



IMPLEMENTASI METODE *NAÏVE BAYES* DENGAN *RANDOM OVERSAMPLING* PADA KLASIFIKASI KELUARGA BERISIKO STUNTING

Yeni Suliswati¹⁾, Tessy Octavia Mukhti^{2,*}, Syafriandi³⁾, Admi Salma⁴⁾

*1,2,3,4) Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Negeri Padang*

**email: tessyoctaviam@fmipa.unp.ac.id*

Abstrak: Stunting masih menjadi salah satu masalah kesehatan serius yang memiliki dampak jangka panjang terhadap tumbuh kembang dan kognitif anak. Keluarga memiliki peran penting dalam mencegah terjadinya stunting, sehingga identifikasi dini keluarga yang berisiko melahirkan anak stunting menjadi langkah awal dalam upaya pencegahan. Penelitian ini bertujuan untuk mengklasifikasikan keluarga berisiko stunting menggunakan metode *Naïve Bayes* serta mengevaluasi pengaruh teknik *Random Oversampling* (ROS) terhadap performa model pada data tidak seimbang. Data pada penelitian ini terdiri dari 7 variabel independen dan 1 variabel dependen yang bersumber dari Perwakilan Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) Sumatera Barat. Hasil evaluasi menunjukkan bahwa model *Naïve Bayes* memiliki akurasi sebesar 92,46% dan sensitivitas 100% serta spesifisitas 69,14% yang menunjukkan kelemahan dalam mengidentifikasi keluarga berisiko. Metode ROS-*Naïve Bayes* menunjukkan peningkatan performa model dimana diperoleh akurasi sebesar 99,87%, sensitivitas 99,83%, dan spesifisitas 100%. Hal ini menunjukkan bahwa implementasi *Naïve Bayes* dengan ROS efektif dalam mengatasi ketidakseimbangan data dan meningkatkan performa model. Faktor utama yang memengaruhi risiko stunting meliputi keikutsertaan KB modern, sanitasi, usia ibu dan jumlah anak.

Kata Kunci: Stunting; *Naïve Bayes*; *Random Oversampling*; Klasifikasi

Abstract: *Stunting remains one of the most serious health issues with long-term impacts on children's growth and cognitive development. Families play a crucial role in stunting prevention, making the early identification of families at risk of having stunted children a critical first step. This study aims to classify families at risk of stunting using the Naïve Bayes method and to evaluate the impact of the Random Oversampling (ROS) technique on model performance in imbalanced data. The dataset used in this study consists of seven independent variables and one dependent variable, sourced from the West Sumatra Representative Office of the National Population and Family Planning Agency (BKKBN). Evaluation results show that the Naïve Bayes model achieved an accuracy of 92.46%, sensitivity of 100%, and specificity of 69.14%, indicating a weakness in identifying at-risk families. In contrast, the ROS-Naïve Bayes model demonstrated improved performance, achieving 99.87% accuracy, 99.83% sensitivity, and 100% specificity. These findings indicate that implementing Naïve Bayes with ROS is effective in handling data imbalance and enhancing model performance. The main factors influencing the risk of stunting include participation in modern family planning, sanitation, maternal age, and number of children.*

Keywords: *Stunting; Naïve Bayes; Random Oversampling, Classification*



PENDAHULUAN

Stunting merupakan salah satu masalah pada balita yang disebabkan oleh kekurangan gizi kronis. Di Indonesia masih banyak masalah kesehatan yang serius salah satunya stunting. Kondisi stunting dimulai sejak periode sebelum kehamilan atau pada tahap prakonsepsi (Lemaking et al., 2022). Stunting tidak hanya disebabkan oleh kekurangan gizi kronis. Faktor lain penyebab stunting yaitu faktor maternal atau faktor yang berasal dari ibu terutama selama periode 1000 hari pertama kehidupan yang mencakup masa kehamilan hingga anak berusia 2 tahun (Susanti, 2023). Selain faktor maternal, faktor keluarga seperti tingkat pendidikan orang tua dan jenis pekerjaan orang tua juga menjadi dasar pemicu kejadian stunting (Nisa & Azinar, 2024). Orang tua dengan pendidikan yang tinggi akan meningkatkan kesempatan mendapatkan pekerjaan dan pendapatan yang cukup sehingga memperoleh kehidupan yang layak.

Berdasarkan data dari BKKBN, tingkat prevalensi stunting di Indonesia menunjukkan penurunan dari 21,6% pada tahun 2022 menjadi 21,5% di tahun 2023. Meskipun mengalami penurunan, angka prevalensi stunting masih lebih tinggi dibandingkan batas prevalensi stunting suatu negara yang ditetapkan WHO yaitu kurang dari 20%. Angka prevalensi stunting di Sumatera Barat pada tahun 2023 tercatat sebesar 23,6%. Angka tersebut lebih rendah dari tahun 2022 yaitu sebesar 25,2%. Sementara itu, prevalensi stunting di Kota Payakumbuh mengalami peningkatan dari 17,8% di tahun 2022 menjadi 19,8% di tahun 2023.

Dalam rangka penurunan angka stunting, pemerintah berfokus pada upaya pengurangan risiko stunting dalam keluarga melalui berbagai bentuk bantuan dan pendampingan. Beberapa penelitian terdahulu menunjukkan bahwa faktor sosial ekonomi, sanitasi, dan pola asuh berperan besar dalam menentukan risiko stunting pada anak (Apriluana & Fikawati, 2018; Suarnianti, 2020; Yunita et al., 2022). Salah satu cara yang dapat dilakukan dalam upaya pengurangan risiko stunting dalam keluarga yaitu identifikasi dini keluarga yang berisiko melahirkan anak stunting. Untuk itu, diperlukan pemodelan yang mampu mengklasifikasikan keluarga berdasarkan karakteristiknya. Pendekatan yang dapat dilakukan yaitu klasifikasi, dimana setiap keluarga direpresentasikan oleh sekumpulan faktor dan dikategorikan ke dalam kelas tertentu. Klasifikasi dapat membantu dalam mengkategorikan keluarga yang berisiko melahirkan anak stunting. Terdapat berbagai metode klasifikasi seperti, *Naïve Bayes*, Regresi Logistik, *Random Forest*, *Decision Tree*, dan *Support Vector Machine*.

Salah satu metode klasifikasi yang dapat digunakan yaitu metode *Naïve Bayes*. Menurut Afrianti et al. (2020), *Naïve Bayes* merupakan teknik klasifikasi statistik yang memprediksi probabilitas keanggotaan yang akan masuk ke dalam kelas tertentu. Metode *Naïve Bayes* memanfaatkan teorema *Bayes* untuk menghitung probabilitas keanggotaan suatu kelas terhadap fitur-fitur yang diberikan. *Naïve Bayes* merupakan algoritma klasifikasi yang sangat efektif dan efisien, *easy to use*, dan dikenal memiliki akurasi yang tinggi dengan perhitungan yang sederhana (Afrianti et al., 2020; Demir & Şahin, 2022). Metode *Naïve Bayes* memiliki beberapa kelebihan diantaranya dapat menangani data kuantitatif dan diskrit, menangani nilai hilang



dengan mengabaikan atribut tersebut saat menghitung peluang estimasi, dan kuat terhadap fitur yang tidak relevan (Peling et al., 2017).

Salah satu tantangan dalam klasifikasi yaitu data tidak seimbang (Fitriani et al., 2021). Data tidak seimbang terjadi ketika jumlah objek dalam satu kelas lebih besar daripada jumlah objek data lainnya. Data tidak seimbang dapat menyebabkan kinerja klasifikasi menjadi berkurang. Terdapat empat pendekatan untuk mengatasi data tidak seimbang yaitu pendekatan pada level data, pendekatan level algoritma, pendekatan yang peka terhadap biaya, dan pendekatan berbasis ensambel (Fernández et al., 2018). Pada penelitian ini data tidak seimbang diatasi dengan pendekatan level data menggunakan metode *Random Oversampling* (ROS).

Penelitian yang dilakukan oleh Fitriani et al. (2021) menunjukkan peningkatan kinerja model setelah dikombinasikan dengan *Random Oversampling*. Dalam penelitian tersebut, ROS dikombinasikan dengan *Naïve Bayes* untuk memprediksi keberhasilan KB IUD di Kabupaten Kendal. Hasilnya, menunjukkan kombinasi ROS-*Naïve Bayes* efektif dalam menangani data tidak seimbang. Penelitian lainnya yang dilakukan oleh Pang et al. (2019) menunjukkan bahwa penggunaan ROS dapat meningkatkan performa klasifikasi pada data tidak seimbang. Dalam penelitian tersebut, ROS digunakan untuk menambah jumlah sampel pada kelas minoritas dalam deteksi malware berbasis lalu lintas jaringan. Hasilnya metode ROS efektif dalam memperbaiki distribusi data dan meningkatkan kemampuan model dalam mengenali kasus pada kelas minoritas.

Permasalahan utama dalam penelitian ini adalah bagaimana menerapkan konsep probabilitas bersyarat pada algoritma *Naïve Bayes* dalam mengklasifikasikan keluarga berisiko stunting, serta bagaimana pengaruh teknik *Random Oversampling* terhadap distribusi data dan performa klasifikasi pada kondisi data tidak seimbang. Penelitian ini didasarkan pada pemahaman bahwa penerapan teknik *Random Oversampling* dapat menangani ketidakseimbangan kelas pada data dan meningkatkan performa model klasifikasi.

Tujuan penelitian ini ialah untuk mengevaluasi penerapan metode *Naïve Bayes* dengan teknik *Random Oversampling* (ROS) dalam mengklasifikasikan keluarga berisiko stunting di Kota Payakumbuh. Selain itu, penelitian ini juga bertujuan untuk mengidentifikasi faktor-faktor yang memengaruhi kemungkinan risiko stunting dalam keluarga. Melalui penelitian ini, pembaca diharapkan dapat memahami penerapan konsep dasar probabilitas, pemodelan klasifikasi dengan data tidak seimbang, serta pendekatan praktis dalam mengoptimalkan akurasi prediksi melalui teknik penyeimbangan data.

METODE PENELITIAN

Penelitian ini merupakan penelitian terapan. Penelitian terapan adalah ialah penelitian yang diawali dengan analisis teori dan diikuti dengan pengambilan data. Penelitian terapan bertujuan untuk memberikan solusi konkret atau mengatasi masalah dalam kehidupan sehari-hari. Penelitian ini menggunakan data sekunder yang diperoleh dari Perwakilan Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) Sumatera Barat. Data pada



penelitian ini terdiri dari 19832 objek dengan 1 variabel dependen (Y) dan 7 variabel independen (X). Penjelasan masing-masing variabel dapat dilihat pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Nama	Keterangan	Kategori
X_1	Air Minum	Keluarga sasaran tidak memiliki sumber air minum yang layak	0 = Tidak 1 = Ya
X_2	Jamban	Keluarga sasaran tidak memiliki jamban yang layak	0 = Tidak 1 = Ya
X_3	PUS Terlalu Muda	Pasangan Usia Subur yang berusia kurang dari 21 tahun	0 = Tidak 1 = Ya
X_4	PUS Terlalu Tua	Pasangan Usia Subur yang berusia lebih dari 35 tahun	0 = Tidak 1 = Ya
X_5	PUS Terlalu Dekat	Pasangan Usia Subur yang masa kehamilan pertama dengan berikutnya kurang dari 2 tahun	0 = Tidak 1 = Ya
X_6	PUS Terlalu Banyak	Pasangan Usia Subur yang memiliki anak lebih dari 2	0 = Tidak 1 = Ya
X_7	KB modern	Bukan peserta KB modern	0 = Tidak 1 = Ya
Y	KRS	Keluarga Berisiko Stunting	0 = Tidak 1 = Ya

Klasifikasi

Klasifikasi merupakan salah satu metode dalam *machine learning* yang dapat mengkategorikan objek ke dalam kelas berdasarkan karakteristik atau ciri tertentu. Salah satu manfaat model klasifikasi yaitu sebagai alat penjelas (*explanatory tool*) untuk membedakan objek dari kelas yang berbeda (Tan et al., 2014). Contohnya, model dapat menjelaskan ciri-ciri keluarga yang berisiko melahirkan anak stunting atau keluarga yang tidak berisiko.

Menurut Gorunescu (2011), klasifikasi didasarkan oleh 4 komponen utama yaitu: *class* atau variabel dependen, *predictors* atau variabel independen, *training dataset* atau data pelatihan, dan *testing dataset* atau data uji. Data uji pada klasifikasi dimanfaatkan untuk membentuk model. Model yang dihasilkan kemudian digunakan untuk memprediksi kelas dari data uji. Hasil prediksi dari data uji digunakan sebagai evaluasi kinerja model klasifikasi berdasarkan jumlah catatan uji yang diprediksi benar atau salah. Menurut Id (2021), tidak ada aturan pasti tentang pembagian data *training* dan data *testing*. Pada penelitian ini pembagian data latih dan data uji yang digunakan adalah 80%:20%.

Ketidakseimbangan Data

Ketidakseimbangan data (*imbalance dataset*) adalah perbedaan jumlah kelas yang signifikan antara kelas minoritas dengan kelas mayoritas (Sir & Soepranoto, 2022). Ketidakseimbangan data dapat menyebabkan model klasifikasi cenderung memilih kelas mayoritas pada saat klasifikasi. Secara umum, suatu dataset dikatakan tidak seimbang apabila



distribusi jumlah data pada masing-masing kelas tidak merata. Ketidakseimbangan ini menjadi masalah ketidakseimbangan data apabila terdapat perbedaan rasio yang signifikan antara satu kelas dengan kelas lainnya (Fernández et al., 2018).

Burnaev et al. (2015) memperkenalkan *Imbalance Ratio* (IR) untuk mengukur derajat ketidakseimbangan data suatu dataset.

$$IR = \frac{|C_0|}{|C_1|} \quad (1)$$

dimana C_0 merupakan jumlah dataset dengan label mayoritas pada dataset dan C_1 merupakan jumlah dataset dengan label minoritas. Nilai $IR \geq 1$ dan semakin besar nilainya menunjukkan semakin tidak seimbang dataset.

Naïve Bayes

Metode *Naïve Bayes* diperkenalkan oleh Thomas Bayes, seorang ahli matematika dan statistik asal Inggris yang memiliki keahlian di bidang statistika dan teori probabilitas. *Naïve Bayes* merupakan metode klasifikasi yang berakar pada teorema Bayes dengan asumsi bahwa pengaruh nilai atribut pada kelas tertentu tidak bergantung pada nilai atribut lainnya (Peling et al., 2017). Asumsi independensi dibuat untuk menyederhanakan komputasi yang terlibat. Asumsi independensi ini jarang terpenuhi dalam aplikasi dunia nyata, sehingga algoritma ini disebut “*Naïve*” (naif).

Jika diketahui sebuah dimensi vektor $X = (x_1, x_2, \dots, x_n)$ dan terdapat m kelas C_1, C_2, \dots, C_m , maka teorema *Bayes* dirumuskan sebagai berikut (Han et al., 2012)

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$

Keterangan:

- $P(C_i|X)$: *posterior* (probabilitas C_i berdasarkan kondisi X)
- $P(X|C_i)$: *likelihood* (probabilitas X berdasarkan kondisi C_i)
- $P(C_i)$: *prior* (probabilitas kelas C_i)
- $P(X)$: *evidence* (probabilitas total X)

Probabilitas prior kelas dapat diestimasi dengan $P(C_i) = C_{i,D}/D$. Dimana $C_{i,D}$ merupakan jumlah data latih dari kategori C_i dan D adalah jumlah total dari data latih.

Asumsi naif dari independensi bersyarat dibuat untuk menyederhanakan perhitungan atau mengurangi komputasi dalam mengevaluasi $P(X|C_i)$. Klasifikasi yang berkembang dengan asumsi ini dikenal sebagai klasifikasi *Naïve Bayes*. Dengan demikian,

$$\begin{aligned} P(X|C_i) &= \prod_{k=1}^n P(x_k|C_i) \\ &= P(x_1|C_i) \times P(x_2|C_i) \times \dots \times P(x_n|C_i) \end{aligned}$$

Probabilitas $P(X|C_i)$ dapat diestimasi dari data latih. Jika atribut adalah kategorik maka dapat dihitung dengan



$$P(X|C_i) = \frac{|C_{i,D,k}|}{|C_{i,D}|}$$

dimana $C_{i,D,k}$ adalah jumlah data pelatihan kelas C_i pada data yang memiliki nilai x_k . Sedangkan $C_{i,D}$ adalah jumlah data training dengan kelas C_i . Jika atribut bernilai kontinu maka atribut bernilai kontinu diasumsikan memiliki distribusi *Gaussian* dengan rata-rata μ dan standar deviasi σ , didefinisikan sebagai berikut.

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\mu\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

maka dari itu $P(x_k|C_i) = g(x_k, \mu C_i, \sigma C_i)$.

Random Oversampling

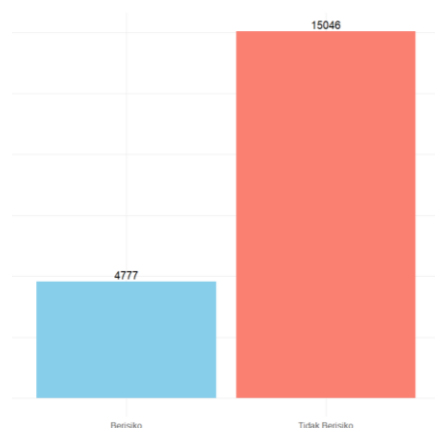
Random Oversampling (ROS) merupakan teknik penanganan data yang sederhana. Teknik *Random Oversampling* mengatasi ketidakseimbangan data dengan cara memilih sampel secara acak dari kelas minoritas dan menduplikasi berulang kali sampai data menjadi seimbang. Menurut Burnaev et al. (2015), *Random Oversampling* juga disebut dengan *bootstrap oversampling*, dimana ukuran kelas minoritas dikembangkan dengan metode *bootstrap*.

Mekanisme *Random Oversampling* secara alami mengikuti prinsip dasarnya, yaitu dengan menambahkan himpunan E yang terdiri dari sampel kelas minoritas. Contoh dari kelas minoritas dipilih secara acak S_{min} , kemudian contoh tersebut diduplikasi dan ditambahkan ke dalam himpunan data S . Cara ini menghasilkan jumlah total contoh dalam S_{min} meningkat sebesar $|E|$, sehingga jumlah data dalam kelas minoritas bertambah dan distribusi kelas dalam dataset menjadi lebih seimbang (He & Garcia, 2009).

HASIL DAN PEMBAHASAN

Rasio Ketidakseimbangan

Ketidakseimbangan suatu kelas merupakan kondisi dimana jumlah objek pada suatu kelas lebih banyak dibandingkan kelas lainnya. Distribusi frekuensi risiko stunting dapat dilihat pada Gambar 1.





Gambar 1. Distribusi Risiko Stunting

Gambar 1. menunjukkan distribusi risiko stunting antara keluarga tidak berisiko dengan keluarga berisiko. Berdasarkan data keluarga berisiko stunting di Kota Payakumbuh tahun 2023 terlihat bahwa terdapat 15046 keluarga yang tidak berisiko melahirkan anak stunting dan terdapat 4777 keluarga yang berisiko melahirkan anak stunting. Berdasarkan Persamaan (1) dapat dihitung IR dari data yaitu sebesar 3,15 yang menunjukkan adanya ketidakseimbangan kelas data.

Penanganan Ketidakseimbangan Data dengan *Random Oversampling*

Metode *Random Oversampling* (ROS) digunakan untuk meningkatkan jumlah kelas secara seimbang. Penanganan ketidakseimbangan kelas data dilakukan pada data latih. Rasio dataset sebelum dan sesudah penanganan dengan ROS dapat dilihat pada Tabel 2.

Tabel 2. Rasio dataset dengan nilai IR

Metode	Risiko Stunting		IR
	Ya	Tidak	
Tanpa ROS	3808	12050	3,1644
ROS	12050	12050	1

Pada Tabel 2. data latih tanpa ROS terdiri dari 15858 kejadian risiko stunting, dengan 3808 termasuk kategori berisiko dan 12050 termasuk kategori tidak berisiko. Nilai IR sebesar 3,1644 menunjukkan bahwa rasio ukuran sampel data latih tidak seimbang. Sementara itu setelah dilakukan penanganan ketidakseimbangan data dengan ROS, diperoleh nilai IR sebesar 1 yang menunjukkan rasio ukuran data latih sudah seimbang. Dapat dikatakan bahwa metode ROS mampu mengatasi ketidakseimbangan data.

Hasil Klasifikasi *Naïve Bayes* dan ROS-*Naïve Bayes*

Model klasifikasi yang dibangun pada penelitian ini yaitu metode *Naïve Bayes*, sedangkan metode penanganan ketidakseimbangan data yang digunakan yaitu metode ROS. Model yang diperoleh dari pembelajaran menggunakan data latih kemudian di uji pada data uji. Hasil pengujian menggunakan metode *Naïve Bayes* dapat dilihat pada Tabel 3.

Tabel 3. Confussion Matrix *Naïve Bayes*

Prediksi	Nilai Aktual	
	Tidak Berisiko	Berisiko
Tidak Berisiko	2996	299
Berisiko	0	670

Berdasarkan Tabel 3. Model *Naïve Bayes* mampu mengklasifikasikan keluarga tidak berisiko stunting dengan sangat baik. Hal ini terbukti dari 2996 keluarga yang tidak berisiko diprediksi dengan tepat. Selain itu, model juga cukup baik dalam mengidentifikasi keluarga yang



berisiko, di mana sebanyak 670 keluarga yang berisiko berhasil diprediksi dengan tepat. Namun, masih terdapat 299 keluarga yang berisiko tetapi diprediksi tidak berisiko. Tabel 3. menunjukkan bahwa model lebih akurat dalam mengenali kelas mayoritas yaitu keluarga tidak berisiko. Hal ini dapat terjadi karena adanya ketidakseimbangan kelas pada data. Selanjutnya hasil pengujian metode *Naïve Bayes* setelah dilakukan penanganan ketidakseimbangan data menggunakan metode ROS dapat dilihat pada Tabel 4.

Tabel 4. *Confussion Matrix* ROS-*Naïve Bayes*

Prediksi	Nilai Aktual	
	Tidak Berisiko	Berisiko
Tidak Berisiko	2991	0
Berisiko	5	969

Tabel 4. menunjukkan bahwa model ROS-*Naïve Bayes* memiliki kinerja yang sangat baik dalam memprediksi keluarga berisiko maupun keluarga yang tidak berisiko. Model mampu memprediksi 2991 keluarga yang tidak berisiko secara tepat dan hanya 5 keluarga tidak berisiko diprediksi sebagai berisiko. Sementara itu, 969 keluarga berisiko diprediksi dengan tepat. Hal ini menunjukkan bahwa model lebih sensitif terhadap kelas minoritas yaitu keluarga berisiko dikarenakan penggunaan *Random Oversampling*. Berdasarkan Tabel 3. dan Tabel 4, dapat dihitung nilai akurasi, spesifisitas, dan sensitivitas yang disajikan pada Tabel 5.

Tabel 5. Perbandingan Hasil Klasifikasi

Model	Akurasi	Sensitivitas	Spesifisitas
<i>Naïve Bayes</i>	92,46%	100%	69,14%
ROS- <i>Naïve Bayes</i>	99,87%	99,83%	100%

Tabel 5. Menunjukkan perbandingan performa model *Naïve Bayes* dan ROS-*Naïve Bayes* berdasarkan tiga metrik utama, yaitu akurasi, sensitivitas, dan spesifisitas. Model *Naïve Bayes* menunjukkan akurasi sebesar 92,46% dengan sensitivitas 100% dan spesifisitas 69,14%. Sensitivitas yang tinggi menunjukkan bahwa model *Naïve Bayes* mampu memprediksi keluarga tidak berisiko dengan sangat baik, namun nilai spesifisitas yang relatif rendah menunjukkan model lemah dalam memprediksi keluarga berisiko.

Model ROS-*Naïve Bayes* mengalami peningkatan akurasi mencapai 99,87% dan sensitivitas mencapai 99,83% serta spesifisitas menjadi 100%. Hal ini menunjukkan penggunaan *Random Oversampling* dalam menyeimbangkan distribusi kelas dan dapat meningkatkan performa model. Peningkatan kinerja model ROS-*Naïve Bayes* mampu membuat model belajar representasi yang lebih seimbang antara kelas mayoritas dengan kelas minoritas yaitu keluarga tidak berisiko dan keluarga berisiko.

Selanjutnya, berdasarkan model menggunakan ROS-*Naïve Bayes* akan dilihat faktor-faktor yang berpengaruh terhadap keluarga berisiko stunting. Pengaruh faktor dapat dilihat dengan membandingkan nilai probabilitas bersyarat $P(X|Y)$. Faktor memiliki pengaruh yang



kuat jika nilai $P(X = 1|Y = 1)$ jauh lebih tinggi dibandingkan $P(X = 1|Y = 0)$. Perbandingan probabilitas bersyarat dari faktor-faktor risiko terhadap risiko stunting disajikan pada Tabel 6.

Tabel 6. Perbandingan Probabilitas Bersyarat Faktor-Faktor Terhadap Risiko Stunting

Nama	$P(X=1 Y=1)$	$P(X=1 Y=0)$	Selisih
Air Minum	0,0203	0	0,0203
Jamban	0,2825	0	0,2825
PUS Terlalu Muda	0,0047	0,0024	0,0023
PUS Terlalu Tua	0,4666	0,1889	0,2777
PUS Terlalu Dekat	0,0196	0,0142	0,0054
PUS Terlalu Banyak	0,5349	0,3035	0,2314
KB Modern	0,8254	0,2945	0,5309

Berdasarkan Tabel 6. dapat diketahui bahwa faktor bukan peserta KB modern merupakan indikator yang paling kuat dalam membedakan keluarga berisiko dan tidak berisiko, dengan selisih probabilitas bersyarat sebesar 0,5309. Faktor lain yang juga memiliki pengaruh besar yaitu akses jamban tidak layak dengan selisih 0,2825 dan usia ibu terlalu tua dengan selisih 0,2777. Selain itu, jumlah anak terlalu banyak juga menunjukkan pengaruh cukup besar dengan selisih probabilitas bersyarat sebesar 0,2314. Sementara itu, faktor seperti akses air minum tidak layak, jarak kelahiran terlalu dekat, dan usia ibu terlalu muda menunjukkan selisih yang kecil. Dalam penelitian ini, ketiga faktor tersebut menunjukkan pengaruh yang kecil terhadap klasifikasi keluarga berisiko stunting.

SIMPULAN

Hasil penelitian menunjukkan bahwa penerapan *Random Oversampling* (ROS) pada metode *Naïve Bayes* mampu meningkatkan performa model dalam mendeteksi keluarga berisiko stunting. Model *Naïve Bayes* memiliki sensitivitas yang tinggi namun spesifisitas rendah, sehingga kurang optimal dalam membedakan keluarga berisiko. Sebaliknya, model ROS-*Naïve Bayes* mampu meningkatkan akurasi serta menghasilkan sensitivitas dan spesifisitas yang tinggi dan seimbang. Hal ini menunjukkan bahwa penerapan *Random Oversampling* dalam menyeimbangkan distribusi kelas pada data latih dapat memperbaiki kemampuan generalisasi model terhadap kelas. Selain itu, faktor yang berpengaruh signifikan terhadap keluarga berisiko stunting yaitu KB Modern (X_7). Faktor lain yang juga berpengaruh signifikan yaitu Jamban (X_2), Terlalu Tua (X_4), dan Terlalu Banyak (X_6). Sementara itu, faktor seperti Air Minum (X_1), Terlalu Dekat (X_5), dan Terlalu Muda (X_3) menunjukkan pengaruh yang relatif kecil terhadap klasifikasi.



UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Perwakilan Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) Sumatera Barat atas ketersediaan dan akses terhadap data yang digunakan dalam penelitian ini. Data tersebut berperan penting dalam mendukung analisis dan pencapaian hasil penelitian ini.

DAFTAR PUSTAKA

- Afrianti, E., Fthoni, & Heroza, R. I. (2020). Text classification with Naïve Bayes Classifier (NBC) for grouping report description and recovery time duration of PT.PLN (persero) WS2JB Palembang area. *E-Journal*, 12(1), 1955–1961.
- Apriluana, G., & Fikawati, S. (2018). Analisis faktor-faktor risiko terhadap kejadian stunting pada balita. *Jurnal Departemen Gizi Fakultas Kesehatan Masyarakat*, 28(4), 247–256. <https://doi.org/10.22435/mpk.v28i4.472>.
- Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN). (n.d.). *Satuan tugas percepatan penurunan stunting*. Diakses 22 Mei 2025, dari <https://satgasstunting.info/web/index.php>
- Burnaev, E., Erofeev, P., & Papanov, A. (2015). Influence of resampling on accuracy of imbalanced classification. *Eighth International Conference on Machine Vision (ICMV 2015)*, 9875(987521), 1–5. <https://doi.org/10.1117/12.2228523>.
- Demir, S., & Şahin, E. K. (2022). Evaluation of oversampling methods (OVER, SMOTE, and ROSE) in classifying soil liquefaction dataset based on SVM, RF, and Naïve Bayes. *European Journal of Science and Technology*, 34, 142–147. <https://doi.org/10.31590/ejosat.1077867>.
- Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). *Learning from imbalanced data sets*. Switzerland: Springer Nature.
- Fitriani, R. D., Yasin, H., & Tarno, T. (2021). Penanganan klasifikasi kelas data tidak seimbang dengan Random Oversampling Pada Naive Bayes (Studi Kasus: Status Peserta KB IUD di Kabupaten Kendal). *Jurnal Gaussian*, 10(1), 11–20. <https://doi.org/10.14710/j.gauss.v10i1.30243>.
- Gorunescu, F. (2011). Data mining concepts, models and technique. In *Springer* (Vol. 12). Berlin, Germany: Springer-Verlag.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: concepts and techniques* (Third). USA: Elsevier Inc.
- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284. <https://doi.org/10.1109/TKDE.2008.239>.
- Id, I. D. (2021). *Machine Learning: teori, sudi kasus dan implementasi menggunakan python* (1st ed.). Riau: UR PRESS.
- Lemaking, V. B., Manimalai, M., & Djogo, H. M. A. (2022). Hubungan pekerjaan ayah, pendidikan ibu, pola asuh, dan jumlah anggota keluarga dengan kejadian stunting pada balita di Kecamatan Kupang Tengah, Kabupaten Kupang. *Ilmu Gizi Indonesia*, 5(2), 123–132. <https://doi.org/10.35842/ilgi.v5i2.254>.
- Nisa, W. K., & Azinar, M. (2024). Karakteristik keluarga berisiko stunting pada anak usia 7-24 bulan di wilayah kerja Puskesmas Bandarharjo dengan pendekatan case control. *Kritis*,



- 33(1), 17–36. <https://doi.org/10.24246/kritis.v33i1p17-36>.
- Pang, Y., Chen, Z., Peng, L., Ma, K., Zhao, C., & Ji, K. (2019). A signature-based assistant random oversampling method for malware detection. *Proceedings 2019 18th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/13th IEEE International Conference on Big Data Science and Engineering, TrustCom/BigDataSE 2019*, 256–263. <https://doi.org/10.1109/TrustCom/BigDataSE.2019.00042>
- Peling, I. B. A., Arnawan, I. N., Arthawan, I. P. A., & Janardana, I. G. N. (2017). Implementation of data mining to predict period of students study using naive bayes algorithm. *International Journal of Engineering and Emerging Technology*, 2(1), 53–57. <https://doi.org/10.18201/ijisae.2019252786>.
- Sir, Y. A., & Soepranoto, A. H. H. (2022). Pendekatan resampling data untuk menangani masalah ketidakseimbangan kelas. *Jurnal Komputer Dan Informatika*, 10(1), 31–38. <https://doi.org/10.35508/jicon.v10i1.6554>.
- Suarnianti. (2020). Faktor risiko stunting : Literatur Review. *Jurnal Ilmiah Kesehatan*, 15(2), 144–147.
- Susanti, R. (2023). Analisis faktor maternal terhadap keluarga berisiko stunting sebagai upaya peningkatan analisis data di BKKBN Kalimantan Timur. *Jurnal Nasional Pengabdian Masyarakat*, 4(1), 7–17. <https://doi.org/10.47747/jnpm.v4i1.1089>.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2014). *Introduction to data mining*. USA: Pearson.
- Yunita, A., Asra, R. H., Nopitasari, W., Putri, R. H., & Fevria, R. (2022). Hubungan sosial ekonomi dengan kejadian stunting pada balita. *Semnas Bio 2022*, 812–819. <https://doi.org/10.24036/prosemnasbio/vol2/519>.