JURNAL PENDIDIKAN INDONESIA

Jurnal Pendidikan Indonesia

p-ISSN : 2745-7141 e-ISSN : 2746-1920 Vol. 6 No. 6 2025

Investigasi Performa Artificial Neural Networks Dengan Resampling dan Pembobotan Kelas Pada Data Berat Lahir Rendah Tidak Seimbang

Alfensi Faruk
Universitas Sriwijaya, Indonesia
Email: alfensifaruk@unsri.ac.id

INFO ARTIKEL

Diterima :
Direvisi :
Disetujui :

Kata kunci:
Berat lahir rendah; artificial neural network; resampling; pembobotan kelas

ABSTRAK

Berat lahir rendah (BLR) merupakan salah satu indikator krusial dalam memantau kesehatan bayi baru lahir. Upaya identifikasi dini terhadap risiko BLR penting dilakukan, namun tantangan muncul ketika data yang tersedia tidak seimbang, dengan jumlah kasus BLR jauh lebih sedikit dibandingkan bayi dengan berat lahir normal. Penelitian ini mengevaluasi kinerja model Artificial Neural Networks (ANN) dalam memprediksi BLR menggunakan dua pendekatan penyeimbangan data, yakni teknik resampling dan pembobotan kelas (class weighting). Dataset memuat data ibu yang pernah melahirkan dengan distribusi label tidak seimbang (6.8% BLR). Model ANN dirancang menggunakan fungsi aktivasi ReLU dan sigmoid serta dioptimasi dengan algoritma Adam. Evaluasi performa ANN dilakukan berdasarkan tiga metrik: akurasi, skor F1, dan AUC. Hasil menunjukkan bahwa teknik resampling berhasil menyesuaikan skor akurasi yang terlalu yang diakibatkan ketidakseimbangan data. Pembobotan kelas tanpa resampling juga memberikan hasil yang cukup setara dengan hasil resampling dari beberapa teknik resampling. Studi perbandingan teknik resampling pada kasus data BLR yang digunakan juga memperlihatkan bahwa teknik SMOTEEN memberikan performa model yang terbaik. Akan tetapi, ketika teknik dikombinasikan dengan pembobotan kelas, teknik SMOTE memperlihatkan kualitas prediksi yang lebih baik. Implikasi dari hasil ini dapat digunakan sebagai acuan dalam pengembangan sistem prediktif BLR di layanan kesehatan primer.

Keywords:
Low birth weight; artificial neural network; resampling; class weighting

ABSTRACT

Low birth weight (LBW) is a crucial indicator in monitoring the health of newborns. Early identification of the risk of low weight (LW) is important, but challenges arise when the available data are unbalanced, with the number of LW cases being much smaller than those of babies with normal birth weight. This study evaluates the performance of Artificial Neural Networks (ANNs) in predicting LW using two data balancing approaches: resampling and class weighting techniques. The dataset contains data on mothers who have given birth with an unbalanced label distribution (6.8% LW). The ANN model was designed using the ReLU and sigmoid activation functions and optimized with the Adam algorithm. The ANN performance evaluation was conducted based on three key metrics: accuracy, F1 score, and AUC. The results showed that the resampling technique successfully adjusted the excessive accuracy score caused by data imbalance. Class

weighting without resampling also yielded results that were reasonably equivalent to those obtained from several resampling techniques. A comparative study of resampling techniques in the used LW data also showed that the SMOTEEN technique provided the best model performance. However, when the technique is combined with class weighting, the SMOTE technique shows better prediction quality. The implications of these results can serve as a reference for developing predictive BLR (Birth Weight Low) systems in primary healthcare services.

PENDAHULUAN

Berat Lahir Rendah (BLR) masih menjadi salah satu permasalahan utama dalam bidang kesehatan ibu dan anak, khususnya di negara berkembang (Barker et al., 2019). Bayi yang lahir dengan berat kurang dari 2.500 gram dikategorikan sebagai BLR, sesuai dengan batasan yang ditetapkan oleh World Health Organization (WHO, 2014). Kondisi ini dapat menyebabkan peningkatan risiko terhadap gangguan tumbuh kembang (Sharma & Dutta, 2017), komplikasi kesehatan pada masa neonatal, hingga kematian bayi (Gagnon et al., 2020). Selain itu, BLR juga terkait dengan peningkatan risiko penyakit kronis di kemudian hari, seperti hipertensi dan diabetes (Barker et al., 2019). Oleh karena itu, prediksi dini terhadap kemungkinan terjadinya BLR sangat penting untuk mendukung upaya pencegahan dan perbaikan kualitas layanan kesehatan maternal dan neonatal (Patel & Shaikh, 2021). Beberapa penelitian menunjukkan bahwa faktor-faktor seperti status gizi ibu dan pemantauan kehamilan yang tepat berperan besar dalam mencegah terjadinya BLR (Hoffman et al., 2018). Prediksi BLR dapat dilakukan melalui pemantauan rutin pada faktor risiko yang ada selama kehamilan, yang pada gilirannya dapat mengurangi tingkat kejadian BLR (Moll & Cunningham, 2020).

Pemanfaatan teknologi kecerdasan buatan (*artificial intelligence*) dalam bidang kesehatan telah berkembang pesat, salah satunya melalui penerapan *Artificial Neural Networks* (ANN). ANN merupakan algoritma *Machine Learning* (ML) yang dirancang untuk mengenali pola dari data kompleks dan bersifat non-linier (LeCun *et al*., 2015). Dalam praktiknya, ANN telah digunakan dalam berbagai studi prediksi medis, termasuk klasifikasi penyakit dan deteksi kondisi risiko. Akan tetapi, tantangan yang sering muncul dalam implementasi ANN pada data klinis adalah masalah ketidakseimbangan kelas (*class imbalance*), dimana jumlah sampel dari kelompok normal jauh lebih besar dibandingkan dengan kelompok kasus BLR. Ketimpangan ini berdampak pada penurunan sensitivitas model terhadap kelas minoritas (He & Garcia, 2009).

Beberapa pendekatan telah dikembangkan untuk mengatasi ketidakseimbangan data tersebut, antara lain melalui teknik dan pembobotan kelas *(class weighting*). Teknik *resampling*, seperti *Synthetic Minority Over-Sampling Technique* (SMOTE) (Chawla *et al*., 2002), bertujuan untuk menambah data sintetis pada kelas minoritas, sementara SMOTE *and Edited Nearest Neighbours* (SMOTEENN) (Muntasir Nishat *et al*., 2022) menggabungkan teknik *oversampling* dan *undersampling* dengan menambah dan mengurangi jumlah data pada kelas. Di sisi lain, pendekatan pembobotan kelas memberikan bobot kesalahan yang lebih besar pada kelas minoritas

dalam proses pelatihan model, sehingga model dapat memberikan perhatian lebih terhadap kelas yang jarang muncul.

Penelitian ini bertujuan untuk melakukan investigasi terhadap performa ANN dalam mengklasifikasi data BLR yang tidak seimbang, menggunakan teknik *resampling* dan pembobotan kelas. Evaluasi dilakukan menggunakan metrik yang relevan terhadap data tidak seimbang, seperti akurasi, skor F1, dan *Area Under The Curve* (AUC). Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem prediksi BLR berbasis kecerdasan buatan yang lebih akurat dan aplikatif dalam dunia kesehatan.

Penelitian ini bertujuan untuk melakukan investigasi terhadap performa ANN dalam mengklasifikasi data BLR yang tidak seimbang, menggunakan teknik resampling dan pembobotan kelas. Evaluasi dilakukan menggunakan metrik yang relevan terhadap data tidak seimbang, seperti akurasi, skor F1, dan Area Under The Curve (AUC). Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem prediksi BLR berbasis kecerdasan buatan yang lebih akurat dan aplikatif dalam dunia kesehatan. Berbeda dari penelitian sebelumnya, studi ini secara sistematik membandingkan pendekatan resampling dan pembobotan kelas dalam konteks ANN untuk kasus BLR di Indonesia.

METODE

Data

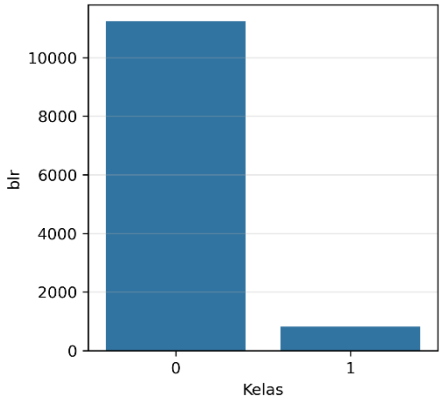
Penelitian ini menggunakan data sekunder hasil Survei Demografi dan Kesehatan Indonesia (SDKI) tahun 2012 (BPS, 2012). Data mencakup atribut-atribut ibu yang melahirkan seperti usia, paritas, riwayat kehamilan, status gizi, tekanan darah, serta hasil akhir berupa berat lahir bayi. Data telah melalui proses *cleaning*, termasuk penanganan nilai hilang dan penghapusan duplikasi. Kelas target terbagi menjadi dua kategori, yaitu normal ($\geq$ 2.500 gram) dan BLR ($<$ 2.500 gram). Distribusi awal data menunjukkan ketidakseimbangan kelas, dengan persentase kasus BLR sekitar 6.8% dari total data.

Pra-Pemrosesan Data

Data awal yang didapatkan berjumlah 45.607 orang wanita yang pernah menikah dan berusia diantara 15-49 tahun. Akan tetapi, setelah dilakukan pra-pemrosesan terhadap data tersebut diperoleh 12.055 data yang memenuhi kriteria, yaitu wanita yang pernah menikah dan melahirkan anak pada rentang waktu bulan Mei 2007 hingga bulan Juli 2012. Data ini juga telah digunakan dalam penelitian yang dilakukan oleh Faruk dkk. (Faruk *et al.*, 2018) dan Eliyati dkk. *(Eliyati et al*., 2019). Untuk karakteristik ibu melahirkan yang digunakan sebagai prediktor dalam penelitian ini mengikuti hasil penelitian Faruk dkk. (Faruk *et al*., 2018), yaitu tempat tinggal, zona waktu, indeks kekayaan, pendidikan ibu, pendidikan ayah, usia ibu, pekerjaan ibu, dan jumlah anak. Terdapat satu prediktor kontinu, yaitu usia, yang kemudian variabel tersebut distandarisasi, sedangkan prediktor lainnya berjenis kategori.

Distribusi dari label, yaitu berat lahir rendah yang dinotasikan dengan ‘blr’ dapat dilihat pada Gambar 1. Label memiliki dua kelas, yaitu 0 (tidak BLR) dan 1 (BLR), dengan kelas 0 dan

kelas 1 berturut-turut sebanyak 11.232 dan 823. Data selanjutnya dibagi secara acak menjadi dua kelompok, yaitu 70% data latih (*train*) dan 30% data uji (*test*). Dalam pembagian data tersebut, proporsi kedua kelas labelnya masih tetap dipertahankan, sehingga diperoleh data latih sebanyak 8.438 (kelas 0 = 7.862 dan kelas 1 = 576) dan data uji sebanyak 3.617 (kelas 0 = 3.370 dan kelas 1 = 247). Secara umum, rasio kelas 0 dan kelas 1 pada data latih dan uji tersebut adalah kurang lebih 14:1, yang mengindikasikan bahwa proporsi kedua kelas label pada data ini tidak seimbang.



Gambar 1. Distribusi kelas label 'blr' sebagai target dari ANN
Sumber : Data Olahan

Seluruh eksperimen dalam penelitian ini dilakukan dengan pendekatan yang menjaga validitas hasil melalui proses **k-fold cross-validation (CV)** pada data latih. Penelitian menggunakan nilai **k=5**, yang berarti data latih dibagi menjadi lima subset dan model dilatih serta divalidasi sebanyak lima kali dengan kombinasi subset yang berbeda. Pendekatan ini bertujuan untuk mengurangi bias akibat pembagian data dan memberikan estimasi performa model yang lebih dapat digeneralisasi. Selain itu, pemilihan arsitektur terbaik dari Artificial Neural Networks (ANN) juga didasarkan pada performa rata-rata dari hasil validasi silang ini.

Setiap eksperimen, termasuk pada penerapan teknik resampling dan pembobotan kelas, dilakukan secara **berulang** untuk meminimalkan fluktuasi hasil akibat proses pelatihan model yang bersifat stokastik. Uji coba berulang ini memastikan bahwa performa model yang diperoleh tidak bersifat kebetulan atau overfitted terhadap satu subset data. Dengan demikian, kombinasi antara validasi silang dan uji coba berulang menjamin bahwa model ANN yang dievaluasi mampu melakukan prediksi secara konsisten pada data uji yang tidak terlihat sebelumnya, serta memiliki ketahanan terhadap variasi distribusi data.

Arsitektur Artificial Neural Network

Sebelum melakukan prediksi dengan ANN, langkah pertama yang dilakukan adalah menentukan nilai-nilai *hyper-parameter* sebagai arsitektur ANN yang digunakan. Jumlah unit pada *input layer* dari arsitektur ANN yang digunakan sama dengan banyaknya prediktor, yaitu delapan. Sementara itu, jumlah unit pada *output* layer hanya satu. Karena output dari ANN yang diinginkan berupa peluang, maka *activation function* pada *output layer* adalah fungsi *sigmoid*. Penelitian ini

juga menggunakan *binary crossentropy* sebagai fungsi *loss*. Untuk menentukan nilai *hyper-parameter* lainnya, penelitian ini menggunakan prosedur k*-fold Cross-Validation* (CV) pada data latih dengan nilai $k = 5$. Penggunaan *k-fold* CV dalam penentuan arsitektur ANN bertujuan untuk mengurangi bias pada saat membagi data awal menjadi data latih dan data uji, serta untuk meningkatkan performa model saat melakukan prediksi pada data uji tersebut.

Hasil yang diperoleh berdasarkan proses k*-fold* CV tersebut disajikan pada *Tabel* 1. Dari tabel tersebut terlihat bahwa jumlah *hidden layer* dan jumlah unit di setiap *hidden layer* relatif kecil. Hal tersebut mengindikasikan bahwa arsitektur ANN yang digunakan cukup sederhana. Nilai $\lambda = 0{,}00001$ juga mengindikasikan bahwa model tidak terlalu memerlukan regularisasi karena nilai L^2*-regularization* yang diberikan sangat minimal. Jumlah *epoch* = 100 dan jumlah *batch* =10 memperlihatkan bahwa arsitektur ANN yang digunakan tidak memerlukan proses pelatihan yang cukup panjang. Seluruh proses komputasi dari arsitektur ANN dalam penelitian ini menggunakan *platform* ML dari Python, yaitu Tensorflow 2.19.0 (Abadi *et al*., 2016).

Tabel 1. Arsitektur ANN berdasarkan *k-fold* CV pada data latih

Sumber : Data Olahan

Hyper-parameter	**Nilai**
Jumlah *hidden layer*	2
Jumlah unit pada setiap *hidden layer*	(16;16)
Activation Function* untuk setiap *hidden layer	ReLu
Learning rate	0,01
***L^2-regularization* (λ)**	0,00001
Jumlah *epoch*	100
Jumlah *batch*	10
Optimizer	Adam

Penanganan Ketidakseimbangan Kelas

Untuk menangani masalah ketidakseimbangan data, penelitian ini melakukan pendekatan berikut:

1. *Resampling*, yang mencakup SMOTE (Chawla *et al.*, 2002), *Adaptive Synthetic Sampling Approach* (ADASYN) (He *et al*., 2008), SMOTE *and Tomek links* (SMOTETomek) (Wang *et al*., 2019), dan SMOTEENN (Muntasir Nishat *et al*., 2022).
2. Pembobotan kelas, yaitu teknik modifikasi fungsi *loss* dengan memberikan bobot yang lebih tinggi pada kelas BLR, sehingga kesalahan pada kelas minoritas memiliki pengaruh lebih besar dalam pembaruan bobot selama pelatihan. Misalkan label memiliki dua kelas, yakni $i = 0{,}1$, formula pembobotan kelas dapat didefinisikan sebagai berikut:

$$weight_i = \frac{n}{k \times n_i},$$

dimana $weight_i$ adalah bobot untuk kelas ke-i , n adalah jumlah total sampel, n_i adalah jumlah sampel kelas ke-i, dan k adalah jumlah kelas.

Metrik Evaluasi

Pada saat mengimplementasikan ANN, salah satu tahapan penting adalah mengevaluasi performanya pada data uji yang independen dari data latih. Metrik yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Akurasi, yang didefinisikan sebagai proporsi dari banyaknya prediksi yang benar terhadap total banyaknya prediksi yang dilakukan. Nilai akurasi ini biasanya dinyatakan dalam persen (0% sampai dengan 100%). Semakin besar nilai akurasi semakin baik performa model.
2. Skor F1 (kelas 1), merupakan skor yang menyeimbangkan nilai *false positive* dan *false negative*. Skor F1 yang digunakan dalam artikel ini adalah skor untuk kelas 1 dari label sebagai target prediksi. Rentang skor F1 terletak pada interval 0 dan 1, dimana semakin besar nilainya mengindikasikan semakin baiknya performa model.
3. AUC, merupakan luas daerah di bawah kurva *receiver operating characteristic* (ROC). Metrik ini memberikan ukuran kemampuan model dalam membedakan antar kelas, atau sering disebut sebagai kapabilitas diskriminasi. Nilai AUC terletak diantara 0,5 sampai dengan 1, semakin besar nilainya semakin baik. Nilai AUC=1 menyatakan bahwa model memiliki kapabilitas diskriminasi yang sempurna, sedangkan nilai AUC=0,5 menyatakan bahwa performa diskriminasi model tidak lebih dari suatu tebakan acak (*random guessing*).

HASIL DAN PEMBAHASAN

Performa ANN Tanpa Resampling dan Pembobotan Kelas

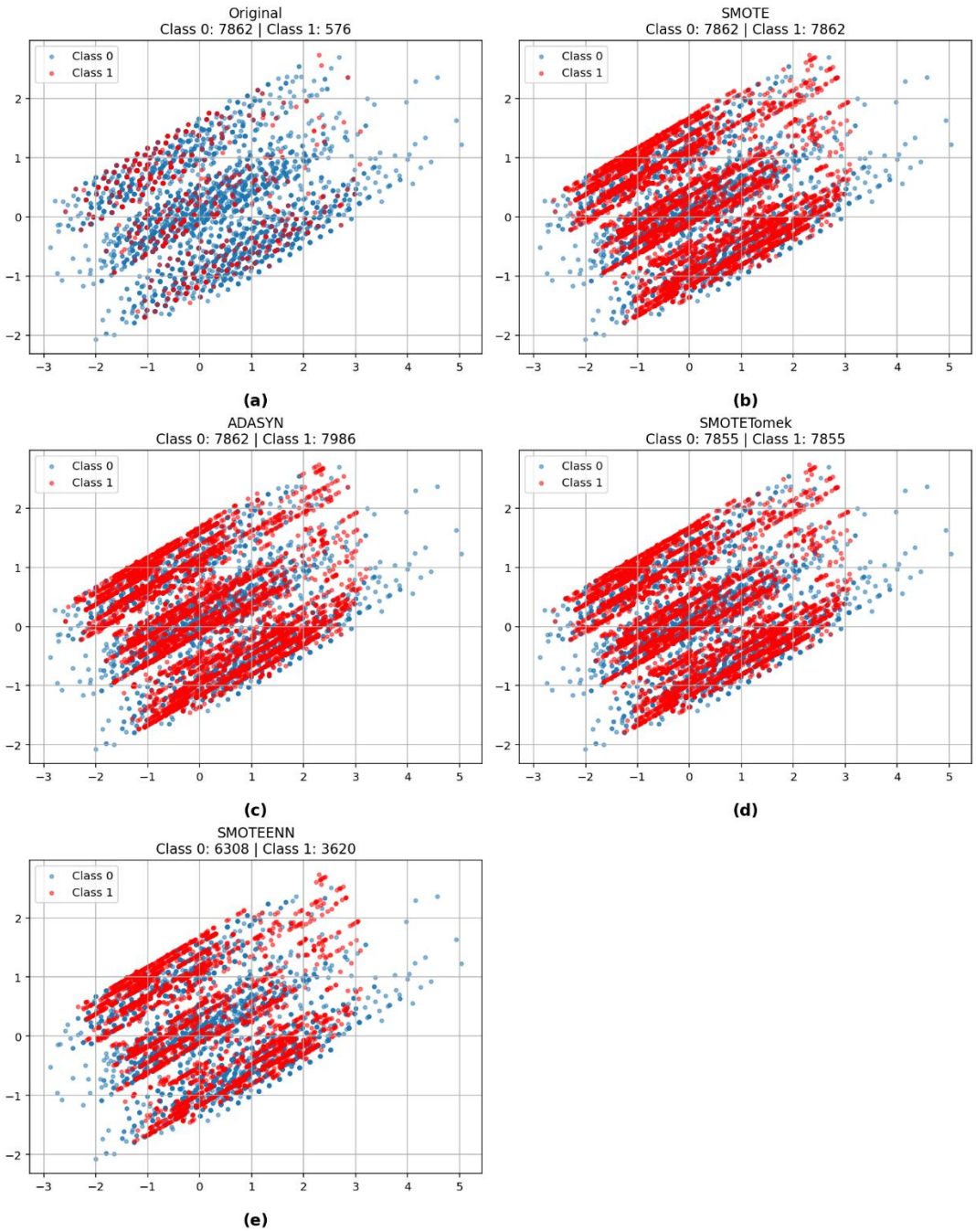
Setelah diperoleh arsitektur ANN, langkah selanjutnya adalah melakukan pelatihan model dari ANN pada data latih berdasarkan arsitektur tersebut. Prediksi kemudian dilakukan pada data uji, yang selanjutnya performa dari prediksi tersebut dievaluasi menggunakan tiga metrik, yaitu akurasi, skor F1 (untuk kelas 1 sebagai target prediksi), dan AUC. Berdasarkan komputasi yang telah dilakukan, performa ANN diberikan oleh nilai-nilai metrik pada data uji, yaitu akurasi (93,14%), skor F1 (0), dan AUC (0,54). Terlihat bahwa nilai akurasi yang cukup tinggi. Namun, skor F1 yang bernilai nol mengindikasikan bahwa ANN sama sekali tidak dapat memprediksi kelas 1 dari label 'blr'. Sementara itu, nilai AUC mendekati nilai ambang batas 0.5, sehingga dapat dikatakan bahwa performa model dalam mengklasifikasi secara benar data-data pada data uji hampir sama dengan tebakan acak. Oleh karena itu, terdapat indikasi bahwa nilai akurasi yang cukup tinggi, yaitu di atas 90%, adalah bias akibat kontribusi dari performa model dari kelas 0 yang sangat mendominasi.

Hasil yang hampir sama juga dilaporkan dalam penelitian yang dilakukan Faruk *et al.* (Faruk et al., 2018) dan Eliyati *et al*. (Eliyati et al., 2019), walaupun kedua penelitian tersebut tidak melaporkan skor F1. Berdasarkan hasil dari penelitian tersebut, menggunakan *Random Forests*, *Support Vector Machines* (SVM), dan model regresi logistik, skor akurasi sekitar 93% dan nilai AUC mendekati 0,5. Apabila kurang hati-hati, hal tersebut dapat membuat seseorang melakukan kesalahan dalam menilai performa dari model yang digunakan. Jika dilihat distribusi kelas dari 'blr' pada Gambar 1, proporsi terjadinya BLR (kelas 1) jauh lebih sedikit dari proporsi tidak terjadinya BLR (kelas 0).

Dalam bagian selanjutnya, dibahas implementasi dari beberapa teknik yang umum digunakan dalam literatur untuk mengatasi permasalahan ketidakseimbangan label pada data. Teknik *resampling* ini bertujuan untuk meningkatkan pelatihan model dan evaluasi performa model dengan cara menambah dan/atau mengurangi sampel baru berdasarkan sampel yang ada.

Pengaruh Teknik Resampling Terhadap Performa ANN

Dalam penelitian ini, terdapat empat teknik yang digunakan, yaitu SMOTE, ADASYN, SMOTETomek, dan SMOTEENN. Sebelum dilakukan pelatihan model pada data latih yang telah dilakukan *resampling*, terlebih dahulu diberikan visualisasi *principle component analysis* (PCA) dari data latih tanpa *resampling* (*original*) dan data latih dengan *resampling* menggunakan keempat teknik *resampling* tersebut (Gambar 2). Seluruh proses komputasi dan visualisasi dari *resampling* tersebut dilakukan menggunakan modul-modul dari Python, yaitu *imbalanced-learn* versi 0.13.0 dan Matplotlib versi 3.10.0.



Gambar 2. Visualisasi PCA data latih (a) *original*, dan data latih dengan *resampling* menggunakan metode (b) SMOTE, (c) ADASYN, (d) SMOTETomek, dan (e) SMOTEENN. Sumber : Data Olahan

Berdasarkan Gambar 2a, distribusi data *original* memperlihatkan kelas 0 (warna biru) sangat mendominasi kelas 1 (warna merah) sehingga banyak data dengan kelas 1 yang tumpang tindih dengan kelas 0. Selain itu, tidak ada batas yang jelas bagi model untuk mempelajari pola kelas 1. Hasil *resampling* menggunakan SMOTE, ADASYN, dan SMOTETomek (Gambar 2b-d)

memiliki pola yang hampir sama, yaitu kelas 0 dan kelas 1 masih saling tumpang tindih dan sangat padat. Namun, hasil *resampling* yang diperlihatkan oleh SMOTEENN (Gambar 2e) lebih baik dari hasil *resampling* metode lainnya. Banyak *noise* dan tumpang tindih antar kelas yang telah dihilangkan, sehingga batas antar kedua kelas lebih jelas. Selain itu, distribusi dari data kelas 1 juga tidak terlalu padat.

Pelatihan model ANN dilakukan terhadap data latih dengan *resampling* berdasarkan keempat metode tersebut. Selanjutnya, evaluasi performa pada data uji menggunakan metrik akurasi, F1, dan AUC, bersama-sama dengan performa pelatihan model pada data latih *original* disajikan di dalam Tabel 2.

Tabel 2. Evaluasi performa ANN pada data uji hasil dari pelatihan model pada data latih *original* dan empat teknik berdasarkan metrik akurasi, F1, dan AUC.

Sumber : Data Olahan

Teknik	Akurasi (%)	F1	AUC
Original	93,14	0	0,54
SMOTE	68,32	0,12	0,49
ADASYN	62,01	0,12	0,50
SMOTETomek	77,11	0,11	0,50
SMOTEENN	75,48	0,13	0,51

Berdasarkan Tabel 2, data *original* memiliki tingkat akurasi tertinggi (93,14%), namun hasil ini bias dikarenakan proporsi dari kelas target 1 yang jauh lebih kecil dari kelas 0. Untuk prediksi ANN pada data uji berdasarkan data latih dengan *resampling*, terjadi penurunan penurunan akurasi karena model mencoba untuk menyeimbangkan kedua kelas label. Penurunan terjadi pada nilai AUC dari ANN yang dilatih dengan data hasil *resampling*. Penurunan AUC ini tidak bermakna karena sangat kecil dan tetap nilainya di sekitar ambang batas tebakan acak. Berdasarkan skor F1, terlihat terjadi peningkatan performa model yang signifikan setelah dilakukannya *resampling* pada data latih, dengan peningkatan paling tinggi diberikan oleh metode SMOTEENN.

Hasil yang diperoleh di dalam Tabel 2 juga sejalan dengan penelitian Qadrini dkk. (Qadrini et al., 2022) yang menggunakan SVM pada klasifikasi dara tidak seimbang, yang memperlihatkan terjadinya penyesuaian nilai akurasi dari di atas 95% sebelum *resampling* menurun menjadi sekitar 76% setelah *resampling* denganm teknik SMOTE. Selain itu, performa diskriminasi tidak berubah signifikan setelah *resampling* yang diperlihatkan oleh nilai AUC yang tetap berada disekitar 0,5. Menggunakan SVM dan *Decision Tree* (DT) pada prediksi kejadian serangan jantung, terjadi juga penyesuaian nilai akurasi yang di atas 90% menjadi sekitar 75% setelah dilakukan teknik *resampling* (Febriyanti & Baita, 2025).

Performa ANN dengan Kombinasi Pembobotan Kelas dan Teknik Resampling

Selain teknik *resampling*, penelitian ini juga mengaplikasikan metode pembobotan kelas (*class weighting*) untuk mengatasi data dengan label yang tak seimbang pada ANN. Kombinasi teknik *resampling* dengan pembobotan kelas diterapkan baik pada data latih *original* dan data latih

hasil *resampling*. Hasil evaluasi performa ANN pada data uji hasil dari pelatihan model pada data latih dengan pembobotan kelas ditampilkan dalam Tabel 3.

Berdasarkan hasil pada Tabel 3, terlihat bahwa terjadi penurunan akurasi dari hampir semua teknik dan data *original*, kecuali SMOTE yang akurasinya naik menjadi 71.19%. Hal dikarenakan ANN telah menyesuaikan dengan kontribusi dari kelas 1 yang semakin besar. Nilai skor F1 data *original* dan SMOTETomek naik menjadi 0,13 sedangkan terjadi penurunan pada SMOTEENN. Akan tetapi, sulit membedakan performa ANN antar teknik *resampling* berdasarkan skor F1, karena perbedaan di antara skor F1 tersebut sangat minimal. Sementara itu, nilai AUC tidak terpengaruh oleh pembobotan kelas yang diperlihatkan nilainya sama persis dengan Tabel 2. Secara keseluruhan, SMOTE sedikit lebih baik dari teknik *resampling* lainnya, diikuti oleh SMOTEENN dan data *original*.

Tabel 3. Evaluasi performa ANN pada data uji hasil dari pelatihan model pada data latih *original* dan data latih hasil *resampling* yang telah dilakukan pembobotan kelas berdasarkan metrik akurasi, F1, dan AUC.

Sumber : Data Olahan

Teknik Resampling	Akurasi (%)	F1	AUC
Original	63,92	0,13	0,54
SMOTE	71,19	0,12	0,50
ADASYN	58,47	0,12	0,51
SMOTETomek	61,99	0,13	0,52
SMOTEENN	65,03	0,11	0,50

KESIMPULAN

Penelitian ini membahas bagaimana proses prediksi BLR menggunakan ANN, dimana proporsi dari kedua kelas pada label tidak seimbang. Berdasarkan hasil dan pembahasan, prediksi dengan ANN pada data latih tanpa *resampling* tidak dianjurkan. Nilai akurasi yang tinggi adalah bias akibat dari hasil prediksi yang didominasi oleh kelas label yang bukan target. Ketika hanya menggunakan teknik *resampling* saja, terjadi penyesuaian skor akurasi yang menjadi tidak terlalu tinggi. Walaupun begitu, pada saat menggunakan teknik pembobotan kelas saja tanpa *resampling*, performa ANN tidak berbeda jauh dengan hasil *resampling*. Hasil implementasi juga memperlihatkan bahwa teknik SMOTEEN memberikan hasil yang terbaik. Akan tetapi, ketika teknik *resampling* dikombinasikan dengan pembobotan kelas, metode SMOTE memperlihatkan kualitas prediksi yang lebih baik. Walaupun prediksi pada data BLR tak seimbang ini memiliki akurasi yang relatif baik, namun performa model pada kelas minoritasnya masih tidak baik yang diperlihatkan oleh nilai F1 untuk kelas 1 yang rendah. Begitu juga dengan kemampuan diskriminasi ANN, yang diindikasikan oleh nilai AUC, pada data BLR tak seimbang ini dapat dikatakan setara dengan tebakan acak. Untuk penelitian selanjutnya, disarankan studi dapat difokuskan pada salah satu teknik saja yang paling cocok dengan karateristik data yang digunakan dan dikombinasikan dengan teknik pembobotan kelas. Kemudian, dilakukan pelatihan

menggunakan metode ML lainnya yang dirancang khusus untuk mengatasi data tak seimbang, seperti *balanced random forest* dan *XGBoost*.

DAFTAR PUSTAKA

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., & Devin, M. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv Preprint arXiv:1603.04467*.

BPS. (2012). *Survey Demografi dan Kesehatan Indonesia Tahun 2012* [Dataset].

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.

Eliyati, N., Faruk, A., Kresnawati, E. S., & Arifieni, I. (2019). *Support vector machines for classification of low birth weight in Indonesia*. *1282*(1), 012010.

Faruk, A., Cahyono, E. S., Eliyati, N., & Arifieni, I. (2018). Prediction and classification of low birth weight data using machine learning techniques. *Indonesian Journal of Science and Technology*, *3*(1), 18–28.

Febriyanti, G. A. P., & Baita, A. (2025). Comparison of Support Vector Machine and Decision Tree Algorithm Performance with Undersampling Approach in Predicting Heart Disease Based on Lifestyle. *Journal of Applied Informatics and Computing*, *9*(2), 318–327.

He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). *ADASYN: Adaptive synthetic sampling approach for imbalanced learning*. 1322–1328.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Barker, D. J. P., Thornburg, K. L., & Osmond, C. (2019). The origins of health inequalities in low birth weight and preterm birth. Lancet, 376(9748), 1581-1589. https://doi.org/10.1016/S0140-6736(10)61191-4

Gagnon, A. J., Nault, L., & Tran, N. (2020). The impact of low birth weight on neonatal mortality: A cohort study in Indonesia. International Journal of Pediatrics, 2020, 1-7. https://doi.org/10.1155/2020/1857345

Hoffman, H. J., Sorensen, C., & Simon, M. (2018). Maternal factors influencing the incidence of low birth weight in developing countries. BMC Public Health, 18(1), 450. https://doi.org/10.1186/s12889-018-5342-6

Moll, H., & Cunningham, S. A. (2020). Predicting low birth weight: An evaluation of prenatal care practices. Journal of Maternal-Fetal & Neonatal Medicine, 33(7), 1147-1154. https://doi.org/10.1080/14767058.2019.1638314

Patel, R. P., & Shaikh, I. (2021). The role of prenatal care in preventing low birth weight: Evidence from randomized clinical trials. Journal of Obstetrics and Gynecology, 63(5), 378-384. https://doi.org/10.1016/j.jogc.2020.12.028

Sharma, S., & Dutta, S. (2017). Low birth weight: Current perspectives and management options. Journal of Neonatal-Perinatal Medicine, 10(4), 379-388. https://doi.org/10.3233/NPM-160167Muntasir Nishat, M., Faisal, F., Jahan Ratul, I., Al-Monsur, A., Ar-Rafi, A. M., Nasrullah, S. M., Reza, M. T., & Khan, M. R. H. (2022). A Comprehensive Investigation of the Performances of Different Machine Learning Classifiers With SMOTE-ENN Oversampling Technique and Hyperparameter Optimization for Imbalanced Heart Failure Dataset. *Scientific Programming*, *2022*(1), 3649406.

Qadrini, L., Hikmah, H., & Megasari, M. (2022). Oversampling, Undersampling, Smote SVM dan Random Forest pada Klasifikasi Penerima Bidikmisi Sejawa Timur Tahun 2017. *Journal of Computer System and Informatics (JoSYC)*, *3*(4), 386–391.

Wang, Z., Wu, C., Zheng, K., Niu, X., & Wang, X. (2019). SMOTETomek-based resampling for personality recognition. *IEEE Access*, *7*, 129678–129689.

World Health Organization. (2014). *Global nutrition targets 2025: Low birth weight policy brief*. World Health Organization.