistem Informasi Geografis," *Jurnal Geosaintek*, vol. 9, no. 2, p. 80, Aug. 2023, DOI: 10.12962/j25023659.v9i2.17431.
- [9] R. Rachman, R. N. Handayani, and I. Artikel, "Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM," *JURNAL INFORMATIKA*, vol. 8, no. 2, 2021, DOI: 10.31294/ji.v8i2.10494.
- [10] A. S. Kamaruddin, M. F. Hadrawi, Y. B. Wah, dan S. Aliman, "An Evaluation of Nature-Inspired Optimization Algorithms and Machine Learning Classifiers for Electricity Fraud Prediction," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 32, no. 1, pp. 468–477, Oct. 2023, DOI: 10.11591/ijeecs.v32.i1.pp468-477.
- [11] L. Qadrini, "Undersampling dan K-Fold Random Forest Untuk Klasifikasi Kelas Tidak Seimbang," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 4, Mar. 2023, DOI: 10.47065/bits.v4i4.3141.
- [12] E. Nugroho et al., "Manajemen Dan Pengurangan Risiko Bencana Melalui Pengembangan Desa Tangguh Bencana (DESTANA)," *Bookchapter Kesehatan Masyarakat Universitas Negeri Semarang*, no. 3, pp. 92–113, Feb. 2023, DOI: 10.15294/km.v1i3.98.
- [13] N. N. Kamil dan D. Syam'ani, "Pemertaan Bencana Tanah Longsor di DTA Barabai Menggunakan Metode Ekstraksi NDBI, SAVI, dan CBI," *Jurnal Sylva Scientiae* Vol. 07 No. 2, Apr. 2024, DOI: 10.20527/jss.v7i2.12316.
- [14] Y. Religia, G. T. Pranoto, dan I. M. Suwancita, "Analysis of the Use of Particle Swarm Optimization on Naive Bayes for Classification of Credit Bank Applications," *JISA (Jurnal Informatika dan Sains)*, Vol. 07 No. 2, Des. 2021, DOI: 10.31326/jisa.v4i2.946.



- [15] Wilandini, D. and Purwantoro, P, “Penerapan algoritma naïve bayes dalam mengklasifikasikan media sosial untuk mengamati trend kuliner,” *J. teknologi terpadu*, vol. 8, no. 1, hlm. 31–39, Jul 2022. DOI: 10.54914/jtt.v8i1.535.
- [16] Sopian Aji, “Analisis Sentiment Masyarakat Menggunakan Penggabungan Algoritma Naive Bayes Dan Particle Swarm Optimization,” *IJCIT (Indonesian Journal on Computer and Information Technology)*, Vol 7 No. 2, 2022. DOI: 10.31294/ijcit.v7i2.14086.
- [17] D. Susanti and T. Wahyuni, “Analisi Potensi Bencana Alam Tanah Longsor Kabupaten Majalengka Menggunakan Algoritma Naive Bayes Clasiffier,” *INFOTECH journal*, vol. 9, no. 2, pp. 299–306, Jul. 2023, DOI: 10.31949/infotech.v9i2.5645.
- [18] J. Dixon and M. Rahman, “Statistical Analysis of Imbalanced Classification with Training Size Variation and Subsampling on Datasets of Research Papers in Biomedical Literature,” *Mach Learn Knowl Extr*, vol. 5, no. 4, pp. 1953–1978, Dec. 2023, DOI: 10.3390/make5040095.
- [19] H. Setiani, A. Sunyoto, A. Nasiri, and U. Amikom, “Metode Naïve Bayes dan Particle Swarm Optimization untuk Klasifikasi Penyakit Jantung”, *EXPLORE*, Vol 12, No. 2, Jul 2022, DIO: 10.35200/explore.v12i2.566.
- [20] J. Prasetya, “Penerapan Klasifikasi Naive Bayes Dengan Algoritma Random Oversampling Dan Random Udersampling Pada Data Tidak Seimbang Cervical Cancer Risk Factors ”, *Leibniz Jurnal Matematika*, vol. 2, no. 2, 2022, DOI: 10.59632/leibniz.v2i2.173.
- [21] H. B. Jatmiko, N. T. Kurniadi, and D. Maulana, “Optimasi Naïve Bayes Dengan Particle Swarm Optimization Untuk Analisis Sentimen Formula E-Jakarta Optimization of Naïve Bayes With Particle Swarm Optimization for Sentimen Analysis of Jakarta E-Prix,” *Journal Automation Computer Information System*, Vol 2 No. 1, Jun 2022, DOI:10.47134/jacis.v2i1.35.
- [22] Y. Liu, J. Huang, R. Xiao, S. Ma, and P. Zhou, “Research on a Regional Landslide Early-Warning Model Based on Machine Learning—A Case Study of Fujian Province, China,” *Forests*, vol. 13, no. 12, Dec. 2022, DOI:10.3390/f13122182.
- [23] 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.
- [24] D. Saputra, W. Irmayani, D. Purwaningtias, and J. Sidauruk, “A Comparative Analysis of C4.5 Classification Algorithm, Naïve Bayes and Support Vector Machine Based on Particle Swarm Optimization (PSO) for Heart Disease Prediction,” *International Journal of Advances in Data and Information Systems*, vol. 2, no. 2, Nov. 2021, DOI: 10.25008/ijadis.v2i2.1221.
- [25] Nurfaiz Fathurrahman Yasien, Felia Yustika, Intan Permatasari, and Muthiah Sari, “Aplikasi Geospasial Untuk Analisis Potensi Bahaya Longsor Menggunakan Metode Weighted Overlay,” *Jurnal Geosains dan Remote Sensing*, vol. 2, no. 1, pp. 33–40, May 2021, DOI: 10.23960/jgrs.2021.v2i1.47.
- [26] M. Khairy, T. M. Mahmoud, and T. Abd El-Hafeez, “The Effect of Rebalancing Techniques on the Classification Performance in Cyberbullying Datasets,” *Research Square*, Vol 36 No Jun 2022, DOI: 10.21203/rs.3.rs-1730456/v1.