



Klasifikasi Kejadian Berat Bayi Lahir Rendah Menggunakan Metode *K-Nearest Neighbor* (Studi Kasus Puskesmas Silo II Kabupaten Jember)

Faiz Robbani*, Mudafiq Riyan Pratama, Ervina Rachmawati, Mochammad Choirur Roziqin

Manajemen Informasi Kesehatan, Jurusan Kesehatan, Politeknik Negeri Jember

*Correspondence: Faiz Robbani

Email: faizrobbani.fr@gmail.com

Received: 27-08-2024

Accepted: 07-11-2024

Published: 19-08-2025



Copyright: © 2025 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license

(<http://creativecommons.org/licenses/by/4.0/>).

Abstrak: Tingginya angka kematian neonatal (AKN) di Indonesia dipengaruhi oleh kasus berat badan lahir rendah (BBLR), yang menjadi penyebab utama kematian neonatal, terutama di Kabupaten Jember, Jawa Timur, pada 2020–2022. Kondisi ini berisiko menghambat pencapaian target SDGs 2030 untuk menurunkan AKN. Penelitian ini bertujuan melakukan klasifikasi kejadian BBLR menggunakan metode *K-Nearest Neighbor* (K-NN). Data berasal dari riwayat pemeriksaan kehamilan di Puskesmas Silo II, Kabupaten Jember, dengan 1341 data kotor, dan 705 data yang digunakan setelah preprocessing, melibatkan 15 variabel. Pengujian akurasi metode K-NN dilakukan menggunakan confusion matrix dan nilai AUC ROC. Data risiko kejadian BBLR mencakup usia ibu, status kelahiran, berat bayi lahir, dan faktor lainnya. Hasil penelitian menunjukkan bahwa model terbaik dalam klasifikasi kejadian BBLR dengan metode K-NN menggunakan perbandingan rasio data uji dan data latih 90:10 dengan type shuffled sampling dan nilai $K=45$ mencapai akurasi 92.86% dan nilai AUC ROC sebesar 0.834, yang tergolong baik. Saran untuk penelitian selanjutnya diharapkan memperhatikan keseimbangan data dengan metode penanganan data imbalanced, yang sering terjadi pada kasus data mining di bidang kesehatan. Hasil penelitian ini juga diharapkan dapat dikembangkan menjadi sistem deteksi dini BBLR untuk mendukung Indonesia mencapai SDGs.

Katakunci: BBLR, Kematian Neonatal, Kinerja, KNN

Pendahuluan

Rekam medis pada fasilitas pelayanan kesehatan menjadi sumber informasi yang memerlukan pengelolaan yang profesional dan kompeten untuk memenuhi kebutuhan berbagai aspek yang meliputi administrasi, hukum, keuangan, penelitian, pendidikan, pendokumentasian, dan kesehatan masyarakat [1]. Dalam melaksanakan pekerjaannya, perekam medis memiliki kewajiban membantu program pemerintah dalam rangka meningkatkan derajat kesehatan masyarakat [2]. Salah satu program pemerintah yaitu upaya pemeliharaan kesehatan bayi dan anak yang ditujukan untuk mempersiapkan generasi sehat, cerdas, dan berkualitas serta untuk menurunkan angka kematian bayi dan anak. Salah satu upaya pemeliharaan kesehatan anak, dilakukan sejak masih dalam kandungan, pada saat dilahirkan, setelah dilahirkan, bahkan sampai berusia 18 tahun. Upaya kesehatan anak diharapkan mampu menurunkan angka kematian anak. Indikator

angka kematian yang berhubungan adalah angka kematian neonatal, angka kematian bayi, dan angka kematian balita [3].

Indonesia beserta dengan negara-negara lain memiliki komitmen untuk mencapai tujuan *Sustainable Development Goals (SDGs)* yang memiliki target pada tahun 2030 dapat menurunkan Angka Kematian Neonatal (AKN) hingga 12 per 1.000 kelahiran hidup [4]. Indonesia memiliki AKN sebesar 15 per 1.000 kelahiran berdasarkan Hasil Survei Demografi dan Kesehatan Indonesia (SDKI) tahun 2017 dan yang menjadi penyebab terbesar pada kejadian kematian neonatal yaitu kondisi berat badan lahir rendah sebesar 7.1%. Berat badan lahir rendah (BBLR) merupakan kelompok berat badan bayi lahir dengan kondisi < 2500 gram yang ditimbang dalam waktu 1 jam pertama setelah lahir [5]. Selain menjadi penyebab terbesar kematian neonatal, berat bayi lahir rendah menyebabkan risiko lebih besar pada bayi untuk stunting dan mengidap penyakit tidak menular seperti diabetes, hipertensi, dan penyakit jantung saat dewasa [6]. Analisis data kesehatan kejadian BBLR tahun 2020-2022 menurut data dari profil Dinas Kesehatan Provinsi Jawa Timur tahun 2020-2022 khususnya di Kabupaten Jember setiap tahunnya BBLR menjadi penyebab terbesar kematian neonatal dengan persentase dalam tabel 1 dibawah ini.

Tabel 1. Persentase Kematian Neonatal Akibat BBLR di Kabupaten Jember

Tahun	Persentase (%)
2020	38.5
2021	40.2
2022	38

Tabel 1, merupakan data dari Dinas Kesehatan Provinsi Jawa Timur yang menunjukkan Kabupaten Jember pada tahun 2020 menjadi peringkat pertama di Jawa Timur dalam kasus kematian neonatal akibat BBLR dengan jumlah kasus 104 kasus dengan persentase 38.5% dari keseluruhan kasus kematian neonatal [7]. Tahun 2021 kasus mengalami penurunan menjadi 88 kasus akan tetapi masih menjadi peringkat pertama di Jawa Timur dan untuk persentase mengalami kenaikan menjadi 40.2 dikarenakan jumlah kasus kematian neonatal juga mengalami penurunan dari 270 kasus di tahun 2020 menjadi 219 pada tahun 2021 [8]. Terjadi peningkatan angka kasus pada tahun 2022 menjadi 90 kasus akan tetapi persentase dilihat dari data Dinas Kesehatan Provinsi Jawa Timur pada tahun 2022 mengalami penurunan menjadi 38% karena kenaikan angka kematian neonatal dari 219 [9]. Pada tahun 2021 menjadi 237 di tahun 2022 dan dari data kabupaten jember masih menjadi peringkat pertama dari tahun 2020 hingga 2022 dalam kasus kematian neonatal akibat BBLR. Persentase kejadian kematian neonatal di Kabupaten Jember tahun 2020 hingga 2022 tergambar dalam tabel 2 berikut.

Tabel 2. Persentase Kematian Neonatal Akibat BBLR di Puskesmas Silo II Kabupaten Jember

Tahun	Persentase (%)
2020	31.25
2021	20

2022

12.5

Dari Tabel 2, Puskesmas Silo II yang merupakan puskesmas yang berada di wilayah kerja Kabupaten Jember dalam Data Profil Kesehatan Kabupaten Jember 2020-2022 tercatat angka kematian neonatal akibat BBLR berdasarkan Dinas Kesehatan Kabupaten Jember (2021) menunjukkan data kematian neonatal akibat BBLR di Puskesmas Silo II pada tahun 2020 sebanyak 5 kasus dari 16 kasus kematian neonatal dengan persentase 31.25% [10]. Tahun 2021 mengalami penurunan kasus menjadi 20% dengan angka 2 kasus dari 10 kasus kematian neonatal menurut data dari Dinas Kesehatan Kabupaten Jember dalam profil kesehatan kabupaten jember 2021 [11]. Data dalam profil kesehatan jember 2022 dari Dinas Kesehatan Kabupaten Jember menunjukkan penurunan kembali kasus kematian neonatal akibat BBLR di puskesmas silo II menjadi 1 kasus dari 8 kasus kematian neonatal dengan persentase 12.5% [12].

Berdasarkan data di atas, maka dibutuhkan suatu upaya dalam menekan angka kematian neonatal sebagai target untuk mencapai tujuan *Sustainable Development Goals* (SDGs) dalam menurunkan Angka Kematian Neonatal (AKN) hingga 12 per 1.000 kelahiran hidup pada tahun 2030, diperlukan upaya untuk mengurangi kematian neonatal, terutama pada kasus berat bayi lahir rendah. Salah satu upaya yang dapat dilakukan adalah melakukan klasifikasi kejadian BBLR untuk menentukan tindakan yang tepat melalui deteksi dini menggunakan teknologi kecerdasan buatan dan *data mining* berbasis *machine learning*. Salah satu algoritma umum untuk klasifikasi adalah *K-Nearest Neighbor* (*K-NN*). *K-Nearest Neighbor* adalah suatu pendekatan untuk menghitung kedekatan antara kasus baru dengan kasus lama yang berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada [13]. Proses pencarian kedekatan antara kasus baru dan lama dalam algoritma *K-Nearest Neighbor* (*K-NN*) melibatkan perhitungan jarak antara data latih dan data uji [14]. Untuk menguji akurasi metode *K-NN*, digunakan metode *confusion matrix*, yang sering dipakai dalam evaluasi model data *mining* klasifikasi dengan mengukur akurasi atau tingkat kesalahan. Proses pengujian akurasi ini menjadi dasar analisis performa metode *K-NN* dalam klasifikasi kejadian BBLR dan digunakan untuk pengembangan sistem deteksi dini kejadian BBLR dengan data *mining* berbasis *machine learning* kedepannya [15].

Berdasarkan uraian permasalahan tersebut, maka dibutuhkan penelitian untuk melakukan klasifikasi kejadian berat badan lahir rendah menggunakan metode *K-Nearest Neighbor* (Studi Kasus Puskesmas Silo II Kabupaten Jember) yang bertujuan untuk melakukan klasifikasi kejadian berat badan lahir rendah menggunakan metode *K-Nearest Neighbor* (Studi Kasus Puskesmas Silo II Kabupaten Jember) dengan melakukan *preprocessing data*, analisis hasil *preprocessing*, serta transformasi data melalui *encoding* dan normalisasi menggunakan *min-max normalization*, mencari nilai akurasi dan *AUC ROC* terbaik untuk menentukan model paling efektif berdasarkan *sampling type*, *rasio data split*, dan nilai *K*, serta menganalisis hasil menggunakan *confusion matrix* yang mencakup nilai akurasi, *presisi*, *recall*, dan *F1 Score* sehingga dapat menjadi dasar untuk memecahkan masalah untuk melakukan klasifikasi kejadian berat bayi lahir rendah menggunakan metode *K-Nearest Neighbor* yang hasilnya diharapkan dapat menjadi dasar pengembangan

sistem pendukung keputusan untuk melakukan deteksi dini kejadian berat bayi lahir rendah berbasis data medis dengan metode *K-Nearest Neighbor* pada penelitian berikutnya.

Metode Penelitian

Jenis/Desain Penelitian

Jenis penelitian yang digunakan adalah penelitian kuantitatif. Penelitian kuantitatif digunakan untuk meneliti populasi atau sampel tertentu, teknik pengambilan sampel pada umumnya dilakukan secara random, pengumpulan data menggunakan instrumen penelitian, analisis data bersifat kuantitatif dengan tujuan untuk menguji hipotesis yang telah ditetapkan dengan kejadian berat bayi lahir rendah menggunakan metode *K-Nearest Neighbor* (Studi Kasus Puskesmas Silo II Kabupaten Jember).

Unit Analisis

Subjek penelitian ini adalah data riwayat pemeriksaan kehamilan ibu hamil, petugas rekam medis, ahli gizi, dan bidan yang mencatat riwayat pemeriksaan kehamilan ibu hamil di Posyandu Wilayah Kerja Puskesmas Silo II. Sedangkan objek dalam penelitian ini yaitu data berdasarkan riwayat pemeriksaan kehamilan yang di dapat dari kohort ibu hamil, register ibu hamil, dan riwayat pemeriksaan kehamilan lain yang tercatat oleh bidan di Posyandu Wilayah Kerja Puskesmas Silo II.

Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini yaitu studi literatur dan dokumentasi.

Metode Analisis Data

Pada penelitian ini terdapat serangkaian langkah untuk untuk memecahkan masalah. Dimulai dengan penyediaan data, kemudian melakukan serangkaian *preprocessing* data. Tahapan sebelum klasifikasi, data dibersihkan, diintegrasikan, dilakukan proses imputasi data. Selanjutnya dilakukan analisis data setelah proses *preprocessing* sebelum proses transformasi dilakukan untuk merubah data agar bisa sesuai dengan bentuk data yang dapat digunakan dalam proses berikutnya, dan dinormalisasikan. Pengujian dilakukan dengan membagi dataset menjadi data latih dan data uji. Setelah membagi data, klasifikasi menggunakan metode yang digunakan dalam *data mining* dapat dilakukan. Langkah terakhir adalah melakukan evaluasi hasil.

Dalam proses *preprocessing* dilakukan *cleaning data*. *Cleaning data* dilakukan untuk penyesuaian data agar sesuai dengan kebutuhan data dan melindungi data yang tergolong dalam data privasi agar tidak melanggar etik penelitian khususnya dalam data kesehatan yang privasinya harus dijaga. Proses berikutnya dilakukan analisis data secara kuantitatif sebelum dilakukan transformasi data dengan proses konversi atau encoding. Dilanjutkan dengan proses normalisasi data menggunakan *min-max normalisasi* agar data menjadi sesuai dengan kebutuhan data untuk digunakan pada proses data mining.

Proses data mining menggunakan metode *K-Nearest Neighbor* dilakukan setelah data sesuai. Evaluasi hasil pengujian data mining dengan metode *K-Nearest Neighbor* menggunakan metode *confusion matrix* untuk melihat nilai akurasi beserta *presisi*, *recall*, *F1-*

Score dan Nilai AUC menggunakan ROC Curve. Pengambilan Kesimpulan dilakukan dengan melihat nilai dan menginterpretasikan nilai yang ditemukan dan dinarasikan dengan mengaitkan dengan referensi dan teori yang ada.

Hasil dan Pembahasan

Preprocessing, Analisis, dan Tranformasi Data

Penelitian dimulai dengan melakukan pengambilan data riwayat kehamilan pada posyandu di wilayah kerja puskesmas Silo II sebagai sumber data dengan sistemasi pengambilan data dari kohort ibu hamil, register ibu hamil, dan riwayat pemeriksaan kehamilan lain yang tercatat oleh petugas posyandu menggunakan link *spreadsheet* yang dibagikan. Pada tahapan *preprocessing* data dilakukan beberapa proses terhadap data. Proses pertama adalah melakukan proses akumulasi data dari beberapa sumber menjadi satu kesatuan dataset. Proses berikutnya adalah membersihkan data (*data cleaning*). Proses dalam *cleaning data* terdiri dari beberapa tahapan yang terdiri dari memeriksa data dan mengeleminasi data yang memiliki *missing value*, memeriksa dan membenahi data yang tidak konsisten (inkonsisten), proses imputasi data, dan memeriksa dan mengeleminasi variabel yang tidak digunakan dalam penelitian [16]. Proses ketiga, merupakan proses menghilangkan inskonsistensi data. Proses empat, merupakan proses imputasi. Proses kelima, melakukan pengecekan ulang pada setiap variabel dan melakukan pembersihan pada variabel dengan data yang tidak memiliki keragaman data yang dibutuhkan dalam proses data *mining* karena akan berpengaruh pada jarak kedekatan (jarak euclidian), karena metode *K-Nearest Neighbor* melakukan klasifikasi dengan melihat jarak kedekatan kasus baru dengan kasus lama.

Dataset yang telah dilakukan proses *preprocessing* maka dapat terbaca dan dilakukan analisis terhadap data yang sudah bersih. Setelah data dianalisis dilakukan proses transformasi atau mengubah variabel kategori menjadi bentuk yang dapat diolah oleh algoritma atau disebut dengan konversi atau *encoding*. Proses transformasi setelah dilakukan konversi atau *encoding* dengan bantuan tabel konversi adalah proses normalisasi data menggunakan metode *min-max* normalisasi dilakukan pada dataset untuk mengubah data asli ke dalam rentang nilai baru, sehingga menghasilkan nilai keseimbangan tolak ukur sebelum dan sesudah proses. Setelah melalui tahap normalisasi data menggunakan *min-max* normalisasi maka dataset siap digunakan pada proses berikutnya dalam proses data *mining* menggunakan metode *K-Nearest Neighbor*. *Min-max Normalization* adalah metode normalisasi data yang bertujuan mengubah data asli ke dalam rentang nilai baru, sehingga menghasilkan nilai keseimbangan tolak ukur sebelum dan sesudah proses [17]. Berikut adalah gambaran dataset ternormalisasi menggunakan *min-max normalitation*:

NO	Pekerjaan	Pendidikan	Jumlah Kehamilan	Usia Kandungan	Jarak Kehamilan	Status Gizi (LILA)	Penyakit Bawaan	Hamil Ganda (Kembar)	Perdarahan Antepartum	Preeklamsia	Ketuban Pecah Dini (KPD)	Anemia	Hipertensi	Cacat Bawaan	BBLR
1	0	0.6	0.50	1.00	0.50	1	0	0	0	0	0	0	0	0	NORMAL
2	0	0.4	0.50	0.00	1.00	1	0	0	0	0	0	0	0	0	NORMAL
3	0	0.6	0.00	0.00	0.50	1	0	0	0	0	1	0	0	0	NORMAL
4	0	0.4	0.50	1.00	0.50	1	0	0	0	0	0	0	0	0	NORMAL
5	0	0.6	0.50	1.00	0.50	0	0	0	0	0	0	0	0	0	NORMAL
6	0	0.4	0.00	1.00	0.50	1	0.2	0	0	0	0	0	0	0	NORMAL
7	0	0.2	0.50	0.00	1.00	1	0	0	0	1	0	0	0	0	NORMAL
8	0	0.2	0.50	1.00	0.50	1	0	0	0	0	0	1	0	0	NORMAL
9	0	0.6	0.00	0.00	0.50	0	0	0	0	0	0	0	0	0	NORMAL
10	0	0.2	0.50	0.00	0.50	1	0	0	0	0	0	0	0	0	BBLR
...
705	0	0.4	0.00	1.00	0.50	1	0	0	0	0	0	0	0	0	NORMAL

Gambar 1. Gambaran Dataset Ternormalisasi Menggunakan *Min-Max Normalitation*

Dari gambar 1., diatas merupakan dataset yang bersumber dari riwayat pemeriksaan kehamilan Puskesmas Silo II periode 2020-2022 menunjukkan bahwa faktor-faktor pekerjaan, jumlah kehamilan, pendidikan, usia kandungan, jarak kehamilan, status gizi (LILA), penyakit bawaan, hamil ganda, perdarahan antepartum, dan ketuban pecah dini berpengaruh signifikan terhadap kejadian BBLR (Berat Badan Lahir Rendah). Guru, IRT, kehamilan preterm, KEK, kehamilan ganda, perdarahan antepartum, dan ketuban pecah dini memiliki persentase kejadian BBLR yang lebih tinggi. Faktor pekerjaan yang menuntut, pendidikan rendah, kehamilan pertama, dan jarak kehamilan yang berisiko juga meningkatkan risiko BBLR. Penelitian ini mendukung teori-teori yang ada tentang pengaruh kondisi kerja, paritas, pendidikan, usia kehamilan, jarak antar kehamilan, status gizi, penyakit bawaan, hamil ganda, perdarahan antepartum, dan ketuban pecah dini terhadap risiko BBLR [4].

Pengujian Nilai Akurasi Untuk Rasio Data Split, Sampling Type, dan Nilai K

Dataset yang telah melalui tahapan *preprocessing* menghasilkan data bersih sebanyak 705 data dari 1341 data kotor yang terdiri dari data normal sebanyak 664 data dan data kejadian BBLR sebanyak 41 data. Data yang diujikan dalam keseluruhan dataset sebanyak 705 data. Dataset kemudian dilakukan split data untuk mencari rasio data split terbaik. Rasio data split yang diteliti adalah 90:10, 80:20, 70:30, 60:40 dan 50:50 yang akan diujikan pada dataset. Proses berikutnya setelah melakukan pengujian pada data split akan diujikan untuk melakukan klasifikasi untuk kejadian berat bayi lahir rendah. Pengujian rasio *data split* pada dataset sebanyak 705 data untuk melakukan pembagian antara *data training* dan *data testing* akan dilakukan menggunakan *tools split data RapidMiner*. *RapidMiner* merupakan aplikasi yang sering digunakan untuk untuk pemrosesan data mining karena memiliki banyak algoritma klasifikasi, pengelompokkan, dan analisis regresi dalam melakukan penambangan data [13].

Rasio yang akan dibandingkan pada penelitian ini yaitu rasio perbandingan 90:10, 80:20, 70:30, 60:40 dan 50:50 menggunakan 3 sampling type yang tersedia pada *tools RapidMiner* yang terdiri dari *linier sampling*, *shuffled sampling* dan *stratified sampling*. Pengujian ini juga menggunakan nilai K 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, dan 55 untuk diujikan dengan tujuan untuk mencari nilai paling efektif untuk klasifikasi kejadian berat bayi lahir rendah. Hasil pengujian pengaruh rasio data latih dan data uji, nilai K serta *type sampling* pada algoritma KNN dengan akurasi tertinggi dapat dilihat pada tabel 1 dibawah ini:

Tabel 3. Akurasi Tertinggi Rasio Data Split dan Nilai K dalam Setiap *Type Sampling* Pada Algoritma KNN

Metode Sampling	Rasio Data Split	Accuracy	Nilai K
<i>Linier Sampling</i>	60:40	93.97	7,9,31,33,35,37,39,41,43,45,47,49,51,53,55
<i>Shuffled Sampling</i>	50:50	95.45	35
<i>Stratified Sampling</i>	60:40 dan 80:20	94.33	31 dan 3

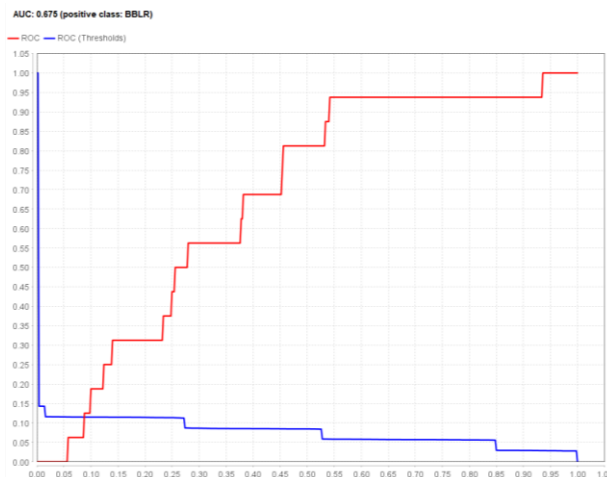
Berdasarkan tabel 1., akurasi tertinggi *rasio data split* dan nilai K dalam setiap *type sampling* pada algoritma K-NN. Terdapat fluktuasi nilai akurasi yang bervariasi, mengalami penurunan dan peningkatan. Hal ini terjadi karena distribusi data dalam satu kelas tidak terstruktur dengan baik. Sehingga data tetangga dalam satu kelas yang sama dapat tersebar. Berdasarkan hasil pengujian, dapat ditarik kesimpulan bahwa *rasio data split*, nilai K dan *type sampling* memiliki pengaruh terhadap nilai akurasi. Pada pengujian nilai akurasi tertinggi terdapat pada rasio 50:50 mencapai 95.45% dengan nilai K = 35 menggunakan tipe *sampling shuffled sampling*.

Pengujian AUC Untuk Rasio Data Split, Sampling Type, dan Nilai K

Pengujian AUC untuk rasio data *split*, *sampling type*, dan nilai K pada penelitian ini yaitu nilai AUC ROC model dengan akurasi paling tinggi dan evaluasi nilai AUC ROC untuk keseluruhan model penelitian.

A. Nilai AUC ROC Model Dengan Akurasi Paling Tinggi

ROC (*Receiver Operating Characteristic*) *curve evaluation* merupakan visualisasi dan pengelola klasifikasi berdasarkan kinerja menggunakan dua kelas keputusan positif atau negatif yang dipetakan ke salah satu elemen dari himpunan pasangan. Nilai AUC (*Area Under Curve*) digunakan untuk mengukur akurasi secara umum. Semakin mendekati 1 maka akan semakin baik uji pemodelannya [18]. Pada tahap ini dilakukan pengujian dan evaluasi menggunakan kurva ROC yang berdasarkan nilai *confusion matrix* untuk mengetahui nilai AUC. Hasil menggunakan kurva ROC untuk melihat nilai AUC pada klasifikasi data kejadian berat bayi lahir rendah menggunakan metode K-NN dapat di lihat pada kurva berikut:



Gambar 2. Kurva ROC dan Nilai AUC untuk K-NN Rasio 50:50 dan K = 35 dengan *Shufflef Sampling*

Gambar 2. menunjukkan *AUC* (*Area Under the Curve*) dari *ROC* (*Receiver Operating Characteristic*) pada pengujian dengan model rasio data split 50:50 dengan *shuffled sampling* menggunakan K = 35 menghasilkan nilai 0.675. *ROC curve* adalah alat evaluasi yang berguna untuk memvisualisasikan seberapa baik model mampu memisahkan kelas positif dan negatif pada berbagai *threshold* pengklasifikasi. Hasil *AUC* sebesar 0.675 menandakan bahwa model memiliki kinerja yang lemah menurut tabel 2.3 mengenai pengkategorian nilai *AUC* [19]. Semakin besar nilai *AUC*, semakin baik pula kemampuan diskriminatif model. Dalam penelitian ini dikarenakan model dengan rasio data 50:50 dengan K = 35 menggunakan metode tipe *sampling shuffled* dengan akurasi maksimal dari model lainnya sebesar 95.45% masuk dalam kategori lemah maka perlu dilihat untuk proporsi data karena ditakutkan tergolong dalam kasus *imbalanced data* [4]. Dilihat dari proporsi data dari tabel 4.8 menunjukkan komposisi 5.8% data kasus kejadian BBLR dengan angka 41 kasus dan 94.2% data NORMAL dengan angka 664 data. Berdasarkan proporsi data maka dataset tergolong dalam kasus *imbalanced data*. Jika dalam kasus *imbalanced data* tingkat akurasi tinggi, namun *sensitivity* atau *specificity* rendah, maka pengklasifikasian dapat dikatakan tidak baik.

B. Evaluasi nilai AUC ROC Untuk Keseluruhan Model Penelitian

Hasil pengujian pengaruh rasio data latih dan data uji, nilai K serta *type sampling* pada algoritma *KNN* dengan nilai *AUC* dalam kurva ROC tertinggi dapat dilihat pada tabel dibawah ini:

Tabel 4. Nilai ROC AUC Maksimal di Setiap *Sampling Type*

Metode Sampling	Rasio Data Split	Accuracy	Nilai K
Linier Sampling	90:10	0.655	23
Shuffeld Sampling	90:10	0.834	45
Stratified Sampling	80:20	0.621	3

Berdasarkan tabel 2., Nilai AUC ROC tertinggi dari variasi model *rasio data split* dan nilai K dalam setiap *type sampling* pada algoritma K-NN. Terdapat fluktuasi nilai AUC yang bervariasi, mengalami penurunan dan peningkatan. Hal ini terjadi karena distribusi data dalam satu kelas tidak terstruktur dengan baik. Sehingga data tetangga dalam satu kelas yang sama dapat tersebar. Berdasarkan hasil pengujian, dapat ditarik kesimpulan bahwa *rasio data split*, nilai K dan *type sampling* memiliki pengaruh terhadap nilai AUC pada kurva ROC. Pada pengujian nilai AUC ROC tertinggi terdapat pada rasio 90:10 mencapai 0.834 dengan nilai K = 45 menggunakan tipe *sampling shuffled sampling*.

Penghitungan Confusion Matrix Pada Model Paling Efektif Untuk Mengetahui Nilai Accuracy, Recall, Precision, dan F1-Score

Confusion Matrix merupakan pengukuran performa untuk permasalahan klasifikasi *machine learning* dimana luaran bisa berbentuk 2 kelas ataupun lebih. Pada prakteknya confusion matrix digunakan untuk menghitung nilai akurasi, presisi, dan *recall*. Nilai akurasi (*accuracy*) menggambarkan seberapa akurat model dalam mengklasifikasikan dengan benar [20]. *Precision* didefinisikan merupakan tingkatan ketepatan antara data yang diharapkan oleh pengguna dengan jawaban yang diberikan oleh sistem [21]. *Recall* merupakan tingkatan keberhasilan sistem dalam menciptakan kembali suatu data [21]. *F1-Score* adalah nilai yang menunjukkan seberapa baik presisi dan *recall* pada model yang dibangun [22]. Pada tahap ini dilakukan pengujian dan evaluasi menggunakan *confusion matrix* untuk mengetahui *accuracy*, *recall*, *precision*, dan *F1-Score*. Hasil uji terbaik pada klasifikasi data kejadian berat bayi lahir rendah dapat di lihat pada tabel berikut:

Tabel 5. Hasil *Confusion Matrix* dengan *Linier Sampling* Menggunakan Rasio Split Data 90:10 dengan K=45

Rasio Data	Variabel	K = 45
90 : 10	TP	65
	FP	5
	FN	0
	TN	0

Keterangan :

- True Positive (TP)* merupakan data positif yang diprediksi benar.
- True Negative (TN)* merupakan data negatif yang diprediksi benar.
- False Positive (FP)* tergolong error tipe 1 artinya data negatif namun diprediksi sebagai data positif.
- False Negative (FN)* tergolong error tipe 2 artinya data positif namun diprediksi sebagai data negatif.

Penjelasan dari tabel 3., diketahui hasil dari pengujian dengan model pengujian menggunakan rasio data split 90:10 dengan *shuffled sampling* dengan nilai K=45. Data total dalam penelitian 705 data sehingga pada model ini 90% yang berjumlah 635 menjadi *data*

training dan 10% yang berjumlah 70 data menjadi *data testing*. Hasil pengujian diketahui dari 70 data uji, 65 data diklasifikasikan benar dan sesuai masuk dalam kategori atau kelas "NORMAL" dalam prediksi yang dilakukan dengan metode Algoritma KNN. Terdapat 5 data yang diprediksi "NORMAL" akan tetapi sebenarnya masuk kedalam kelas atau kategori "BBLR". Hasil tabel *confusion matrix* dalam tabel 3 dapat dijadikan dasar untuk mencari nilai *accuracy*, *recall*, *precision*, dan *F1-Score* secara manual dengan persamaan sebagai berikut:

- a. Perhitungan *accuracy* pada *confusion matrix data testing* :

$$\begin{aligned} \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \\ \text{Accuracy} &= \frac{65 + 0}{65 + 0 + 5 + 0} \\ \text{Accuracy} &= \frac{65}{70} = 0.92857 \end{aligned}$$

- b. Perhitungan *Precision* pada *confusion matrix data testing* :

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP + FP} \\ \text{Precision} &= \frac{65}{65 + 5} = \frac{65}{70} = 0.92857 \end{aligned}$$

- c. Perhitungan *Recall* pada *confusion matrix data testing* :

$$\begin{aligned} \text{Recall} &= \frac{TP}{TP + FN} \\ \text{Recall} &= \frac{65}{65 + 0} = \frac{65}{65} = 1 \end{aligned}$$

- d. Perhitungan *F1Score* pada *confusion matrix data testing* :

$$\begin{aligned} \text{F1Score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\ \text{F1Score} &= 2 \times \frac{0.92857 \times 1}{0.92857 + 1} = 2 \times \frac{0.92857}{1.92857} = 2 \times 0.48148 = 0.9629 \end{aligned}$$

Dari hasil perhitungan *confusion matrix* yang dilakukan pada proses klasifikasi pada modul *classify* dengan metode *K-NN*, maka dihasilkan ringkasan nilai sebagai berikut:

Tabel 6. Hasil Perhitungan *Confusion Matrix*

Indikator	Nilai	Nilai (%)
<i>Accuracy</i>	0.92857	92.86
<i>Precision</i>	0.92857	92.86
<i>Recall</i>	1	100
<i>F1Score</i>	0.9629	96.3

Berdasarkan tabel 4., hasil perhitungan *confusion matrix* dengan metode *K-Nearest Neighbors (K-NN)*, diperoleh nilai akurasi sebesar 92.86%. Ini berarti model *K-NN* berhasil mengklasifikasikan sekitar 92.86% dari total data dengan benar. Akurasi yang relatif tinggi

ini menunjukkan bahwa model memiliki kemampuan yang baik dalam memprediksi kelas-kelas pada *dataset* yang digunakan [23]. Selanjutnya, nilai *precision* sebesar 92.86% menunjukkan bahwa ketika model *K-NN* mengklasifikasikan sebuah data sebagai positif, sekitar 92.86% dari prediksi tersebut benar-benar positif. Dengan kata lain, dari semua prediksi yang model buat sebagai positif, sebagian besar di antaranya memang relevan dan tepat.

Nilai *recall* sebesar 100% menggambarkan bahwa dari semua data yang benar-benar positif dalam *dataset*, model *K-NN* berhasil mengidentifikasi sekitar 100% dari mereka dengan benar. Ini menunjukkan bahwa model memiliki kemampuan baik dalam memprediksi kelas positif [24].

F1-score yang merupakan rata-rata harmonik dari *precision* dan *recall*, diperoleh sebesar 96.3%. *F1-score* merupakan metrik yang menyatukan kedua konsep tersebut menjadi satu nilai. Sebuah *F1-score* yang tinggi menunjukkan keseimbangan yang baik antara *precision* dan *recall*. Secara keseluruhan, hasil ini menunjukkan bahwa model *K-NN* memiliki performa yang baik dalam melakukan klasifikasi pada *dataset* yang digunakan [25].

Kesimpulan

Berdasarkan penelitian Klasifikasi Kejadian Berat Badan Lahir Rendah Menggunakan Metode *K-Nearest Neighbor* (Studi Kasus Puskesmas Silo II Kabupaten Jember) dapat disimpulkan bahwa tingkat akurasi terbaik dengan metode *K-NN* menggunakan perbandingan rasio data uji dan data latih 50:50 dengan *type shuffled sampling* dengan nilai $K=35$ mencapai nilai 95.45%. Nilai AUC pada model dengan akurasi terbaik dengan rasio data split 50:50 dengan *type shuffled sampling* dengan nilai $K=35$ mencapai nilai 0.675 dan tergolong lemah, sehingga dilakukan evaluasi ulang pada seluruh model menggunakan kurva ROC untuk mengetahui Nilai AUC terbaik untuk pemilihan model terbaik dan menghasilkan nilai AUC maksimal berada pada angka 0.834 pada model dengan rasio data split 90:10 dengan nilai $K = 45$ pada *sampling type shuffled sampling*. Selain itu, hasil dari perhitungan *confusion matrix* dengan metode *K-NN* model dengan rasio data split 90:10 dengan nilai $K = 45$ pada *sampling type shuffled sampling* menghasilkan nilai *accuracy* 92.86%, *precision* 92.86%, *recall* 100%, dan *F1Score* 96.3%. Dari hasil yang diperoleh telah tersebut, model *K-Nearest Neighbor* (*K-NN*) pada tingkat akurasi yang tinggi, nilai *precision*, *recall*, dan *F1Score* yang memuaskan, serta AUC yang termasuk dalam kategori baik, menunjukkan potensi model ini masih dapat diimplementasikan dalam aplikasi *machine learning*.

Saran yang diperoleh dari penelitian ini sehingga dapat dikembangkan untuk penelitian selanjutnya adalah diharapkan dapat memperhatikan keseimbangan data dalam penelitian antara data kasus dan data normal dalam penelitian berikutnya dengan menambahkan metode untuk menangani kasus *imbalanced data* karena biasanya dalam kasus data *mining* terutama dalam bidang kesehatan kasus *imbalanced data* sering terjadi. Serta, peneliti juga mengusulkan agar hasil penelitian ini dikembangkan menjadi sebuah

sistem deteksi dini berat badan lahir rendah (BBLR) yang dapat membantu mencapai tujuan SDGs (*Sustainable Development Goals*).

Daftar Pustaka

- [1] Kementerian Kesehatan Republik Indonesia, Keputusan Menteri Kesehatan Republik Indonesia Nomor HK.01.07/MENKES/312/2020 tentang Standar Profesi Perkam Medis dan Informasi Kesehatan. Jakarta, Indonesia, 2020.
- [2] Kementerian Kesehatan Republik Indonesia, Peraturan Menteri Kesehatan Republik Indonesia Nomor 55 Tahun 2013 tentang Penyelenggaraan Pekerjaan Perkam Medis. Jakarta, Indonesia, 2013.
- [3] Kementerian Kesehatan Republik Indonesia, Peraturan Menteri Kesehatan Republik Indonesia Nomor 25 Tahun 2014 tentang Upaya Kesehatan Anak. Jakarta, Indonesia, 2014.
- [4] P. R. Sihombing and I. F. Yuliaty, "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] I. Setiawati and A. A. Lase, "Hubungan Berat Badan Lahir dengan Ruptur Perineum Persalinan Normal di Puskesmas Sawo Kecamatan Sawo Kabupaten Nias Utara," *Jurnal Mutiara Kebidanan*, vol. 8, no. 2, pp. 59–64, 2022.
- [6] S. Nurkholivah and Y. Fatmawati, "Faktor Internal Ibu yang Berhubungan dengan Kejadian Berat Badan Lahir Rendah di RSUD dr. R. Soetrasno Rembang," *Jurnal Keperawatan dan Kesehatan Masyarakat Cendekia Utama*, vol. 12, no. 3, pp. 235–247, 2023, doi: 10.31596/jcu.v12i3.1276.
- [7] Dinas Kesehatan Provinsi Jawa Timur, Profil Kesehatan Provinsi Jawa Timur Tahun 2020. Surabaya, Indonesia, 2021.
- [8] Dinas Kesehatan Provinsi Jawa Timur, Profil Kesehatan Provinsi Jawa Timur Tahun 2021. Surabaya, Indonesia, 2022.
- [9] Dinas Kesehatan Provinsi Jawa Timur, Profil Kesehatan Provinsi Jawa Timur Tahun 2022. Surabaya, Indonesia, 2023.
- [10] Dinas Kesehatan Kabupaten Jember, Profil Kesehatan Kabupaten Jember Tahun 2020. Jember, Indonesia, 2021.
- [11] Dinas Kesehatan Kabupaten Jember, Profil Kesehatan Kabupaten Jember Tahun 2021. Jember, Indonesia, 2022.
- [12] Dinas Kesehatan Kabupaten Jember, Profil Kesehatan Kabupaten Jember Tahun 2022. Jember, Indonesia, 2023.
- [13] M. A. R. Sirait, D. Saripurna, and P. S. Ramadhan, "Sistem Pakar Mendiagnosa Penyakit Grandemultipara Menggunakan Metode Algoritma K-Nearest Neighbour," *Jurnal CyberTech*, vol. 2, no. 1, pp. 31–41, 2019.
- [14] A. S. P. Anugerah, I. Indriati, and C. Dewi, "Implementasi Algoritme Fuzzy K-Nearest Neighbor untuk Penentuan Lulus Tepat Waktu (Studi Kasus: Fakultas Ilmu Komputer Universitas Brawijaya)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 4, pp. 1726–1732, 2018.

-
- [15] A. Yudhana, S. Sunardi, and A. J. S. Hartanta, "Algoritma K-NN dengan Euclidean Distance untuk Prediksi Hasil Penggajian Kayu Sengon," *Transmisi: Jurnal Ilmiah Teknik Elektro*, vol. 22, no. 4, pp. 123–129, 2020, doi: 10.14710/transmisi.22.4.123-129.
- [16] D. Aprilla, D. A. Baskoro, L. Ambarwati, and I. W. S. Wicaksana, *Belajar Data Mining dengan RapidMiner*. Jakarta, Indonesia: Gramedia Pustaka Utama, 2013.
- [17] D. Fenny, "Analisis Perbandingan Cosine Normalization dan Min-Max Normalization pada Pengelompokan Terjemahan Ayat Al-Quran Menggunakan Algoritma K-Means Clustering," Bachelor's thesis, Fakultas Sains dan Teknologi, Universitas Islam Negeri Syarif Hidayatullah Jakarta, Jakarta, Indonesia, 2019.
- [18] P. A. Jusia, "Analisis Komparasi Pemodelan Algoritma Decision Tree Menggunakan Metode Particle Swarm Optimization dan Metode AdaBoost untuk Prediksi Awal Penyakit Jantung," in *Proc. Seminar Nasional Sistem Informasi (SENASIF)*, vol. 2, pp. 1048–1056, 2018.
- [19] K. Kristiawan and A. Widjaja, "Perbandingan Algoritma Machine Learning dalam Menilai Sebuah Lokasi Toko Ritel," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 7, no. 1, 2021, doi: 10.28932/jutisi.v7i1.3313.
- [20] P. Romadloni, B. A. Kusuma, and W. M. Baihaqi, "Komparasi Metode Pembelajaran Mesin untuk Implementasi Pengambilan Keputusan dalam Menentukan Promosi Jabatan Karyawan," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, pp. 622–628, 2022, doi: 10.36040/jati.v6i2.5708.
- [21] A. N. M. A. Muslim, B. Prasetyo, E. L. H. Mawarni, A. J. Herowati, Mirqotussa'adah, and S. H. Rukmana, *Data Mining Algoritma C4.5 Disertai Contoh Kasus dan Penerapannya dengan Program Komputer*. Semarang, Indonesia: Universitas Negeri Semarang Press, 2019.
- [22] S. Clara, D. L. Prianto, R. Al Habsi, E. F. Lumbantobing, and N. Chamidah, "Implementasi Seleksi Fitur pada Algoritma Klasifikasi Machine Learning untuk Prediksi Penghasilan pada Adult Income Dataset," in *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya*, vol. 2, no. 1, pp. 741–747, 2021.
- [23] T. Triana, E. Utami, and A. D. Hartanto, "Implementasi Algoritma K-Nearest Neighbor pada Aplikasi Deteksi Risiko Tinggi pada Kehamilan," *Infokes: Jurnal Ilmiah Rekam Medis dan Informatika Kesehatan*, vol. 13, no. 2, pp. 64–71, 2023, doi: 10.47701/infokes.v13i2.3456.
- [24] M. R. Hunafa and A. Hermawan, "Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbor pada Imbalance Class Dataset Penyakit Diabetes," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 3, pp. 1551–1561, 2023, doi: 10.30865/klik.v4i3.1628.
- [25] A. M. Argina, "Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes," *Indonesian Journal of Data and Science*, vol. 1, no. 2, pp. 29–33, 2020, doi: 10.33096/ijodas.v1i2.11.