

Klasifikasi Risiko Kesehatan Ibu Hamil Menggunakan Random Oversampling Untuk Mengatasi Ketidakseimbangan Data

Riska Aryanti¹, Titik Misriati^{2*}, Rahmat Hidayat³

¹Teknik dan Informatika, Ilmu Komputer, Universitas Bina Sarana Informatika, Jakarta, Indonesia

²Teknik dan Informatika, Sistem Informasi Akuntansi, Universitas Bina Sarana Informatika, Jakarta, Indonesia

³Teknik dan Informatika, Sistem Informasi, Universitas Bina Sarana Informatika, Jakarta, Indonesia

Email: ¹riska.rts@bsi.ac.id, ^{2*}titik.tmi@bsi.ac.id, ³rahmat.rhh@bsi.ac.id

Email Penulis Korespondensi: titik.tmi@bsi.ac.id

Abstrak—Ketidakseimbangan data adalah masalah umum dalam klasifikasi, termasuk dalam klasifikasi risiko kesehatan ibu hamil. Ketidakseimbangan data terjadi ketika jumlah sampel dalam kelas positif jauh lebih sedikit dibandingkan dengan kelas negatif. Hal ini dapat menyebabkan model klasifikasi tidak akurat dan cenderung memprediksi kelas mayoritas. Salah satu cara untuk mengatasi masalah ketidakseimbangan data adalah dengan menggunakan teknik random over sampling. Pada penelitian ini, teknik random oversampling diterapkan untuk mengatasi masalah ketidakseimbangan data dalam klasifikasi risiko kesehatan ibu hamil dan particle swarm optimization (PSO) digunakan untuk pembobotan atribut, memperbaiki hasil dari random oversampling dan meningkatkan kinerja model sehingga hasil penelitian menunjukkan bahwa random oversampling dapat meningkatkan akurasi dan mengurangi kesalahan dalam memprediksi kelas minoritas. Selain itu, teknik PSO juga memberikan kontribusi signifikan dalam meningkatkan akurasi model. Hasil pengujian algoritma random forest menggunakan 10 fold cross validation terhadap risiko kesehatan ibu hamil mempunyai akurasi sebesar 80.77% setelah melalui teknik random oversampling tingkat akurasi mencapai 81,86% dan setelah optimalisasi dengan menggunakan teknik PSO terbukti adanya peningkatan sebesar 2.15% sehingga tingkat akurasi mencapai 82,92%.

Kata Kunci: Klasifikasi; Ketidakseimbangan Data; PSO; Random Oversampling; Risiko Kesehatan Ibu Hamil

Abstract—Data imbalance is a common problem in classification, including in maternal health risk classification. Data imbalance occurs when the number of samples in the positive class is much less than the negative class. Data imbalance can cause the classification model to be inaccurate and tend to predict the majority class. One way to overcome the problem of data imbalance is to use the random oversampling technique. In this study, the random oversampling method is applied to overcome the problem of data imbalance in the classification of maternal health risks. Particle swarm optimization (PSO) is used for attribute weighting, improving the results of random oversampling and model performance. The results show that random oversampling can improve accuracy and reduce errors in predicting minority classes. In addition, the PSO technique also significantly contributed to improving the model's accuracy. The results of testing the random forest algorithm using 10-fold cross-validation on the health risks of pregnant women have an accuracy of 80.77%. After going through the random oversampling technique, the accuracy rate reaches 81.86%, and after optimization using the PSO technique, there is an increase of 2.15%, so the accuracy rate reaches 82.92%.

Keywords: Classification; Imbalance Data; PSO; Random Oversampling; Health Risks for Pregnant Women

1. PENDAHULUAN

Kesehatan ibu hamil merupakan aspek penting yang perlu diperhatikan. Menurut World Health Organization (WHO), setiap tahunnya terdapat sekitar 800 wanita meninggal dunia akibat komplikasi kehamilan dan persalinan [1]. Banyak ibu hamil meninggal karena komplikasi kehamilan karena tidak memiliki cukup informasi tentang perawatan kesehatan ibu selama kehamilan dan pasca kehamilan. Hal ini terutama terjadi di daerah pedesaan dan menengah ke bawah di negara berkembang [2]. Kematian ibu hamil dan melahirkan merupakan permasalahan di bidang kesehatan yang sudah lama terjadi, khususnya di negara-negara berkembang. Faktor utama kematian perempuan adalah kematian pada saat melahirkan. Ibu hamil memerlukan pemantauan terus menerus selama periode kehamilan sampai proses melahirkan supaya dapat diketahui faktor risiko yang dialami selama kehamilan. Meskipun identifikasi faktor risiko, pemeriksaan kehamilan, serta pelayanan rujukan dalam upaya pencegahan komplikasi kehamilan telah dilakukan, kemungkinan komplikasi berat terjadi pada saat proses persalinan. Tidak cukupnya informasi yang diterima oleh masyarakat khususnya yang berada di pedesaan yang terpencil, seperti jarak yang sangat jauh dari rumah sakit atau klinik bersalin sehingga informasi yang diperoleh dari dokter untuk ibu hamil tidak tersampaikan secara jelas. Tentu ini menjadi suatu kendala dalam mengatasi risiko kesehatan pada ibu hamil dan harus mencari solusi untuk mengurangi angka kematian. Paling banyak angka kematian setelah melahirkan adalah pendarahan, dimana banyak faktor dari penyebab pendarahan pada ibu hamil [3]. Tentunya ilmu serta pengetahuan yang menjadi dasar seorang wanita dinyatakan hamil untuk memahami dan tahu betul soal kehamilan.

Untuk mengatasi hal ini harus dipantau selama kehamilan untuk perkembangan janin yang sehat dan untuk memastikan persalinan yang tidak berbahaya [2]. Oleh karena itu, klasifikasi risiko kesehatan ibu hamil menjadi sangat penting untuk mencegah dan mengurangi risiko kesehatan yang serius bagi ibu dan bayi yang dikandungnya. Klasifikasi merupakan proses menemukan model untuk menggambarkan data dengan tujuan model tersebut dapat digunakan prediksi yang belum diketahui dari suatu objek [4]. Klasifikasi ini merupakan salah satu tugas dari data mining dalam mencari pola atau model [5]. Namun, masalah ketidakseimbangan data sering terjadi dalam klasifikasi risiko kesehatan ibu hamil, dimana jumlah sampel dalam kelas positif (misalnya, risiko tinggi) jauh lebih sedikit dibandingkan dengan kelas negatif (misalnya, risiko rendah). Hal ini dapat menyebabkan model klasifikasi tidak akurat dan cenderung memprediksi kelas mayoritas. Untuk mengatasi masalah ketidakseimbangan data, teknik oversampling dapat digunakan. Teknik

oversampling memperbanyak sampel kelas minoritas untuk menyeimbangkan jumlah sampel antara kelas mayoritas dan kelas minoritas [6], [7].

Pada penelitian ini, teknik random oversampling diterapkan untuk mengatasi masalah ketidakseimbangan data dalam klasifikasi risiko kesehatan ibu hamil. Hasil penelitian sebelumnya menunjukkan bahwa teknik random oversampling dapat meningkatkan akurasi klasifikasi dalam kasus ketidakseimbangan data [8]–[10]. Untuk mengatasi masalah ketidakseimbangan data, teknik oversampling dapat digunakan. Salah satu teknik oversampling yang sering digunakan adalah random oversampling. Teknik ini memilih sampel acak dari kelas minoritas untuk digandakan dan ditambahkan ke set data pelatihan. Namun, teknik random oversampling dapat menghasilkan banyak sampel yang tidak signifikan dan mempengaruhi kinerja model klasifikasi. Proses dari teknik oversampling ini dilakukan dengan cara penambahan diulang sampai jumlah data kelas minoritas sama dengan jumlah kelas mayoritas [11].

Oleh karena itu, dalam penelitian ini menggabungkan teknik random oversampling dan Particle Swarm Optimization (PSO) untuk mengatasi masalah ketidakseimbangan data dalam klasifikasi risiko kesehatan ibu hamil. PSO digunakan untuk memperbaiki hasil dari random oversampling dan meningkatkan kinerja model. Dengan mengaplikasikan teknik random oversampling dan PSO, diharapkan hasil klasifikasi dapat menjadi lebih akurat dan lebih signifikan.

2. METODOLOGI PENELITIAN

2.1 Dataset

Dataset diambil dari UCI Machine Learning yang terdiri dari 6 atribut dan 1 label yaitu usia, tekanan darah sistolik, tekanan darah diastolik, gula darah, suhu tubuh, denyut jantung dan tingkat risiko yang dapat dilihat pada tabel 1. Atribut-atribut ini merupakan faktor risiko dan signifikan terhadap kesehatan ibu hamil. Dataset berisi 1.014 data, table 2, yang terdiri dari tiga kelas yaitu kelas risiko tinggi sebanyak 272, risiko sedang sebanyak 336, dan risiko rendah sebanyak 406 seperti pada gambar 1.

Tabel 1. Informasi Dataset

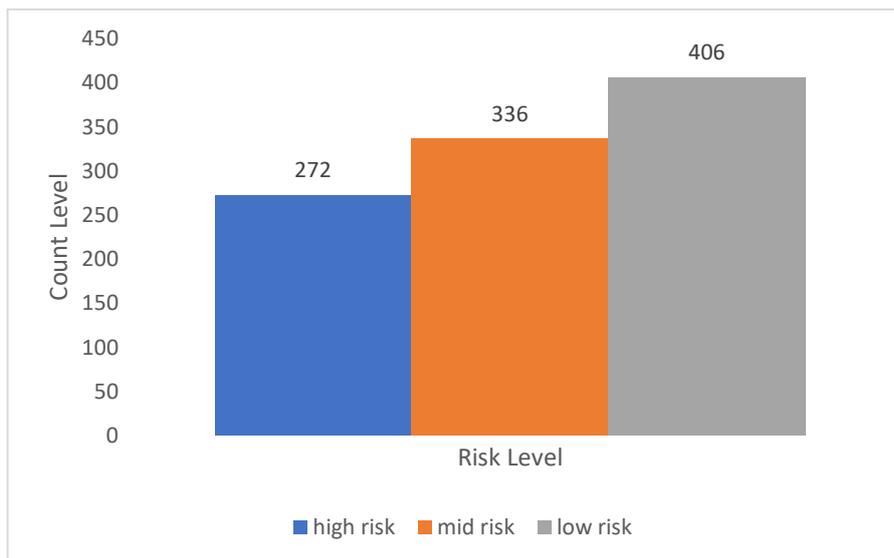
Atribut	Keterangan
Age	Usia ibu hamil
SystolicBP	Nilai atas Tekanan Darah dalam mmHg
DiastolicBP	Nilai bawah Tekanan Darah dalam mmHg
BS	Kadar glukosa darah dalam konsentrasi molar, mmol/L
Body Term	Suhu tubuh ibu hamil
HeartRate	Denyut jantung istirahat normal dalam denyut per menit.
Risk Level	Tingkat risiko

2.2 Preprocessing

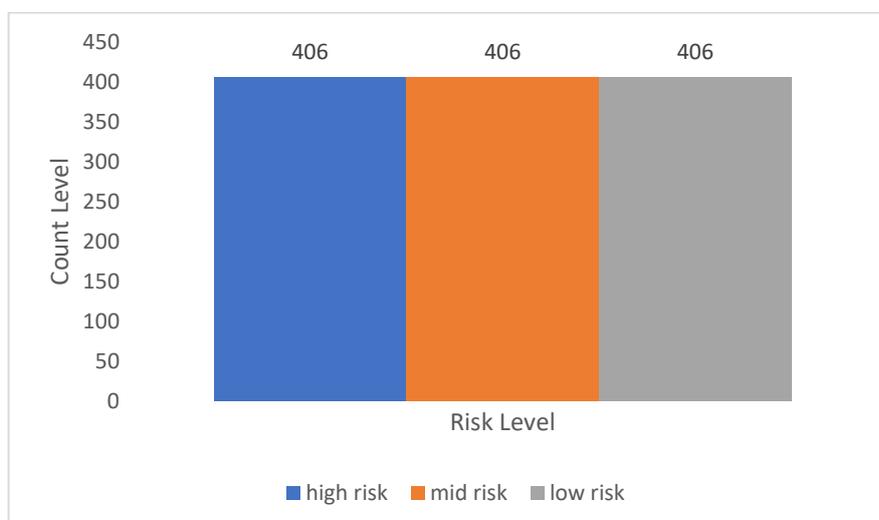
Random oversampling adalah teknik dalam data preprocessing yang digunakan untuk menangani masalah ketidakseimbangan kelas dalam dataset [12]. Proses random oversampling ini dilakukan dengan cara mengulang sampel yang sudah ada secara acak pada kelas minoritas hingga jumlah sampel pada kelas minoritas sama dengan kelas mayoritas [13]. Cara ini memungkinkan untuk menghasilkan dataset yang lebih seimbang dan meningkatkan performa model dalam mengenali kelas minoritas. Resampling dilakukan untuk mencapai hasil terbaik yang diusulkan. Resampling data diterapkan untuk menghasilkan sampel dengan bobot yang sama dan mengukur ketidakpastian tingkat risiko. Random oversampling digunakan untuk menyeimbangkan dataset sehingga semua kelas berisi data yang sama, yaitu 406 data. Dari Gambar 2, dapat dilihat bahwa jumlah sampel untuk setiap kelas seimbang.

Tabel 2. Dataset

No	Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate	RiskLevel
1	25	130	80	15	98	86	high risk
2	35	140	90	13	98	70	high risk
3	29	90	70	8	100	80	high risk
4	30	140	85	7	98	70	high risk
5	35	120	60	06.01	98	76	low risk
6	23	140	80	07.01	98	70	high risk
7	23	130	70	07.01	98	78	mid risk
8	35	85	60	11	102	86	high risk
...
1014	32	120	65	6	101	76	mid risk



Gambar 1. Sebelum Resampling



Gambar 2. Setelah Random Oversampling

2.3 Particle Swarm Optimization

Particle Swarm Optimization adalah salah satu algoritma optimasi yang digunakan dalam masalah optimasi global. Algoritma PSO didasarkan pada prinsip koloni atau kelompok partikel yang bergerak di dalam ruang pencarian solusi untuk mencari solusi yang optimal [14]. Setiap partikel menghitung nilai fungsi tujuan dari posisi saat ini, dan kemudian membandingkannya dengan posisi terbaik yang pernah dicapai oleh partikel itu sendiri dan oleh seluruh kelompok. Kemudian, partikel akan menghitung kecepatan dan posisi baru dengan menggabungkan informasi terbaik dari posisi terbaik pribadi dan posisi terbaik global.

2.3 Cross Validation

Penelitian ini menggunakan 10-fold cross validation yaitu salah satu bentuk dari teknik cross validation yang umum digunakan dalam evaluasi model. Pada 10-fold cross validation, data dibagi menjadi 10 bagian yang sama besar. Kemudian, model dilatih pada 9 bagian dari data dan diuji pada 1 bagian dari data yang tidak digunakan sebelumnya. Proses ini dilakukan sebanyak 10 kali, sehingga setiap bagian dari data digunakan sebagai data uji satu kali. Keuntungan dari menggunakan 10-fold cross validation adalah mencegah overfitting dan meningkatkan kehandalan hasil evaluasi model. Dengan menggunakan teknik ini, model akan diuji pada seluruh data dan menghasilkan kinerja yang lebih objektif. Selain itu, teknik 10-fold cross validation juga memungkinkan penggunaan sebagian besar data untuk pelatihan dan pengujian.

2.4 Random Forest

Random forest merupakan algoritma machine learning yang bekerja dengan cara membuat banyak pohon keputusan secara acak, lalu menggabungkan hasil prediksi dari semua pohon untuk menghasilkan prediksi akhir [12], [15].

2.5 Evaluasi

a. Precision

Precision adalah ukuran seberapa baik model dalam memprediksi kelas positif secara akurat.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

True positive adalah jumlah sampel positif yang diklasifikasikan dengan benar oleh model, sedangkan false positive adalah jumlah sampel negatif yang salah diklasifikasikan sebagai positif oleh model.

b. Recall

Recall adalah ukuran seberapa baik model dalam menemukan semua sampel positif.

recall = true positive / (true positive + false negative)

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

True positive adalah jumlah sampel positif yang diklasifikasikan dengan benar oleh model, sedangkan false negative adalah jumlah sampel positif yang salah diklasifikasikan sebagai negatif oleh model.

c. Akurasi (Accuracy)

Akurasi adalah metrik evaluasi yang mengukur seberapa akurat suatu model dalam mengklasifikasikan data [16]. Akurasi dihitung dengan membandingkan jumlah prediksi yang benar dengan jumlah total prediksi. Persamaan untuk menghitung akurasi adalah:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

TP = true positive, TN = true negative, FP = false positive, FN = false negative

d. Kappa

Kappa adalah matrik evaluasi yang mengukur seberapa baik model dalam memprediksi kelas dengan memperhitungkan kebetulan. Kappa dapat dianggap sebagai ukuran kesepakatan antara prediksi model dengan kelas aktual yang dibandingkan dengan kesepakatan yang diharapkan secara acak. Nilai kappa berkisar dari -1 hingga 1, dimana nilai 1 menunjukkan kesepakatan sempurna dan nilai 0 menunjukkan kesepakatan yang sama dengan kebetulan [17]. Persamaan untuk menghitung kappa sebagai berikut:

$$Kappa = \frac{Po-Pe}{1-Pe} \quad (4)$$

Po = proporsi kesepakatan antara model dan kelas aktual, Pe = proporsi kesepakatan yang diharapkan secara acak

3. HASIL DAN PEMBAHASAN

3.1 Algoritma Random Forest

Nilai akurasi menggunakan algoritma random forest adalah sebesar 80.77%, precision sebesar 81.75%, recall sebesar 80,94% dan nilai kappa 0.706. Pada tabel 3 dapat diketahui prediksi yang tepat yaitu 819 data dan prediksi yang tidak tepat sebanyak 119 data.

```

PerformanceVector:
accuracy: 80.77% +/- 3.84% (micro average: 80.77%)
ConfusionMatrix:
True: high risk  low risk  mid risk
high risk:  241   5      17
low risk:   12   347   88
mid risk:   19   54    231
kappa: 0.706 +/- 0.059 (micro average: 0.706)
    
```

Gambar 3. Hasil Algoritma Random Forest

Tabel 3. Confusion Matrix Algoritma Random Forest

	true high risk	true low risk	true mid risk
pred. high risk	241	5	17
pred. low risk	12	347	88
pred. mid risk	19	54	231

$$\begin{aligned}
 \text{Akurasi} &= \frac{241 + 347 + 231}{241 + 5 + 17 + 12 + 347 + 88 + 19 + 54 + 231} = \frac{819}{1014} = 0.8077 \times 100\% = 80.77\% \\
 P_{(\text{high risk})} &= \frac{241}{241 + 5 + 17} = 0.9163 \times 100\% = 91.63\% \\
 P_{(\text{low risk})} &= \frac{347}{347 + 12 + 88} = 0.7763 \times 100\% = 77.63\% \\
 P_{(\text{mid risk})} &= \frac{231}{231 + 19 + 54} = 0.7599 \times 100\% = 75.99\% \\
 \text{Precision} &= \frac{P_{(\text{high risk})} + P_{(\text{low risk})} + P_{(\text{mid risk})}}{3} = 0.8175 \times 100\% = 81.75\% \\
 R_{(\text{high risk})} &= \frac{241}{241 + 12 + 19} = 0.8860 \times 100\% = 88.60\% \\
 R_{(\text{low risk})} &= \frac{347}{347 + 5 + 54} = 0.8547 \times 100\% = 85.47\% \\
 R_{(\text{mid risk})} &= \frac{231}{231 + 17 + 88} = 0.6875 \times 100\% = 68.75\% \\
 \text{Recall} &= \frac{R_{(\text{high risk})} + R_{(\text{low risk})} + R_{(\text{mid risk})}}{3} = 0.8094 \times 100\% = 80.94\% \\
 \text{Kappa} &= \frac{(819 \times 1014) - 355162}{1014^2 - 355162} = 0.706
 \end{aligned}$$

3.2 Algoritma Random Forest dengan Random Oversampling

Akurasi dari algoritma random forest dengan teknik random oversampling yaitu sebesar 81.85%, precision sebesar 82.32%, recall sebesar 81.86%, dan nilai kappa menjadi 0.728. Teknik ini dapat memperbaiki nilai akurasi klasifikasi dengan persentase kenaikan akurasi sebesar 1.09% serta kenaikan nilai kappa sebesar 2.2%. Pada tabel 4 dapat diketahui prediksi yang tepat yaitu 997 data dan prediksi yang tidak tepat sebanyak 221 data.

PerformanceVector:			
accuracy: 81.85% +/- 4.43% (micro average: 81.86%)			
ConfusionMatrix:			
True:	high risk	low risk	mid risk
high risk:	361	7	20
low risk:	17	340	90
mid risk:	28	59	296
kappa: 0.728 +/- 0.066 (micro average: 0.728)			

Gambar 4. Hasil Algoritma Random Forest dengan Random Oversampling

Tabel 4. Confussion Matrix Algoritma Random Forest dengan Random Oversampling

	true high risk	true low risk	true mid risk
pred. high risk	361	7	20
pred. low risk	17	340	90
pred. mid risk	28	59	296

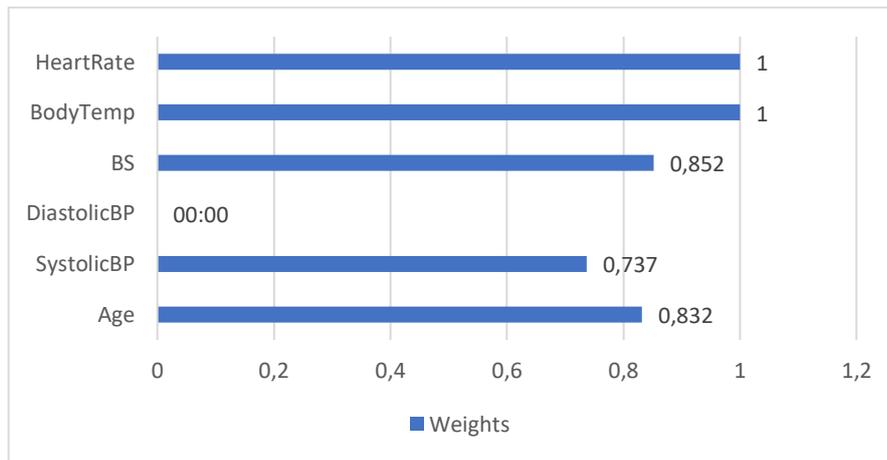
$$\begin{aligned}
 \text{Akurasi} &= \frac{361 + 340 + 296}{361 + 7 + 20 + 340 + 17 + 90 + 296 + 28 + 59} = \frac{997}{1218} = 0.8186 \times 100\% = 81.86\% \\
 P_{(\text{high risk})} &= \frac{361}{361 + 7 + 20} = 0.9304 \times 100\% = 93.04\% \\
 P_{(\text{low risk})} &= \frac{340}{340 + 17 + 90} = 0.7606 \times 100\% = 76.06\% \\
 P_{(\text{mid risk})} &= \frac{296}{296 + 28 + 59} = 0.7728 \times 100\% = 77.28\% \\
 \text{Precision} &= \frac{P_{(\text{high risk})} + P_{(\text{low risk})} + P_{(\text{mid risk})}}{3} = 0.8232 \times 100\% = 82.32\% \\
 R_{(\text{high risk})} &= \frac{361}{361 + 17 + 28} = 0.8892 \times 100\% = 88.92\% \\
 R_{(\text{low risk})} &= \frac{340}{340 + 7 + 59} = 0.8374 \times 100\% = 83.74\% \\
 R_{(\text{mid risk})} &= \frac{296}{296 + 20 + 90} = 0.7291 \times 100\% = 72.91\%
 \end{aligned}$$

$$Recall = \frac{R_{(high\ risk)} + R_{(low\ risk)} + R_{(mid\ risk)}}{3} = 0,8186 \times 100\% = 81,86\%$$

$$Kappa = \frac{(997 \times 1218) - 494508}{1218^2 - 494508} = 0.728$$

3.3 Algoritma Random Forest dengan Random Oversampling dan Particle Swarm Optimization (PSO)

Akurasi dari algoritma Random Forest dengan teknik random oversampling dan PSO adalah 84.41%, precision sebesar 82.93%, recall sebesar 82.92% dan nilai kappa sebesar 0.765. Pembobotan atribut untuk optimasi yang dilakukan dapat dilihat pada gambar 5. Atribut umur menghasilkan bobot sebesar 0.832, Nilai atas Tekanan Darah dalam mmHg 0.737, nilai bawah Tekanan Darah dalam mmHg 0.0, kadar glukosa darah dalam konsentrasi molar mmol/L 0.852, suhu tubuh ibu hamil 1.0, dan denyut jantung istirahat normal dalam denyut per menit 1.0. Pada tabel 6 dapat diketahui prediksi yang tepat yaitu 1010 data dan prediksi yang tidak tepat sebanyak 208 data.



Gambar 5. Bobot Attribute Optimasi

```

PerformanceVector:
accuracy: 82.92% +/- 3.11% (micro average: 82.92%)
ConfusionMatrix:
  True: high risk  low risk  mid risk
  high risk:  375   9       27
  low risk:   15   338    82
  mid risk:   16   59     297
kappa: 0.744 +/- 0.047 (micro average: 0.744)
    
```

Gambar 6. Hasil Algoritma Random Forest dengan Random Oversampling dan PSO

Tabel 6. Confussion Matrix Algoritma Random Forest dengan Random Oversampling dan PSO

	true high risk	true low risk	true mid risk
pred. high risk	375	9	27
pred. low risk	15	338	82
pred. mid risk	16	59	297

$$Akurasi = \frac{375 + 338 + 297}{375 + 15 + 16 + 338 + 9 + 59 + 297 + 16 + 59} = \frac{1010}{1218} = 0.8292 \times 100\% = 82.92\%$$

$$P_{(high\ risk)} = \frac{375}{375 + 9 + 27} = 0.9124 \times 100\% = 91.24\%$$

$$P_{(low\ risk)} = \frac{338}{338 + 15 + 82} = 0.7770 \times 100\% = 77.70\%$$

$$P_{(mid\ risk)} = \frac{297}{297 + 16 + 297} = 0.7984 \times 100\% = 79.84\%$$

$$Precision = \frac{P_{(high\ risk)} + P_{(low\ risk)} + P_{(mid\ risk)}}{3} = 0.8293 \times 100\% = 82.93\%$$

$$R_{(high\ risk)} = \frac{375}{375 + 15 + 16} = 0.9236 \times 100\% = 91.24\%$$

$$R_{(low\ risk)} = \frac{338}{338 + 9 + 59} = 0.8325 \times 100\% = 83.25\%$$

$$R_{(mid\ risk)} = \frac{297}{297 + 16 + 59} = 0.7315 \times 100\% = 73.15\%$$

$$Recall = \frac{R_{(high\ risk)} + R_{(low\ risk)} + R_{(mid\ risk)}}{3} = 0.8292 \times 100\% = 82.92\%$$

$$Kappa = \frac{(1010 \times 1218) - 494508}{1010^2 - 494508} = 0.744$$

Berdasarkan hasil klasifikasi yang dilakukan dengan algoritma random forest diperoleh nilai akurasi sebesar 80.77% dengan nilai kappa 0.706. Setelah dilakukan keseimbangan data dengan random oversampling mengalami peningkatan akurasi sebesar 1.09% dan peningkatan kappa sebesar 2.2%.

Tabel 7. Perbandingan Hasil Akurasi dan Kappa

	RF	RF+random oversampling	Kenaikan
Akurasi	80.77%	81.86%	1.09%
Kappa	0.706	0.728	2.2%

Berikutnya, dilakukan optimasi pada algoritma random forest sehingga menghasilkan nilai akurasi sebesar 82.92% dan kappa 0.744. Hal ini menunjukkan adanya peningkatan akurasi sebesar 1.06% dan peningkatan kappa sebesar 1.6%.

Tabel 8. Perbandingan Hasil Akurasi dengan Optimasi

	RF+random oversampling	RF+random oversampling+PSO	Kenaikan
Akurasi	81.86%	82.92%	1.06%
Kappa	0.728	0.744	1.6%

4. KESIMPULAN

Dalam penelitian ini, peneliti mengusulkan penggunaan teknik Random Oversampling (PSO) untuk mengatasi masalah ketidakseimbangan data pada klasifikasi risiko kesehatan ibu hamil dan Particle Swarm Optimization untuk pembobotan atribut. Hasil pengujian menunjukkan bahwa teknik yang diusulkan mampu meningkatkan kinerja model klasifikasi pada klasifikasi risiko kesehatan ibu hamil dan memiliki akurasi yang lebih tinggi dan tingkat kesalahan klasifikasi yang lebih rendah dibandingkan dengan teknik oversampling dan teknik ensemble lainnya. Hasil pengujian menunjukkan bahwa penggunaan teknik random oversampling dan PSO dapat mengatasi masalah ketidakseimbangan data pada klasifikasi risiko kesehatan ibu hamil dengan lebih baik. Dengan demikian, teknik random oversampling dan PSO dapat digunakan sebagai solusi yang efektif untuk meningkatkan kinerja model klasifikasi pada klasifikasi risiko kesehatan ibu hamil dengan mengatasi masalah ketidakseimbangan data. Teknik ini dapat membantu tenaga medis dalam melakukan identifikasi faktor risiko pada kehamilan yang dapat mempengaruhi kesehatan ibu hamil secara lebih akurat. Selain itu, teknik ini juga dapat membantu meningkatkan kualitas layanan kesehatan dan meningkatkan kualitas hidup ibu dan bayi yang lahir.

REFERENCES

- [1] W. H. Organization, "Maternal Mortality," 2021. <https://www.who.int/news-room/fact-sheets/detail/maternal-mortality>.
- [2] M. Ahmed, M. A. Kashem, M. Rahman, and S. Khatun, "Review and Analysis of Risk Factor of Maternal Health in Remote Area Using the Internet of Things (IoT)," *Lecture Notes in Electrical Engineering*, vol. 632, pp. 357–365, 2020, doi: 10.1007/978-981-15-2317-5_30.
- [3] P. K. Wardani, "Faktor-Faktor yang Mempengaruhi Terjadinya Perdarahan Pasca Persalinan," *Jurnal Aisyah : Jurnal Ilmu Kesehatan*, vol. 2, no. 1, pp. 51–60, 2017, doi: 10.30604/jika.v2i1.32.
- [4] P. R. Sihombing and I. F. Yulianti, "Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia," *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, vol. 20, no. 2, pp. 417–426, 2021, doi: 10.30812/matrik.v20i2.1174.
- [5] Ardiyansyah and P. A. Rahayuningsih, "Penerapan Teknik Sampling Untuk Mengatasi Imbalance," *Jurnal Teknik Informatika Kaputama (JTIK)*, vol. 4, no. 1, pp. 7–15, 2020.
- [6] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-Sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.
- [7] M. Sulistiyono, Y. Pristiyanto, S. Adi, and G. Gumelar, "Implementasi Algoritma Synthetic Minority Over-Sampling Technique Untuk Menangani Ketidakseimbangan Kelas Pada Dataset Klasifikasi," *SISTEMASI: Jurnal Sistem Informasi*, vol. 10, no. 2, pp. 445–459, 2021.
- [8] A. A. Arifiyanti and E. D. Wahyuni, "Smote: Metode Penyeimbang Kelas Pada Klasifikasi Data Mining," *SCAN Jurnal Teknologi Informasi Dan Komunikasi*, vol. 15, no. 1, pp. 34–39.
- [9] G. L. Pritalia, "Analisis Komparatif Algoritme Machine Learning Pada Klasifikasi Kualitas Air Layak Minum," *KONSTELASI: Konvergensi Teknologi Dan Sistem Informasi*, vol. 2, no. 1, pp. 43–55, 2022.
- [10] W. Jaśkowski, P. Liskowski, M. Szubert, and K. Krawiec, "The performance profile: A multi-criteria performance evaluation method for test-based problems," *International Journal of Applied Mathematics and Computer Science*, vol. 26, no. 1, pp. 215–229, Mar. 2016, doi: 10.1515/amcs-2016-0015.

- [11] R. D. Fitriani, H. Yasin, and Tarno, "Penanganan Klasifikasi Kelas Data Tidak Seimbang Dengan Random Oversampling Pada Naive Bayes (Studi Kasus: Status Peserta KB Iud di Kabupaten Kendal)," *Jurnal Gaussian*, vol. 10, no. 1, pp. 11–20, 2021.
- [12] M. P. Paing and S. Choomchuay, "Improved Random Forest (RF) Classifier for Imbalanced Classification of Lung Nodules," in *2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST)*, IEEE, Jul. 2018, pp. 1–4. doi: 10.1109/ICEAST.2018.8434402.
- [13] Q. Dai, J. Liu, and J.-L. Zhao, "Distance-based arranging oversampling technique for imbalanced data," *Neural Comput Appl*, vol. 35, no. 2, pp. 1323–1342, 2023, doi: 10.1007/s00521-022-07828-8.
- [14] Y. E. Achyani, "Penerapan Metode Particle Swarm Optimization Pada Optimasi Prediksi Pemasaran Langsung," *Jurnal Informatika*, vol. 5, no. 1, pp. 1–11, 2018.
- [15] M. Bourel and A. M. Segura, "Multiclass classification methods in ecology," *Ecol Indic*, vol. 85, pp. 1012–1021, Feb. 2018, doi: 10.1016/j.ecolind.2017.11.031.
- [16] K. Iwata and K. Ogasawara, "Assessment of the Efficiency of Non-Invasive Diagnostic Imaging Modalities for Detecting Myocardial Ischemia in Patients Suspected of Having Stable Angina," *Healthcare (Switzerland)*, vol. 11, no. 1, pp. 1–12, 2023, doi: 10.3390/healthcare11010023.
- [17] F. Moons and E. Vandervieren, "Measuring agreement among several raters classifying subjects into one-or-more (hierarchical) nominal categories. A generalisation of Fleiss' kappa," 2023, [Online]. Available: <http://arxiv.org/abs/2303.12502>