

Klasifikasi Berat Badan Lahir Rendah (BBLR) menggunakan Metode *Support Vector Machine* dengan Teknik SMOTE

Anindya Celena Khansa Kirana¹, Muhammad Tanzil Furqon², Achmad Ridok³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹anindyavip@gmail.com, ²m.tanzil.furqon@ub.ac.id, ³acribdokb@ub.ac.id

Abstrak

Salah satu penyebab utama kematian bayi berhubungan dengan peningkatan Angka Kematian Neonatal (AKN) yang mengalami Berat Badan Lahir Rendah (BBLR). BBLR perlu diidentifikasi maupun diprediksi untuk mencegah kematian ketika mengetahui beresiko BBLR. Pada penelitian ini dibangun sebuah sistem klasifikasi sebagai indentifikasi awal BBLR. Data yang digunakan berasal dari data rekam medis persalinan di Puskesmas Ardimulyo Kabupaten Malang periode januari-agustus 2021 yang berupa *imbalanced class*. Pada penelitian ini metode yang digunakan adalah *Support Vector Machine* (SVM) dengan mengombinasikan teknik *Synthetic Minority Oversampling Technique* (SMOTE). Performa metode SVM tanpa teknik SMOTE dan metode SVM dengan teknik SMOTE yang menggunakan kernel linier dan RBF dibandingkan dalam penelitian ini. Dilakukan pengujian dengan menggunakan 3-fold cross validation pada pengujian kernel dan parameter untuk menemukan metode yang terbaik dan pengujian pada data independen terhadap seluruh metode untuk membandingkan keduanya. Berdasarkan pengujian hasil evaluasi yang didapatkan kurang optimal karena mendapat hasil yang rendah. Dengan pengujian pengujian 3-fold cross validation memperoleh hasil yang terbaik pada kernel RBF dengan parameter *lamda* bernilai 0,1, *gamma* bernilai 0,001, *complexity* 20, iterasi maksimum bernilai 100, dan *epsilon* bernilai 0,001. Sedangkan, hasil pengujian data uji menunjukkan metode yang terbaik adalah metode SVM kernel RBF tanpa menggunakan SMOTE menghasilkan yaitu *accuracy* bernilai 0,75, *precision* bernilai 0,5, *recall* bernilai 0,2, dan *f-measure* bernilai 0,2857.

Kata kunci: klasifikasi, berat lahir badan rendah, *imbalanced class*, *support Vector Machine*, *synthetic minority oversampling technique*

Abstract

One of the main causes of infant mortality is associated with an increase in the Neonatal Mortality Rate (AKN) with low birth weight (LBW). LBW needs to be identified and predicted to prevent death when knowing the risk of LBW. In this study, a classification system was built as the initial identification of LBW. The data used comes from medical record data for childbirth at the Ardimulyo Public Health Center, Malang Regency for the January-August 2021 period in the form of *imbalanced class*. In this study, the method used is the *Support Vector Machine* (SVM) by combining the *Synthetic Minority Oversampling Technique* (SMOTE) technique. The performance of the SVM method without the SMOTE technique and the SVM method with the SMOTE technique using a linear kernel and RBF are compared in this study. Tests were carried out using 3-fold cross validation on kernel and parameter testing to find the best method and independent data testing of all methods. to compare the two. Based on testing the evaluation results obtained are less than optimal because they get low results. By testing the 3-fold cross validation test, the best results are obtained on the RBF kernel with the parameter *lamda* of 0.1, *gamma* of 0.001, *complexity* of 20, maximum iteration of 100, and *epsilon* of 0.001. Meanwhile, the results of data testing show that the best method is the RBF kernel SVM method without using SMOTE, which results in *accuracy* of 0.75, *precision* of 0.5, *recall* of 0.2, and *f-measure* of 0.2857.

Keywords: *classification, imbalanced class, low birth weight, synthetic minority oversampling technique, support vector machine*

1. PENDAHULUAN

Indonesia memiliki komitmen dan tujuan Sustainable Development Goals (SDGs) yaitu salah satu targetnya pada tahun 2030 bisa menurunkan Angka Kematian Neonatal (AKN) hingga 12 per 1000 kelahiran hidup, sedangkan dari data Survei Demografi dan Kesehatan (SKDI) 2017 masih tercatat bahwa AKN 15 per 1000 kelahiran hidup. Angka Kematian Neonatal (AKN) merupakan jumlah anak yang dilahirkan pada tahun tertentu dan meninggal saat periode 28 hari pertama kehidupan dan dinyatakan dalam angka per 1.000 kelahiran hidup (Yuliati & Sihombing, 2021).

Peningkatan AKN yang mengalami Berat Bayi Lahir Rendah (BBLR) merupakan salah satu penyebab utama kematian bayi. BBLR harus diidentifikasi dan diprediksi sebelum kelahiran bayi dengan melakukan observasi dari data historis expectant (Samosir, Wilandari, & Yasin, 2015). Pencegahan maupun penanganan untuk meminimalisir kematian saat melahirkan sangat diperlukan ketika ibu hamil diketahui akan melahirkan dengan kondisi beresiko BBLR (Agustin, Setiawan, & Fauzi, 2019). Oleh karena itu, dibutuhkan suatu sistem untuk melakukan klasifikasi sebagai indentifikasi awal BBLR sehingga tenaga kesehatan mengetahui risiko BBLR dengan mudah.

BBLR merupakan bayi yang memiliki berat lahir kurang dari 2500 gram. Berat lahir rendah merupakan berat bayi setelah satu jam lahir kemudian ditimbang. Dari Proverawati, 2010 dalam buku "BBLR (Berat Badan Lahir Rendah)" yang dikutip oleh Febrianti (Febrianti, 2019), terdapat beberapa faktor yang dapat memengaruhi terjadinya permasalahan BBLR yaitu faktor ibu (usia ibu, penyakit, paritas, jarak persalinan, riwayat BBLR, sosial ekonomi, dan lain-lain), faktor janin (infeksi dan cacat bawaan), faktor plasenta serta faktor lingkungan. Beberapa faktor tersebut dapat digunakan untuk mengidentifikasi penyelesaian permasalahan klasifikasi BBLR melalui penelitian.

Klasifikasi merupakan data mining (machine learning) untuk memprediksi keanggotaan dari kelompok data instances (Phyu, 2009). Seiring dengan kepopuleran era data, klasifikasi data juga memiliki beberapa permasalahan, salah satunya yaitu ketidakseimbangan data yang antara kelas yang berbeda, dan saat kondisi ekstrim pada ketidakseimbangan data disebut dengan

imbalanced data atau rare event (Yuliati & Sihombing, 2021). Permasalahan tersebut dapat menimbulkan *missclassification* apabila tidak diatasi dengan tepat. Pada kasus BBLR ini, terjadi ketidakseimbangan pada rasio kelas BBLR dan tidak BBLR membuat sulit dilakukan prediksi. Dalam mengatasi imbalanced data, penelitian oleh Zhu, Lin, dan Zhou (2013) mengusulkan metode Synthetic Minority Over Sampling Technique (SMOTE) untuk menanganinya.

Penelitian yang menerapkan teknik resampling yaitu oversampling untuk menangani imbalanced class dilakukan oleh Sulistyono, Pristyanto, dkk dengan judul Implementasi Algoritma Synthetic Minority Over-Sampling Technique untuk Menangani Ketidakseimbangan Kelas pada Dataset Klasifikasi 2021 menunjukkan bahwa penangan kelas yang tidak seimbang pada tahap pra-pemrosesan data memberikan pengaruh pada nilai akurasi dan g-mean algoritma Naïve Bayes, SVM, KNN, dan Decision Tree. Hasil nilai G-mean paling baik 96,42% untuk algoritma Naïve Bayes, SVM bernilai 99,37%, KNN bernilai 99,53% dan Decision Tree berniali 96,42% pada dataset ecoli 15,8 setelah SMOTE dilakukan sengan 10-fold cross validation (Sulistyono, Pristyanto, Adi, & Gumelar, 2021).

Selain itu, terdapat penelitian yang menggunakan teknik resampling data lainnya yaitu dengan menggunakan teknik Adaptive Synthetic (ADASYN) seperti penelitian yang dilakukan oleh Wahyu, Mursyid, dan Arief pada 2021 dengan judul Pengaruh Algoritma ADASYN dan SMOTE terhadap Performa Support Vector Machine pada ketidak seimbangan dataset Airbnb dari penelitian tersebut menunjukkan bahwa algoritma SMOTE SVM lebih unggul dibandingkan ADASYN SVM dan SVM tanpa oversampling (Hidayat, Ardiansyah, & Setyanto, 2021). Penelitian tersebut menunjukkan bahwa teknik SMOTE lebih baik jika dibandingkan dengan teknik ADASYN.

Sedangkan penelitian untuk menangani klasifikasi BBLR dilakukan oleh beberapa peneliti sebelumnya yang dibahas oleh Organig, Musdalifah, dan Lusiyanti berjudul Klasifikasi Status Gizi Ibu Hamil Untuk Mengidentifikasi Bayi Berat Lahir Rendah (BBLR) Menggunakan Metode Support Vector Machine (SVM) (Studi Kasus di Puskesmas Labuan). Hasil perfarmansi didapatkan sebesar 92% (Organig, 2017).

Dari beberapa penelitian tersebut

menunjukkan bahwa teknik SMOTE dapat mengatasi permasalahan imbalance data pada klasifikasi terbukti dengan meningkatkan hasil akurasi. Selain itu, metode Support Vector Machine dapat menyelesaikan permasalahan klasifikasi. Oleh karena itu, peneliti memutuskan untuk mengimplementasi kedua metode SMOTE untuk menangani imbalanced class yang ada pada dataset BBLR dan melakukan klasifikasi menggunakan metode Support Vector Machine.

2. DASAR TEORI

2.1. BBLR

Low birth weight (LBW) atau Berat Badan Lahir Rendah (BBLR) yang telah didefinisikan oleh WHO sebagai berat badan saat lahir kurang dari 2500 gram. Terdapat 6 faktor yang berhubungan dengan kejadian BBLR, yaitu faktor usia, hipertensi, paritas, pendarahan antepartum, eclampsia/preeclampsia dan KPD (Setiati & Rahayu, 2017). Berikut adalah atribut yang akan digunakan pada penelitian ini berdasarkan faktor-faktor terjadinya BBLR dari beberapa penelitian lainnya, yaitu:

1. Usia

Usia dapat mempengaruhi kejadian BBLR. Menurut sitirarani (2008) umur yang baik untuk hamil adalah 20-35 tahun bagi seorang ibu. Kehamilan dibawah 20 tahun atau lebih dari 35 tahun akan beresiko tinggi.

2. Tinggi badan

Tinggi badan ibu hamil yang kurang dari sama dengan 145 cm memiliki risiko untuk BBLR dibandingkan ibu hamil yang memiliki tinggi badan lebih dari 145 cm (Trihardiani, 2011).

3. Berat badan

Berat badan pada ibu hamil akan bertambah berkisar antara 8-18 Kg. Berat badan dengan proporsi yang baik akan menunjukkan status gizi ibu yang baik sehingga dapat mempengaruhi BBL dan sebaliknya berat badan dengan proporsi tidak baik akan meningkatkan faktor risiko BBLR.

4. Lingkar lengan atas (LILA)

Lingkar lengan atas atau LILA merupakan unsur antropometri untuk menentukan status gizi ibu hamil. Wanita yang mempunyai pola makan baik dengan berat badan normal sebelum kehamilan

maka masalah kehamilan tidak akan terjadi, sedangkan ibu hamil yang mengalami kekurangan gizi dapat menyebabkan kelahiran BBLR pada bayi (Jaya, 2009).

5. Hipertensi

Tekanan darah yang tinggi atau hipertensi menyebabkan penurunan aliran darah ke plasenta yang dapat mempengaruhi persediaan atau distribusi oksigen serta nutrisi pada janin. Hal tersebut dapat memperlambat pertumbuhan dan perkembangan janin sehingga meningkatnya risiko saat melahirkan salah satunya BBLR (Setiati & Rahayu, 2017).

6. Paritas

Paritas yang merupakan banyaknya kelahiran hidup yang dimiliki oleh seorang wanita. Paritas berhubungan dengan BBLR disebabkan banyak pasangan yang tidak mengikuti program KB sehingga jika paritas banyak akan sangat beresiko premature, BBLR, dan kematian janin (Setiati & Rahayu, 2017).

7. Jarak Paritas

Ibu hamil yang mempunyai jarak paritas atau kelahiran lebih dari 2 tahun akan memiliki risiko BBBLR lebih tinggi dibandingkan dengan ibu hamil berjarak paritas lebih dari 2 tahun (Febrianti, 2019).

2.2. Klasifikasi

Klasifikasi adalah bagian dari salah satu proses yang penting dalam data mining. Pengelompokan fitur pada kelas yang sesuai disebut dengan klasifikasi. Sekumpulan data dengan kelas sudah ditentukan dilatih sehingga sebuah pengklasifikasi dapat dibuat (Wibawa, Purnama, Akbar, & Dwiyanto, 2018). Tujuan klasifikasi yaitu memprediksi secara akurat terhadap kelas target pada setiap kasus yang ada dalam data. Tahap klasifikasi adalah membuat sebuah pola yang belum diketahui kelas dari pola tersebut menggunakan proses pembelajaran pada training (Tambun, Furqon, & Widodo, 2018).

2.2. Normalisasi Asimtotik

Normalisasi adalah proses pengolahan data untuk menormalisasikan standar data sehingga berada pada rentang tertentu (Dewi, 2013).

$$X' = \left(\frac{X - \min(X)}{\max(X) - \min(X)} \right) \times 0,8 + 0,1 \quad (1)$$

2.3. Synthetic Minority Oversampling

Technique (SMOTE)

Synthetic Minority Oversampling Technique (SMOTE) merupakan metode yang digunakan untuk menyeimbangkan data yang tidak seimbang dengan oversampling yaitu dengan membuat data sintetis (Chawla, 2002). Pada penelitian ini, data berskala numerik. Berikut merupakan tahapan dari algoritma SMOTE berskala numerik.

1. Tentukan sampel data minoritas yang akan digunakan (T), nilai persentase SMOTE (N) dan tentukan jumlah k-tetangga terdekat.

2. Hitung N dan hitung T untuk jumlah iterasi KNN menggunakan persamaan

$$N = \begin{cases} T = \frac{N}{100} \times T, N = 100 & \text{Jika } N < 100 \\ N, N \geq 100 \end{cases} \quad (2)$$

3. Hitung jumlah data sintesis yang akan dibuat dengan persamaan:

$$N = \text{int}\left(\frac{N}{100}\right) \quad (3)$$

4. Hitung k-tetangga terdekat menggunakan jarak Euclidean dengan persamaan:

$$\text{dist} = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \dots + (X_n - Y_n)^2} \quad (4)$$

5. Pilih k-tetangga secara acak yang terdekat.

6. Hitung perbedaan (dif) antar data yang jaraknya terdekat (x_knn) dengan data yang direplikasi (x_i) dengan persamaan:

$$\text{dif} = (x_{knn} - x_i) \quad (5)$$

7. Tentukan bilangan acak (δ) antara 0 hingga 1

8. Hitung persamaan berikut untuk membuat data sintetis

$$x_{syn} = x_i + \text{dif} \times \delta \quad (6)$$

Keterangan:

x_{syn} : Data sintetis

x_i : Data yang direplikasi

2.4. Support Vector Machine

Support Vector Machine (SVM) merupakan metode regresi maupun pengklasifikasi data berdasarkan dari data sebelumnya serta pemodelannya dilakukan supervise dahulu (Qasthari, 2017). Konsep dari SVM yaitu usaha untuk mencari lokasi hyperplane paling optimal yang berguna sebagai pemisah dua kelas dalam input space (Hasibuan, Mukid, & Prahutama, 2017).

SVM dapat digunakan untuk mengatasi

klasifikasi non linier yaitu ketika hyperplane tidak dengan mudah dipisahkan secara sempurna (Yustihan, Adikara, & Indriati, 2021). Permasalahan tersebut dapat diatasi dengan menggunakan fungsi kernel untuk ditransformasikan ke dalam dimensi yang lebih tinggi sehingga dapat diklasifikasikan.

Fungsi kernel yang digunakan pada SVM secara umum yaitu:

1. Kernel linier

$$K(x_i, x_j) = x_i^T x_j \quad (7)$$

2. Kernel Polynomial

$$K(x_i, x_j) = (x_i^T x_j + 1)^d \quad (8)$$

3. Kernel Radial Basis Function (RBF)

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (9)$$

2.4.1 Sequential Training

Quadratic Programming merupakan salah satu dari beberapa jenis proses training yang dimiliki oleh SVM. Kelemahan dari Quadratic Programming yaitu membutuhkan waktu yang lama dan algoritmanya kompleks. Dari kelemahan tersebut, sequential training diusulkan oleh Vijaykumar dan Wu tahun 1999 untuk menemukan hyperplane yang paling baik untuk memisahkan kelas data menjadi dua seperti tujuan dari SVM (Yustihan, Adikara, & Indriati, 2021). Metode Sequential training lebih optimal karena lebih sederhana dan tidak membutuhkan banyak waktu dalam prosesnya (Agustina, Furqon, & Rahayudi, 2018). Algoritma sequential training yaitu sebagai berikut:

1. Inisialisasi parameter yang digunakan, yaitu:

- Complexity(C): nilai kompleksitas
- Lamda (λ) : nilai scalar
- Gamma (γ): nilai learning rate atau laju pembelajaran
- Epsilon (ϵ) : nilai error
- Nilai Iterasi maksimum

2. Menghitung matrik Hessian dengan persamaan:

$$D_{i,j} = y_i y_j (K(x_i, x_j) + \lambda^2) \quad (10)$$

Keterangan:

- $D_{i,j}$: matrik hessian
- y_i dan y_j : kelas dari data ke-i dan data ke-j
- λ : nilai skalar
- $K(x_i, x_j)$: fungsi kernel

3. Inisialisasi nilai parameter Lagrange Multiplier dengan $\alpha_i = 0$ untuk

menemukan support vector.

- a) Menghitung nilai error (E_i)

dengan persamaan:

$$E_i = \sum_{j=1}^n a_j D_{ij} \quad (11)$$

a_j : nilai Lagrange Multipliers ke-j

- b) Menghitung perubahan nilai Lagrange Multiplier (δa_i) dengan persamaan :

$$\delta a_i = \min \{ \max[\gamma(1 - E_i), -a_i], C - a_i \} \quad (12)$$

- c) Menghitung nilai Lagrange Multipliers (a_i) dengan persamaan:

$$a_i = a_i + \delta a_i \quad (13)$$

4. Melakukan perulangan pada tahap 3 hingga mendapat iterasi yang maksimum atau mencapai persamaan:

$$\max(|\delta a_i|) < \epsilon \quad (14)$$

5. Memperoleh nilai support vector (SV), yaitu nilai SV sama dengan nilai Lagrange Multipliers (a_i) lebih besar dari treshholdSV yang nilainya 0

6. Menghitung nilai bias (b) dengan persamaan:

$$b = -\frac{1}{2} (\sum_{i=0}^n \alpha_i y_i K(x_i, x^-) + \sum_{i=0}^n \alpha_i y_i K(x_i, x^+)) \quad (15)$$

Keterangan:

x_i : Data latih

x^+ : Data latih kelas positif dengan nilai a paling tinggi

x^- : Data latih kelas negatif dengan nilai a paling tinggi

7. Menghitung $f(x)$ yaitu nilai fungsi dengan persamaan

$$f(x) = \sum_{i=0}^n \alpha_i y_i K(x_i, x) + b \quad (16)$$

Keterangan:

$f(x)$: Hasil klasifikasi

y_i : Kelas data latih

2.5. K-Fold Cross Validation

K-fold cross validation adalah metode pengujian untuk mengoptimasi nilai parameter. Selain itu, k-fold cross validation digunakan untuk mengevaluasi dan membandingkan suatu metode dengan cara membagi dataset menjadi data uji dan data latih (Hesay, 2021). Metode ini dilakukan dengan cara membagi dataset sejumlah fold yang ditentukan dengan menggunakan 1 bagian data sebagai data uji dan bagian sisa data lainnya untuk data latih.

2.6. Confussion Matrix

Confussion matrix adalah salah satu metode pengujian pada Supervised Machine Learning untuk mengevaluasi kinerja dari algoritma. Pengukuran kualitas dari suatu klasifikasi dibangun dari sebuah confusion matrix dengan mencatat jumlah hasil klasifikasi yang benar dan salah (Sokolova, Japkowicz, & Szpakowicz, 2006). Pada Tabel 1 berikut akan merepresentasikan confusion matrix dari kasus yang memiliki dua jenis kelas saja.

Tabel 1 Confusion Matrix

Class	Predictive Positive	Predictive Negative
Actual Positive	TP	FN
Actual Negative	FP	FP

Keterangan:

- True Positive (TP): jumlah data kelas positif yang diklasifikasikan dengan tepat.
- True Negative (TN): jumlah data kelas negatif yang diklasifikasikan dengan tepat.
- False Negative (FN): jumlah data kelas negatif yang diklasifikasikan sebagai kelas positif.
- False Positif (FP): jumlah data kelas positif yang diklasifikasikan sebagai kelas negative.

Dari confusion matrix tersebut dapat ditentukan nilai dari kriteria performa klasifikasi dengan perhitungan untuk mengetahui *accuracy*, *precision*, *recall*, dan *f-Measure* dengan penjabaran sebagai berikut:

Accuracy untuk menggambarkan persentase data testing yang diklasifikasikan dengan benar. Akurasi merupakan kriteria yang kurang tepat jika bekerja pada kelas tidak seimbang karena kelas minoritas memiliki sumbangsih yang sedikit pada kriteria ini (Siringoringo, 2018), persamaan accuracy yaitu sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (17)$$

Precision untuk mengukur tingkat ketepatan sistem dalam memberikan data yang relevan dari semua data yang dipilih. Selain itu, merupakan data yang diambil dari informasi yang kurang, persamaan precision yaitu sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \tag{18}$$

Recall merupakan tingkat keberhasilan sistem dari jumlah kesesuaian data yang relevan dan didapatkan dari seluruh data yang relevan, persamaan precision yaitu sebagai berikut:

$$Recall = \frac{TP}{TP+FN} \tag{19}$$

F-Measure merupakan rata-rata dari nilai evaluasi yaitu precision dan recall yang didapatkan dari kombinasi nilai tersebut. Rata-rata kedua nilai evaluasi tersebut cenderung lebih dekat ke yang lebih kecil dari keduanya. Oleh karena itu, nilai *f-measure* yang tinggi memastikan bahwa nilai recall dan presisi cukup tinggi. (Gu, Zhu, & Cai, 2009). Pada kasus distribusi kelas yang tidak seimbang, *f-measure* berperan dalam menghindari kerugian terhadap matriks seperti tingkat kesalahan yang ada (Haryono & Bachtiar, 2020). Persamaan *f-measure* yaitu sebagai berikut:

$$F - Measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{20}$$

3. METODOLOGI

3.1. Pengumpulan Data

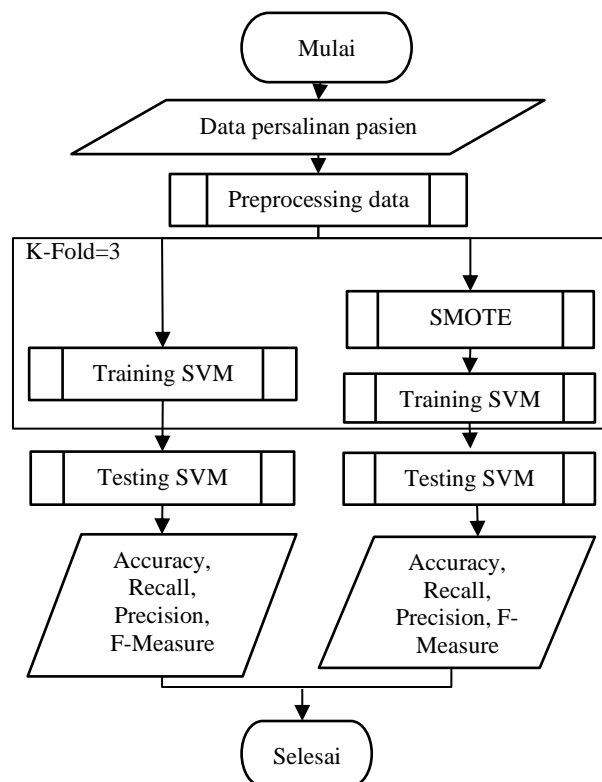
Pengumpulan data dilakukan dengan menggunakan data sekunder yaitu pengambilan data secara tidak langsung. Data yang diambil berasal dari puskesmas Ardimulyo Kabupaten Malang. Data tersebut berasal dari periode januari hingga september 2021. Data bersifat kontinyu berupa numerik. Sebanyak 97 data asli terdiri dari 8 data kelas BBLR serta 89 data kelas tidak BBLR. Dari data asli ditentukan bahwa sebanyak 78 data menjadi data latih dan 20 data uji. Data latih terdiri dari 70 data dengan kelas tidak BBLR dan 6 data dengan kelas BBLR. Data uji terdiri dari 18 data kelas tidak BBLR dan 2 data kelas BBLR.

3.2. Perancangan Algoritma

Sistem yang dibangun berupa sistem untuk melakukan klasifikasi berat badan lahir rendah terhadap data rekam medis persalinan metode yang digunakan adalah Support Vector Machine dan dengan teknik SMOTE untuk menangani imbalanced class. Data yang digunakan berjumlah 98 yang terbagi menjadi 90 tidak BBLR dan 8 BBLR dengan 8 fitur.

Sistem digambarkan secara umum pada Gambar 1 dengan melakukan beberapa tahapan. Sistem yang dibangun terdapat dua proses untuk

membandingkan metode SVM dengan SMOTE dan tanpa SMOTE. Tahap awal yang dilakukan dengan melakukan preprocessing untuk mengolah data dalam menangani missing value dan normalisasi pada data. Hasil normalisasi data yang minor kemudian dilakukan pembagian data dengan k-fold validation pada proses yang pertama dan kedua. Pada proses yang pertama hasil pembagian data k-fold tersebut digunakan untuk tahap training, sedangkan pada tahap kedua hasil k-fold digunakan sebagai input untuk melakukan smote sehingga menduplikasi data minoritas. Data sintesis yang sudah dibentuk dari proses SMOTE kemudian digabungkan dengan data mayor lalu dilakukan pelatihan menggunakan data training pada proses training SVM pada proses kedua. Tahap testing dilakukan untuk kedua proses terhadap data testing untuk mendapatkan hasil klasifikasi berupa kelas BBLR atau tidak BBLR. Sehingga dari hasil klasifikasi tersebut, kemudian dilakukan tahapan perhitungan confusion matrix untuk mendapat output berupa hasil evaluasi berupa nilai *accuracy*, *recall*, *precision*, dan *f-measure* terhadap masing-masing prosesnya.



Gambar 1. Diagram Alir Umum Sistem

4. PENGUJIAN DAN ANALISIS

Pengujian pada penelitian ini dilakukan

berdasarkan perancangan pengujian yang sudah dibuat sebelumnya. Pada pengujian ini menggunakan perbandingan dari hasil perhitungan confusion matrix terhadap pengujian metode SVM dengan SMOTE dan tanpa SMOTE, pengujian kernel, pengujian tiap parameter yang sudah ditentukan. Pengujian dilakukan dengan menggunakan 3-fold cross Validation pada data training yang bertujuan untuk mencari akurasi pelatihan model yang terbaik. Hasil pengujian 3-Fold tersebut kemudian dibandingkan dengan hasil pengujian untuk data testing dengan tujuan untuk membandingkan model mana yang paling baik. Pengujian k-Fold dilakukan dengan membagi data training menjadi 3 fold dimana 1 fold sebagai data uji dan 2 fold lainnya sebagai data latih. Pada pengujian 3-fold ini hasil yang ditampilkan berupa rata-rata dari confusion matriks yang diperoleh. Pengujian diawali dengan mencari nilai SVM yang memiliki confusion matriks dengan nilai tertinggi pada masing-masing metode dengan parameter yang sama dan telah ditentukan.

4.1. Pengujian Kernel Support Vector Machine (SVM) tanpa SMOTE menggunakan 3-Fold Cross Validation

Pengujian kernel pada metode SVM tanpa smote digunakan untuk membandingkan kernel yang paling optimal pada klasifikasi ini. Selain itu, untuk mengetahui hasil klasifikasi metode SVM tanpa SMOTE dengan menggunakan K-fold. Parameter yang digunakan pada pelatihan ini yaitu λ (*lambda*) bernilai 0,5, C bernilai 1 (*complexity*), γ (*gamma*) bernilai 0,005, Iterasi Maksimum bernilai 10, dan ϵ (*epsilon*) bernilai 0,0001. Berikut adalah hasil pengujian kernel linier dan kernel rbf pada metode SVM yang ditunjukkan pada Tabel 2.

Tabel 2. Hasil Pengujian Kernel Support Vector Machine (SVM) tanpa SMOTE Menggunakan 3-Fold Cross Validation

Hasil evaluasi	Tanpa SMOTE dengan SVM Kernel Linear	Tanpa SMOTE dengan SVM Kernel RBF
Accuracy	0,6923	0,8077
Precision	0,2778	0
Recall	0,0953	0
F-measure	0,1407	0

Hasil dari pengujian tersebut menunjukkan nilai confusion matriks terbaik diperoleh dari kernel linier karena memperoleh nilai *precision*,

recall, dan *f-measure* yang lebih besar dibandingkan dengan hasil yang diperoleh dari kernel RBF. Sedangkan, pada kernel RBF didapatkan akurasi terbaik saja.

4.2. Pengujian Kernel Support Vector Machine (SVM) dengan SMOTE menggunakan 3-Fold Cross Validation

Pada pengujian ini diterapkan pada metode SVM dengan SMOTE untuk membandingkan kernel yang terbaik. Nilai parameter yang digunakan pada pengujian ini sama dengan nilai pada pengujian sebelumnya. Berikut merupakan hasil pengujian kernel linier dan kernel rbf pada metode SVM dengan SMOTE yang ditunjukkan pada Tabel 3.

Tabel 3. Hasil Pengujian Kernel Support Vector Machine (SVM) dengan SMOTE Menggunakan 3-Fold Cross Validation

Hasil evaluasi	SMOTE dengan SVM Kernel Linier	SMOTE dengan SVM Kernel RBF
Accuracy	0,5385	0,3718
Precision	0,3889	0,7222
Recall	0,0704	0,0775
F-measure	0,1191	0,1378

Dari table diatas ditunjukkan bahwa hasil yang paling optimal diperoleh dari kernel RBF untuk metode SVM dengan SMOTE karena memiliki hasil *precision*, *recall*, dan *f-measure* yang lebih baik jika dibandingkan dengan hasil pengujian kernel linier yang hanya mendapat nilai akurasi terbaik saja.

4.3. Pengujian Nilai Parameter Support Vector Machine (SVM) menggunakan 3-Fold Cross Validation

Pengujian parameter bertujuan untuk mencari nilai parameter terbaik dari masing-masing metode. Terdapat beberapa parameter yang diuji, yaitu λ (*lambda*), C (*complexity*), γ (*gamma*), Iterasi Maksimum, dan ϵ (*epsilon*). Pada pengujian ini menggunakan 3-Fold Cross Validation sehingga hasil confusion matiks yang ditampilkan merupakan rata-rata dari 3 fold untuk masing-masing metode.

Parameter terbaik untuk setiap metode ditunjukkan pada Tabel 4. Selain itu, hasil evaluasi dari pengujian seluruh parameter SVM untuk setiap metode dengan 3-fold cross validation ditunjukkan pada Tabel 5.

Tabel 4. Parameter Metode Support Vector Machine

(SVM) Terbaik dari Setiap Menggunakan 3-Fold Cross Validation

Metode	Parameter				
	λ	γ	C	Iterasi Maksimum	ϵ
SVM Linier	0,1	1	0,01	10	0,01
SVM RBF	0,1	0,001	0,1	10	0,00001
SVM linier + SMOTE	0,3	0,001	10	800	0,00001
SVM RBF + SMOTE	0,1	0,001	20	1000	0,001

Tabel 5. Hasil Pengujian Nilai Parameter *Support Vector Machine* (SVM) menggunakan 3-Fold Cross Validation

Metode	Evaluasi			
	Accuracy	Precision	Recall	F-measure
SVM Linier	0,6923	0,2778	0,0953	0,1407
SVM RBF	0,6923	0,3889	0,1273	0,1905
SVM linier + SMOTE	0,7564	0,2778	0,0926	0,1347
SVM RBF + SMOTE	0,4872	0,8333	0,1317	0,2122

Berdasarkan Tabel 5 dijabarkan bahwa metode yang paling baik merupakan metode SVM menggunakan SMOTE dengan kernel RBF karena memiliki nilai yang paling tertinggi pada hasil *precision*, *recall*, dan *f-measure* diantara metode yang lainnya. Metode yang memperoleh akurasi tertinggi yaitu metode SVM kernel linier dengan SMOTE. Dari hasil tersebut akan dihitung rata-ratanya dengan pengujian 3-fold cross validation sebelumnya yang tanpa menggunakan parameter optimal hal ini dilakukan untuk mengetahui metode yang paling baik untuk pengujian 3-fold cross validation pada data training. Berikut merupakan hasil rata-rata pengujian 3-fold cross validation untuk setiap metodenya.

Tabel 6. Hasil Evaluasi rata-rata 3-Fold Cross Validation

Metode	Evaluasi rata-rata 3-Fold Cross Validation			
	Accuracy	Precision	Recall	F-measure
SVM Linier	0,6923	0,2778	0,0953	0,1407
SVM RBF	0,6923	0,3889	0,1273	0,1905
SVM linier + SMOTE	0,7564	0,2778	0,0926	0,1347
SVM RBF + SMOTE	0,4872	0,8333	0,1317	0,2122

SVM linier	0,6923	0,2778	0,0953	0,1407
SVM RBF	0,7500	0,1945	0,0637	0,0953
SVM Linier + SMOTE	0,6474	0,3334	0,0815	0,1269
SVM RBF + SMOTE	0,4295	0,7777	0,1046	0,1750

Pengujian dengan 3-fold cross validation dari tabel Tabel 6 menunjukkan hasil rata-rata dari evaluasi terbaik pada metode SVM menggunakan SMOTE dengan kernel RBF karena memiliki hasil paling optimal diantara yang lain dengan perolehan *precision*, *recall*, dan *f-measure* tertinggi yaitu *accuracy* bernilai 0,4295, *precision* bernilai 0,7777, *recall* bernilai 0,1046, dan *f-measure* bernilai 0,1750. Akurasi tertinggi dihasilkan dari metode SVM kernel RBF dengan hasil 0,7500. Hasil pengujian ini kemudian dibandingkan dengan hasil pengujian data indepen untuk mengetahui metode yang mana yang memang terbaik.

4.3. Pengujian Pengaruh Penerapan SMOTE dengan Kernel Linear dan Kernel RBF pada data independen

Pengujian ini digunakan untuk membandingkan pengaruh proses SMOTE pada klasifikasi BBLR menggunakan metode SVM dengan klasifikasi tanpa menggunakan SMOTE menggunakan kernel linear dan kernel RBF pada Tabel 7 dan Tabel 8.

Tabel 7. Hasil Pengujian Pengaruh Tanpa SMOTE dengan SVM Kernel Linear dan Kernel RBF

Hasil evaluasi	Tanpa SMOTE dengan SVM Kernel Linear	Tanpa SMOTE dengan SVM Kernel RBF
Accuracy	0,5	0,75
Precision	0	0,5
Recall	0	0,2
F-measure	0	0,2857

Tabel 8. Hasil Pengujian Pengaruh SMOTE dengan SVM Kernel Linear dan Kernel RBF

Hasil evaluasi	SMOTE dengan SVM Kernel Linier	SMOTE dengan SVM Kernel RBF
Accuracy	0,45	0,25
Precision	0,5	0
Recall	0,0909	0
F-measure	0,1538	0

Dari pengujian di atas membuktikan bahwa penggunaan SMOTE pada klasifikasi BBLR dengan kernel RBF kurang mampu meningkatkan hasil evaluasi dengan seimbang, sedangkan untuk smote dengan kernel linier dapat meningkatkan hasil evaluasi. Hal tersebut dapat terjadi karena pada saat pembentukan data sintesis yang terbentuk tidak memperhatikan data mayoritas sehingga memungkinkan adanya tumpang tindih data yang dihasilkan kurang dapat mewakili kelas minoritas. Berdasarkan tabel di atas ditunjukkan bahwa metode terbaik yaitu SVM kernel RBF karena memiliki hasil evaluasi yang paling baik dan menghasilkan nilai tertinggi dari *accuracy*, *precision*, *recall*, dan *f-measure* dengan *accuracy* sebesar 0,75, *precision* sebesar 0,5, *recall* sebesar 0,2, dan *f-measure* sebesar 0,2857. Hasil pengujian ini jika dibandingkan dengan hasil pengujian 3-fold cross validation tidak sama karena mendapat metode terbaik yang berbeda.

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian klasifikasi berat badan lahir rendah (BBLR) menggunakan metode *Support Vector Machine* dengan teknik SMOTE untuk pengujian 3-fold cross validation untuk klasifikasi Berat Badan Lahir Rendah pada penelitian ini dapat disimpulkan bahwa model parameter ditentukan dengan pengujian pengujian 3-fold cross validation memperoleh hasil yang terbaik pada kernel RBF dengan parameter *lamda* bernilai 0,1, *gamma* bernilai 0,001, *complexity* 20, iterasi maksimum bernilai 100, dan *epsilon* bernilai 0,001. Selain itu, hasil pengujian metode *Support Vector Machine* dengan teknik SMOTE terhadap data testing pada penelitian ini memiliki kinerja yang rendah dibandingkan dengan metode *Support Vector Machine* tanpa smote dengan kernel RBF. Pada pengujian ini metode yang terbaik adalah SVM tanpa SMOTE dengan kernel RBF memperoleh hasil evaluasi yaitu, *accuracy* bernilai 0,75, *precision* bernilai 0,5, *recall* bernilai 0,2, dan *f-measure* bernilai 0,2857.

Saran yang diperoleh dari penelitian ini sehingga dapat dikembangkan untuk penelitian selanjutnya adalah menambahkan jumlah data latih sehingga dapat meningkatkan hasil evaluasi. Selain itu, perlu melakukan model *experiment* pada jumlah data dengan perbandingan jumlah pada pada setiap kelas untuk mengetahui perilaku data dan dapat

menggunakan perbandingan model resampling yang lainnya yaitu seperti kombinasi SMOTE dengan *both sampling* yaitu dengan *oversampling* dan *undersampling*.

6. DAFTAR PUSTAKA

- Agustin, S., Setiawan, B. D., & Fauzi, M. A. (2019). Klasifikasi Berat Badan Lahir Rendah (BBLR) Pada Bayi Dengan Metode Learning Vector Quantization (LVQ). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2929-2936.
- Agustina, W., Furqon, M. T., & Rahayudi, B. (2018). Implementasi Metode Support Vector Machine (SVM) Untuk Klasifikasi Rumah Layak Huni (Studi Kasus: Desa Kidal Kecamatan Tumpang Kabupaten Malang). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3366-3372.
- Chawla, N. V. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* 16, 321–357.
- Dewi, C. (2013). Perbandingan Akurasi Backpropagation Neural Network dan ANFIS Untuk Memprediksi Cuaca. *Journal of Scientific Modeling & Computation*, 7-13.
- Febrianti, R. (2019). Faktor-Faktor Risiko yang Mempengaruhi Kejadian Berat Badan Lahir Rendah (BBLR) di RSUP DR. M. Djamil Padang Tahun 2019. *SCIENTIA JOURNAL*, 464-469.
- Gu, Q., Zhu, L., & Cai, Z. (2009). Evaluation Measures of the Classification Performance of Imbalanced Data Sets . 461–471.
- Haryono, D. A., & Bachtiar, F. A. (2020). Klasifikasi Ulasan Palsu menggunakan MDLText dengan Seleksi Fitur IG dan SMOTE. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1536-1545.
- Hasibuan, C. A., Mukid, M. A., & Prahutama, A. (2017). Klasifikasi Diagnosa Penyakit Demam Berdarah Dengue (DBD) Menggunakan Support Vector Machine (SVM) Berbasis GUI MATLAB. *JURNAL GAUSSIAN*, 171-180.

- Hesay, I. K. (2021). Analisis Sentimen Ulasan Pengunjung Simpang Lima Gumul Kediri menggunakan Metode BM25 dan Neighbor-Weighted K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3160-3169.
- Hidayat, W., Ardiansyah, M., & Setyanto, A. (2021). Pengaruh Algoritma ADASYN dan SMOTE terhadap Performa Support Vector Machine pada Ketidakseimbangan Dataset Airbnb. *Edumatic: Jurnal Pendidikan Informatika*, 11-20.
- Jaya, N. (2009). Analisis Faktor Resiko Kejadian Bayi Berat Lahir Rendah di Rumah Sakit Ibu dan Abak Siti Fatimah Kota Makassar. *Media Gizi Pangan*, 51-54.
- Oganis, C. (2017). Klasifikasi Status Gizi Ibu Hamil untuk Mengidentifikasi Bayi Berat Lahir Rendah (BBLR) Menggunakan Metode Support Vector Machine (SVM) (Studi Kasus di Puskesmas Labuan). *Jurnal Ilmiah Matematika Terapan*, 144 - 151.
- Phyu, T. N. (2009). Survey of Classification Techniques in Data Mining. Hong Kong.
- Qasthari, E. T. (2017). Teknik Pengukuran: Metode Klasifikasi Support Vector Machine (SVM) pada Data Pengukuran.
- Samosir, R. O., Wilandari, Y., & Yasin, H. (2015). Perbandingan Metode Klasifikasi Regresi Logistik Biner dan Radial Basis Function Network pada Berat Bayi Lahir Rendah (Studi Kasus: Puskesmas Pamenang Kota Jambi). *JURNAL GAUSSIAN*, 997-1005.
- Setiati, A. R., & Rahayu, S. (2017). Faktor yang Mempengaruhi Kejadian BBLR (Berat Badan Lahir Rendah) di Ruang Perawatan Intensif Neonatus RSUD DR Moewardi di Surakarta. *Jurnal Keperawatan Global*, 9-20.
- Siringoringo, R. (2018). Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan k-Nearest Neighbor. *ISD*, 44-49.
- Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond Accuracy, F-score and ROC: a Family of Discriminant Measures for Performance Evaluation.
- Sulistiyono, M., Pristyanto, Y., Adi, S., & Gumelar, G. (2021). Implementasi Algoritma Synthetic Minority Over-Sampling Technique untuk Menangani Ketidakseimbangan Kelas pada Dataset Klasifikasi. *Jurnal Sistem Informasi*, 445-459 .
- Tambun, M. S., Furqon, M. T., & Widodo, A. W. (2018). Penerapan Algoritme Jaringan Syaraf Tiruan Backpropagation pada Pengklasifikasian Status Gizi Balita. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3074-3080 .
- Trihardiani, I. (2011). Faktor Risiko Kejadian Berat Badan Lahir Rendah di Wilayah Kerja Puskesmas Singkawang Timur dan Utara Kota Singkawang. 1-55.
- Wibawa, A. P., Purnama, M. G., Akbar, M. F., & Dwiyanto, F. A. (2018). Metode-metode Klasifikasi. *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi* , (pp. 134-138). Malang.
- Yuliati, I. F., & Sihombing, P. R. (2021). Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia. *Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, 41-426.
- Yustihan, S. R., Adikara, P. P., & Indriati. (2021). Analisis Sentimen berbasis Aspek terhadap Data Ulasan Rumah Makan menggunakan Metode Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1017-1023.