

# Application of Naïve Bayes Classifiers for Family Risk Identification and Stunting Intervention Planning

Wildan Indra Kurniawan<sup>1\*</sup>, Joko Triloka<sup>2\*</sup>

<sup>\*</sup>Fakultas Ilmu Komputer, Institut Informatika Dan Bisnis Darmajaya  
[wildan.2421211003p@mail.darmajaya.ac.id](mailto:wildan.2421211003p@mail.darmajaya.ac.id)<sup>1</sup>, [joko.triloka@darmajaya.ac.id](mailto:joko.triloka@darmajaya.ac.id)<sup>2</sup>

## Article Info

### Article history:

Received 2025-08-09

Revised 2025-09-03

Accepted 2025-09-20

### Keyword:

*Data mining,  
Early detection,  
Algorithm performance,  
Stunting prevention.*

## ABSTRACT

Stunting remains a significant public health concern influenced by a combination of social, economic, and environmental factors. This study aims to implement the Naïve Bayes algorithm to support the determination of appropriate intervention strategies for families identified as being at risk of stunting in Metro City. Risk data were obtained from the BKKBN Metro City and underwent preprocessing steps, including handling missing values, encoding categorical variables, and feature selection. The dataset was then divided into training, validation, and testing subsets to develop and evaluate models using three Naïve Bayes variants: Gaussian, Multinomial, and Bernoulli. Evaluation metrics of accuracy, precision, recall, and F1-score indicate that the Multinomial Naïve Bayes model achieved the best performance with 99% accuracy, followed by the Bernoulli Naïve Bayes model with 98% accuracy. Both models effectively classified families at risk of stunting with minimal misclassification, while the Gaussian Naïve Bayes variant demonstrated lower performance with an accuracy of 60%. These results highlight the potential of the Naïve Bayes algorithm, particularly the Multinomial and Bernoulli models, as practical and efficient tools to support data-driven decision-making for stunting interventions.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

## I. PENDAHULUAN

Stunting merupakan salah satu permasalahan serius dalam bidang kesehatan masyarakat, khususnya di Indonesia. Stunting adalah kondisi gagal tumbuh pada anak balita akibat kekurangan gizi kronis yang berlangsung dalam waktu lama, terutama pada masa 1.000 hari pertama kehidupan. Anak yang mengalami stunting umumnya memiliki tinggi badan yang lebih rendah dibandingkan anak seusianya, serta rentan terhadap penyakit dan gangguan perkembangan kognitif. Berbagai faktor turut memengaruhi risiko stunting dalam keluarga, antara lain kondisi sanitasi yang buruk, pola hidup yang tidak sehat, adanya perokok dalam keluarga, serta minimnya akses terhadap fasilitas dasar seperti air bersih dan jamban yang layak [1], [2].

Dalam konteks penanganan stunting, pemerintah Indonesia melalui Dinas Pemberdayaan Perempuan, Perlindungan Anak, Pengendalian Penduduk, dan Keluarga Berencana (PPAPKB) telah melakukan intervensi terhadap keluarga

yang masuk dalam kategori berisiko stunting. Intervensi ini dikelompokkan ke dalam dua jenis, yaitu intervensi spesifik dan intervensi sensitif, berdasarkan faktor-faktor risiko seperti kebiasaan buang air besar sembarangan, kepemilikan jamban tidak layak, status bantuan sosial, dan kondisi anemia pada anggota keluarga perempuan [3].

Berbagai penelitian sebelumnya telah mengeksplorasi pemanfaatan algoritma data mining dalam klasifikasi dan prediksi data, seperti penelitian oleh [4] meningkatkan kinerja metode klasifikasi machine learning, yaitu K-Nearest Neighbor (KNN) dan Decision Tree (C4.5) dengan fitur particle swarm optimization (PSO) pada kasus penyakit jantung. Temuan mereka menunjukkan bahwa algoritma K-NN berkinerja sangat baik dengan PSO, mencapai akurasi 89,09%, presisi 89,61%, recall 90,79%, dan nilai AUC 0,935. Penelitian lainnya menggunakan model ANN dan CNN diusulkan untuk memprediksi apakah pelanggan di industri ritel akan berubah di masa depan [5]. Hasil penelitian mereka menunjukkan bahwa model prediksi churn berbasis deep

learning yang diusulkan memiliki kinerja klasifikasi yang lebih baik. Model CNN menghasilkan 97.62% tingkat akurasi yang menghasilkan klasifikasi dan keberhasilan prediksi yang lebih baik dibandingkan model lain yang dibandingkan.

Meskipun berbagai algoritma seperti KNN, Decision Tree, dan CNN telah banyak digunakan untuk klasifikasi dan prediksi, metode Naïve Bayes dipilih dalam penelitian ini karena keunggulannya dalam menangani dataset berdimensi besar dan bersifat independen, serta kemampuannya dalam mengelola data kategorikal dengan efisien [6] [7]. Keunggulan lain dari Naïve Bayes adalah kecepatan komputasi yang tinggi, serta kemudahan dalam implementasi, yang membuatnya sangat cocok untuk masalah dengan data yang terbatas atau tidak sempurna, seperti data keluarga berisiko stunting. Naïve Bayes juga lebih mudah diinterpretasikan dibandingkan algoritma lain, sehingga dapat mendukung pengambilan keputusan berbasis data yang lebih transparan dan dapat dijelaskan kepada pihak terkait dalam kebijakan publik [8].

Algoritma Naïve Bayes merupakan metode klasifikasi probabilistik yang berdasarkan pada Teorema Bayes dengan asumsi independensi antar fitur, yang terkenal karena kesederhanaan, efisiensi komputasi, dan kemampuannya menghasilkan klasifikasi yang memadai. Keunggulannya dalam menangani dataset berdimensi besar dan fitur kategorikal menjadikannya populer di berbagai bidang, termasuk pengolahan data keluarga berisiko stunting [6], [7]. Beberapa varian utama dari Naïve Bayes, yaitu Bernoulli dan Multinomial, dipilih berdasarkan karakteristik data; Bernoulli Naïve Bayes efektif untuk data dengan fitur biner, sedangkan Multinomial lebih cocok untuk data dengan distribusi frekuensi fitur yang kompleks, seperti yang sering ditemukan dalam data stunting yang melibatkan berbagai kategori sosial-ekonomi dan kesehatan [9], [10]. Selain itu, Naïve Bayes menawarkan kecepatan komputasi tinggi dan kemudahan implementasi, menjadikannya ideal untuk menangani data terbatas atau tidak sempurna, seperti yang sering dijumpai pada data stunting. Dengan kemampuan ini, Naïve Bayes mendukung pengambilan keputusan berbasis data yang lebih transparan dan efisien, serta memudahkan implementasi dalam sistem pendukung keputusan berbasis data mining untuk intervensi stunting. Meskipun penggunaannya dalam penanganan stunting masih terbatas, metode ini memiliki potensi besar untuk membantu Dinas PPPAPKB dalam merencanakan intervensi yang lebih tepat sasaran, berdasarkan karakteristik keluarga yang teridentifikasi berisiko.

Meskipun Naïve Bayes telah banyak diterapkan dalam berbagai aplikasi, seperti deteksi serangan Distributed Denial of Service (DDoS) dan intrusi jaringan [11], [12], [13], [14], penerapannya dalam stunting masih terbatas. Penelitian ini bertujuan untuk mengisi gap tersebut dengan menggunakan Naïve Bayes untuk klasifikasi risiko stunting, yang memungkinkan perencanaan intervensi lebih akurat dan tepat sasaran bagi keluarga berisiko. Selain itu, peneliti mengaplikasikan teknik Recursive Feature Elimination

(RFE), yang terbukti efektif dalam meningkatkan akurasi model dengan mengoptimalkan jumlah fitur relevan, sehingga memperbaiki kinerja klasifikasi dan mendukung pengambilan keputusan yang lebih tepat.

Secara praktis, output model Naïve Bayes yang mengklasifikasikan keluarga berisiko stunting dapat diintegrasikan ke dalam sistem pengambilan keputusan yang digunakan oleh pemerintah daerah terkait. Hasil klasifikasi ini dapat digunakan untuk menentukan prioritas bantuan dan jenis intervensi yang paling sesuai, mempercepat respons terhadap masalah kesehatan masyarakat, serta memaksimalkan sumber daya yang tersedia. Integrasi ini menegaskan bagaimana penggunaan data mining berbasis Naïve Bayes dapat memberikan solusi berbasis data yang efisien dan efektif dalam penanganan stunting di tingkat kebijakan publik.

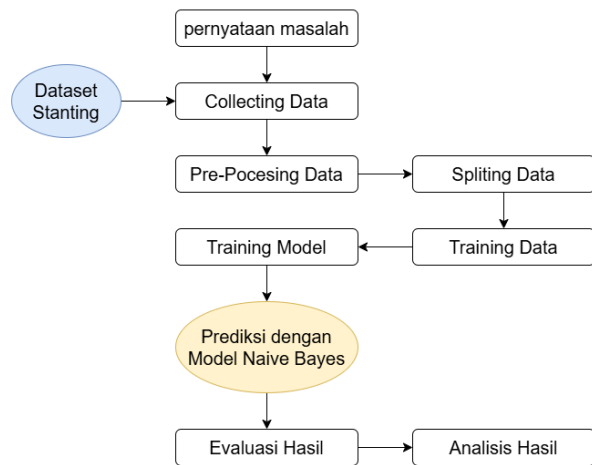
Penelitian ini bertujuan memberikan kontribusi baru dengan menerapkan Naïve Bayes untuk menentukan intervensi keluarga berisiko stunting secara data-driven, yang dapat mendukung kebijakan kesehatan masyarakat yang lebih tepat sasaran. Kebaruan penelitian ini terletak pada penerapan metode Naïve Bayes dalam konteks penanganan stunting berbasis data keluarga, yang masih jarang dijadikan fokus dalam penelitian sebelumnya. Selain itu, penelitian ini berkontribusi dalam menyediakan solusi berbasis data mining yang dapat diintegrasikan langsung ke dalam sistem pengambilan keputusan di tingkat pemerintahan, khususnya di Dinas PPPAPKB. Dengan demikian, hasil penelitian ini diharapkan dapat meningkatkan efisiensi dan efektivitas intervensi kesehatan melalui pemilihan dan penyesuaian intervensi yang lebih terarah berdasarkan karakteristik keluarga yang berisiko.

## II. METODE

Penelitian menggunakan algoritma Naïve Bayes sebagai metode klasifikasi utama untuk menentukan intervensi bagi keluarga berisiko stunting. Prosedur penelitian disajikan pada Gambar 1, yang memperlihatkan alur mulai dari pengumpulan data, pra-pemrosesan, pelatihan model, prediksi, hingga evaluasi hasil. Setiap tahap dilaksanakan secara sistematis untuk menghasilkan model yang valid dan dapat diandalkan dalam mendukung pengambilan keputusan intervensi berbasis data.

### A. Collecting Data

Data yang digunakan dalam penelitian ini diperoleh dari instansi resmi, yaitu BKKBN Kota Metro, Provinsi Lampung yang menyediakan data terkait keluarga berisiko stunting. Dataset tersebut mencakup berbagai variabel penting yang berkaitan dengan kondisi sosial-ekonomi, kesehatan, dan lingkungan keluarga. Data ini kemudian digunakan sebagai dasar untuk analisis menggunakan algoritma Naïve Bayes.



Gambar 1. Tahapan penelitian

### B. Pre-Processing Data

Tahap pra-pemrosesan data dilakukan untuk membersihkan dan menyiapkan data agar layak digunakan dalam proses analisis dan pemodelan. Prosesnya meliputi identifikasi dan penanganan data yang hilang (*missing values*), penghapusan duplikasi, normalisasi data serta transformasi data seperti encoding variabel kategorikal menjadi bentuk numerik agar dapat diproses oleh algoritma yang digunakan. Pra-pemrosesan juga mencakup seleksi fitur untuk memilih atribut yang relevan dan mengurangi dimensi data sehingga meningkatkan kualitas dan efisiensi model.

### C. Splitting Data

Setelah data siap, langkah selanjutnya adalah membagi dataset menjadi beberapa bagian untuk keperluan pelatihan dan pengujian model. Pembagian bertujuan agar model dapat dilatih pada sebagian data (training set) dan diuji performanya pada data yang belum pernah dilihat sebelumnya (testing set). Umumnya, data juga dapat dipisahkan menjadi subset validasi (validation set) untuk menyetel parameter model dan mencegah overfitting. Proporsi pembagian data disesuaikan dengan kebutuhan dan karakteristik dataset

### D. Prediksi dengan Model Naïve Bayes

Pada tahap prediksi, model klasifikasi dilatih menggunakan data pelatihan yang telah disiapkan sebelumnya. Proses pelatihan bertujuan untuk membangun model yang mampu mengenali pola dan hubungan antar fitur dalam data sehingga dapat mengklasifikasikan data dengan baik. Setelah model selesai dilatih, dilakukan tahap prediksi menggunakan data pengujian untuk mengevaluasi kemampuan model dalam mengklasifikasikan data yang belum pernah dilihat sebelumnya. Hasil prediksi berupa label kelas yang dihasilkan oleh model berdasarkan probabilitas yang dihitung selama pelatihan.

### E. Evaluasi Hasil

Evaluasi dilakukan dengan membandingkan hasil prediksi model dengan data aktual menggunakan confusion matrix. Dari matriks ini dihitung metrik-metrik evaluasi yaitu akurasi, presisi, recall, dan F1-score yang memberikan gambaran komprehensif tentang performa masing-masing model Naïve Bayes.

## III. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil analisis dan pembahasan mengenai penerapan algoritma Naïve Bayes untuk menentukan intervensi bagi keluarga berisiko stunting di Kota Metro. Data yang digunakan telah melalui tahap pra-pemrosesan dan pemilihan fitur. Model dikembangkan dengan tiga varian Naïve Bayes, kemudian dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score.

Dataset yang digunakan merupakan data stunting Kota Metro, Provinsi Lampung, yang diperoleh dari instansi BKKBN Kota Metro. Gambar 2 menunjukkan contoh dataset yang digunakan dalam penelitian ini.

	NDI	Name	baduta	balita	PUS	PUS_hamil	DP53	DP52	DP54	...	FLTS1	FLTS2	FLTS3	Ibu_SLTP	post_3	post_2	post_1	post_4	potensi_stunting	No
count	6.970000e+02	697	697	697	697	697	697	697	697	...	697	697	697	697	697	697	697	697	697	697.000000
unique	Naïv	646	2	2	2	2	2	2	2	...	2	2	2	2	2	2	2	2	2	Naïv
top	Naïv	SUPRIYANTO	X	X	V	X	X	X	X	...	X	X	X	X	X	V	X	X	V	Naïv
freq	Naïv	6	655	580	695	672	691	693	696	693	...	699	691	641	611	693	576	678	454	692
mean	1.920773e+15	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	...	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv
std	7.122702e+14	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	...	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv
min	8.720202e+14	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	...	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv
25%	1.972020e+15	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	...	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv
50%	1.972020e+15	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	...	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv
75%	1.972020e+15	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	...	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv
max	1.000000e+16	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	...	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv	Naïv

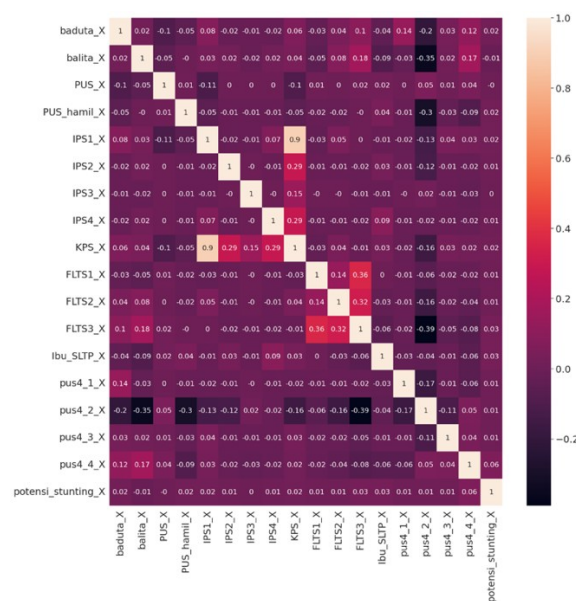
Gambar 2. Dataset stunting

Dataset berisi 697 entri yang merepresentasikan data keluarga di Kota Metro, Provinsi Lampung dengan fokus pada identifikasi faktor-faktor risiko stunting. Mayoritas kolom merupakan variabel kategorikal seperti status baduta, balita, pasangan usia subur (PUS), serta indikator sosial-ekonomi dan sanitasi, yang didominasi nilai "X" dan "V". Kolom Nama memiliki 646 nilai unik, dengan nama "SUPRIYANTO" paling sering muncul. Terdapat dua kolom numerik, yaitu No dan potensi\_stunting, di mana potensi\_stunting memiliki nilai rata-rata 349 dan menunjukkan variasi yang cukup tinggi. Beberapa kolom seperti Ibu\_SLTP memiliki missing value, yang nantinya diperbaiki pada tahap praproses data. Secara keseluruhan, dataset telah relevan untuk digunakan dalam penerapan algoritma Naïve Bayes guna memprediksi dan menentukan intervensi bagi keluarga berisiko stunting di wilayah Kota Metro.

Selanjutnya, dilakukan pra-pemrosesan data yang dimulai dengan identifikasi nilai kosong menggunakan fungsi `isnull().sum()` untuk mengetahui kolom yang memiliki *missing values*. Selanjutnya, dilakukan *one-hot encoding* terhadap variabel kategorikal menggunakan `pd.get_dummies(data, drop_first=True)` agar seluruh fitur dapat dibaca oleh algoritma Naïve Bayes dalam bentuk

numerik biner. Setelah data dikonversi ke bentuk numerik, dilakukan *feature correlation analysis* dengan menghitung matriks korelasi menggunakan `.corr()`, lalu divisualisasikan dalam bentuk heatmap menggunakan Seaborn. Visualisasi ini bertujuan untuk melihat hubungan antar fitur dan mengidentifikasi variabel yang relevan terhadap `potensi_stunting_X`, sehingga dapat mendukung pemilihan fitur dalam model klasifikasi. Visualisasi dalam bentuk heatmap ditunjukkan pada Gambar 3.

Berdasarkan *heatmap* korelasi yang ditampilkan pada Gambar 3, menunjukkan beberapa hal terkait hubungan antar fitur dan `potensi_stunting`. Fitur `PUS_X` menunjukkan korelasi yang kuat dengan indikator sosial ekonomi seperti `IPS1_X` (0.79) dan `IPS2_X` (0.29), mengindikasikan bahwa status pasangan usia subur sangat terkait dengan faktor-faktor sosial yang mempengaruhi kondisi keluarga, seperti pendapatan atau akses terhadap fasilitas kesehatan. Hal ini menunjukkan bahwa status pasangan usia subur dapat menjadi faktor kunci dalam merencanakan intervensi kesehatan berbasis risiko stunting. Di sisi lain, variabel `FLTS3_X` memiliki korelasi yang cukup kuat dengan `FLTS1_X` (0.36) dan `FLTS2_X` (0.32), yang mengindikasikan adanya hubungan antara kondisi sanitasi dan fasilitas keluarga, yang berperan dalam menentukan risiko stunting. Kondisi sanitasi yang buruk, seperti akses terbatas ke air bersih dan fasilitas sanitasi yang tidak memadai, perlu menjadi fokus utama dalam perencanaan intervensi yang bertujuan menurunkan angka stunting.



Gambar 3. Visualisasi distribusi dataset dengan *feature correlation analysis*

Di sisi lain, beberapa fitur seperti `pus4_2_X` dan `pus4_3_X` menunjukkan korelasi negatif yang cukup signifikan, terutama dengan nilai -0.39, yang menunjukkan adanya hubungan terbalik antara beberapa faktor yang dapat mempengaruhi risiko stunting. Walaupun hubungan antara fitur-fitur ini dengan `potensi_stunting_X` relatif lemah, seperti

yang terlihat pada korelasi antara `baduta_X`, `balita_X`, dan `potensi_stunting_X` (0.02 dan 0.04), masih terdapat indikasi bahwa faktor-faktor tersebut berpotensi memengaruhi stunting secara tidak langsung.

Setelah melalui tahap preprocessing dan *feature correlation analysis*, proses dilanjutkan dengan implementasi algoritma Naïve Bayes untuk klasifikasi keluarga berisiko stunting. Sebelum model dibangun, dataset terlebih dahulu dibagi menjadi tiga bagian: 70% data latih (training), 10% data validasi (validation), dan 20% data uji (testing). Pembagian ini dilakukan untuk memastikan model belajar dengan baik, serta dapat mengevaluasi secara objektif pada data yang tidak pernah dilihat sebelumnya. Model kemudian dikembangkan menggunakan tiga varian algoritma Naïve Bayes, yaitu GaussianNB, MultinomialNB, dan BernoulliNB, sesuai dengan karakteristik distribusi data.

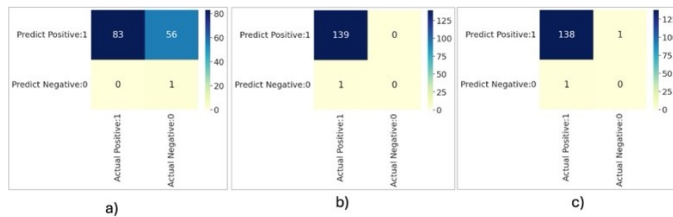
Hasil prediksi yang diperoleh dari model ini dapat digunakan untuk menentukan jenis intervensi yang sesuai bagi keluarga yang berisiko stunting, seperti program gizi atau peningkatan akses sanitasi. Misalnya, keluarga dengan status sosial ekonomi rendah yang terdeteksi berisiko stunting dapat diprioritaskan untuk program bantuan gizi dan sanitasi. Dengan cara ini, model ini tidak hanya memberikan klasifikasi, tetapi juga membantu dalam merencanakan intervensi yang lebih terarah dan efisien. Model dilatih menggunakan data latih, disesuaikan dan disetting melalui data validasi sebagaimana ditunjukkan pada Tabel 1.

TABEL 1  
IMPLEMENTASI VARIAN NAÏVE-BAYES DARI LIBRARY SCIKIT-LEARN

Gaussian Naïve Bayes	Multinomial Naïve Bayes	Bernoulli Naïve Bayes
<ul style="list-style-type: none"> <li>GaussianNB</li> </ul> GaussianNB()	<ul style="list-style-type: none"> <li>MultinomialNB</li> </ul> MultinomialNB()	<ul style="list-style-type: none"> <li>BernoulliNB</li> </ul> BernoulliNB()

Setelah model Naïve Bayes dilatih menggunakan dataset yang telah diproses, langkah berikutnya adalah melakukan prediksi untuk menentukan intervensi yang diperlukan bagi keluarga berisiko stunting di Kota Metro. Data uji yang sebelumnya belum dilihat oleh model akan digunakan untuk memprediksi apakah keluarga tersebut berisiko mengalami stunting. Hasil prediksi ini akan memberikan nilai biner, di mana nilai 1 menunjukkan keluarga yang berisiko stunting dan membutuhkan intervensi, sementara nilai 0 menunjukkan keluarga yang tidak berisiko.

Hasil klasifikasi ini kemudian dapat diintegrasikan dalam sistem pengambilan keputusan yang digunakan oleh pemerintah daerah terkait, memungkinkan penentuan prioritas bantuan dan jenis intervensi yang paling sesuai, mempercepat respons terhadap masalah kesehatan masyarakat, serta memaksimalkan sumber daya yang tersedia. Hasil prediksi disajikan pada Gambar 4.



Gambar 4. Hasil Prediksi Dataset dengan Algoritma Naïve Bayes. a) Gaussian Naïve Bayes. b) Multinomial Naïve Bayes. c) Bernoulli Naïve Bayes

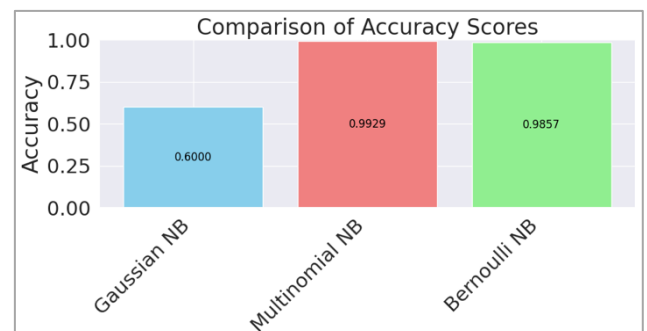
Berdasarkan Gambar 4 yang diperoleh hasil untuk ketiga model, model Gaussian Naïve Bayes menunjukkan kinerja yang baik dengan 83 true positives (TP), yang berarti model berhasil mengidentifikasi 83 keluarga yang benar-benar berisiko stunting sebagai berisiko. Namun, model ini juga menghasilkan 56 false positives (FP), yang mengindikasikan bahwa 56 keluarga yang sebenarnya tidak berisiko stunting, salah diklasifikasikan sebagai berisiko. Hal ini menunjukkan adanya kecenderungan model untuk memberikan prediksi positif pada keluarga yang tidak berisiko, yang dapat mengarah pada pemborosan sumber daya dalam intervensi. Imbalance data mungkin menjadi faktor penyebab tingginya jumlah false positives pada model ini, terutama jika terdapat lebih banyak keluarga yang tidak berisiko dibandingkan yang berisiko. Selain itu, penting untuk dicatat bahwa model ini tidak memiliki false negatives (FN), yang berarti semua keluarga yang benar-benar berisiko stunting teridentifikasi dengan benar. Kelemahan utamanya adalah dalam false positives, yang mengindikasikan perlunya penyesuaian lebih lanjut pada model untuk mengurangi tingkat kesalahan klasifikasi, serta evaluasi lebih lanjut menggunakan cross-validation atau data baru untuk memastikan generalizability model.

Selanjutnya, pada model Multinomial Naïve Bayes, hasilnya lebih baik dibandingkan dengan Gaussian Naïve Bayes. Model ini berhasil mengidentifikasi 139 true positives (TP), yang merupakan jumlah tertinggi di antara ketiga model. Selain itu, model ini hanya menghasilkan 1 false positive (FP), yang menunjukkan bahwa hampir tidak ada keluarga yang tidak berisiko stunting yang salah diklasifikasikan sebagai berisiko. Ini menunjukkan bahwa model ini sangat efisien dalam mengklasifikasikan keluarga yang berisiko stunting, dengan sangat sedikit kesalahan pada keluarga yang tidak berisiko. Selain itu, tidak ada false negatives (FN), yang berarti model ini sangat akurat dalam mendeteksi keluarga yang berisiko stunting. Secara keseluruhan, Multinomial Naïve Bayes memberikan hasil yang sangat akurat dan efisien, terutama dalam menghindari kesalahan klasifikasi pada keluarga yang tidak berisiko. Keunggulan model ini juga didukung oleh penggunaan data kategorikal yang lebih sesuai dengan karakteristik distribusi data stunting. Multinomial Naïve Bayes memberikan kinerja yang lebih baik, terutama pada data yang memiliki distribusi frekuensi fitur yang lebih kompleks.

Model Bernoulli Naïve Bayes menunjukkan hasil yang hampir serupa dengan Multinomial Naïve Bayes. Dengan 138

true positives (TP), model juga sangat efektif dalam mengidentifikasi keluarga yang berisiko stunting. Model ini menghasilkan 1 false positive (FP), yang menunjukkan bahwa hanya sedikit keluarga yang tidak berisiko stunting yang salah diklasifikasikan sebagai berisiko. Seperti halnya Multinomial Naïve Bayes, tidak ada false negatives (FN), yang berarti model tidak melewatkan keluarga yang berisiko stunting. Bernoulli Naïve Bayes juga memberikan hasil yang sangat baik dalam mengklasifikasikan keluarga berisiko stunting dengan sedikit kesalahan. Namun, meskipun keduanya hampir identik dalam hal performa, Multinomial Naïve Bayes sedikit lebih unggul karena distribusi frekuensi fitur pada data stunting lebih kompleks dan cocok dengan karakteristik model tersebut.

Setelah mengevaluasi kinerja masing-masing model, selanjutnya disajikan hasil perbandingan akurasi model Gaussian, Multinomial, dan Bernoulli Naïve Bayes sebagaimana ditunjukkan pada Gambar 5.



Gambar 5. Perbandingan Akurasi Model Naïve Bayes

Berdasarkan perbandingan nilai akurasi ketiga model Naïve Bayes pada Gambar 5, dapat dilihat bahwa Multinomial dan Bernoulli Naïve Bayes memberikan kinerja yang sangat baik, dengan akurasi masing-masing sebesar 99.29% dan 98.57%. Kedua model menunjukkan kemampuan yang sangat baik dalam mengklasifikasikan keluarga berisiko stunting dengan sedikit kesalahan pada keluarga yang tidak berisiko.

Dari hasil prediksi dataset pada Gambar 4 didapatkan juga perbandingan Precision, Recall, dan F1-Score untuk ketiga model Naïve Bayes sebagaimana disajikan pada Tabel 2.

TABEL 2  
PERHITUNGAN CONFUSION MATRIKS

Model NB	Perhitungan Confusion Matriks		
	Precision	Recall	F1-Score
Gaussian	0.597	1.0	0.747
Multinomial	0.592	1.0	0.744
Bernoulli	0.597	1.0	0.747

Berdasarkan pada Tabel 2, dapat dianalisis bahwa hasil evaluasi menunjukkan adanya perbedaan perspektif antara metrik klasifikasi. Dari sisi recall, ketiga model Naïve Bayes mampu mendeteksi seluruh keluarga berisiko stunting (recall = 1.0), sehingga tidak ada kasus positif yang terlewat, dengan



nilai F1-score relatif seimbang di kisaran 0,74–0,75, di mana Gaussian dan Bernoulli sedikit lebih unggul dibandingkan Multinomial. Namun, dari sisi akurasi keseluruhan, Multinomial Naive Bayes menunjukkan performa terbaik dengan nilai 0,9929, diikuti Bernoulli 0,9857, sementara Gaussian hanya 0,6000. Hal ini mengindikasikan bahwa Multinomial lebih efisien dalam meminimalkan kesalahan prediksi secara total, sedangkan Gaussian dan Bernoulli lebih menonjol dalam keseimbangan presisi dan recall pada kelas positif. Dengan demikian, pemilihan model terbaik bergantung pada prioritas yaitu, untuk tujuan utama pencegahan stunting melalui deteksi dini tanpa melewatkan keluarga berisiko, maka Gaussian atau Bernoulli lebih relevan, sedangkan jika menekankan efisiensi sistem dan pengurangan false positive, multinomial menjadi pilihan yang lebih tepat.

Secara teoretis, Gaussian Naive Bayes lebih sesuai untuk data numerik dengan asumsi distribusi normal. Namun, keterbatasannya dalam menangani data kategorikal yang dominan pada dataset stunting membuat kinerjanya kurang optimal. Sebaliknya, Multinomial dan Bernoulli Naive Bayes lebih mampu menangkap pola distribusi frekuensi dan variabel biner, sehingga menghasilkan performa yang lebih tinggi dan lebih tepat digunakan karena lebih sesuai dengan distribusi frekuensi data stunting.

#### IV. KESIMPULAN

Berdasarkan evaluasi terhadap tiga algoritma Naive Bayes, penelitian ini menemukan bahwa pemilihan jenis algoritma sangat mempengaruhi kinerja klasifikasi. Model yang menggunakan pendekatan distribusi yang sesuai dengan karakteristik data, seperti Multinomial Naive Bayes dan Bernoulli Naive Bayes, mampu menghasilkan performa yang jauh lebih baik dibandingkan dengan Gaussian Naive Bayes. Secara keseluruhan, hasil penelitian ini menegaskan pentingnya memahami sifat data sebelum menentukan metode klasifikasi yang akan digunakan. Temuan ini juga menunjukkan bahwa model dengan pendekatan yang sesuai dengan distribusi data, seperti Multinomial Naive Bayes, sedikit lebih unggul karena lebih cocok dengan data kategorikal frekuensi yang dominan pada dataset ini. Penelitian ini juga mencatat bahwa Gaussian Naive Bayes kurang efektif pada data ini, yang mungkin disebabkan oleh ketidaksesuaian tipe data dengan model tersebut.

Penelitian selanjutnya disarankan untuk melakukan pengujian pada data dengan karakteristik berbeda serta cross-validation untuk memastikan generalizability model. Perhatian juga harus diberikan terhadap imbalance data, yang bisa mempengaruhi hasil, khususnya pada Gaussian Naive Bayes, yang menghasilkan false positives tinggi. Selain itu, penelitian ini merekomendasikan perbandingan dengan baseline model seperti Logistic Regression atau Decision Tree untuk membandingkan kinerja Naive Bayes dan memberikan justifikasi lebih kuat tentang keunggulannya. Penelitian selanjutnya dapat menggunakan metrik evaluasi

tambahan, seperti AUC dan kappa, untuk memberikan gambaran yang lebih komprehensif mengenai kinerja model dalam penanganan stunting.

#### DAFTAR PUSTAKA

- [1] F. N. Ainy, T. Susanto, L. A. Susumaningrum., The relationship between environmental sanitation of family and stunting among under-five children: A cross-sectional study in the public health center of Jember, Indonesia., *Nursing Practice Today*, 8 (3), pp. 173 – 178, 2021. DOI: 10.18502/npt.v8i3.5932
- [2] Z. Zaatmal, R. Restila, S. Sundari & R. Palupi., The influence of environmental sanitation on stunting., *Jurnal Kesehatan Lingkungan*, 16 (1), pp. 59 – 67, 2024. DOI: 10.20473/jkl.v16i1.2024.59-67
- [3] D. K. Sari, A. Rahmiwati & R. Flora., Policy brief effectiveness of specific nutrition intervention programs as efforts to prevent stunting in Indonesia: literature review., *Media Publikasi Promosi Kesehatan Indonesia*, 7 (6), pp. 1446–1450, 2024. DOI: 10.56338/mpki.v7i6.5247
- [4] K. Wabang, O. D. Nurhayati, & Farikhin., Application of the Naive Bayes classifier algorithm to classify community complaints. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 6(5), 872-876, 2022. <https://doi.org/10.29207/resti.v6i5.4498S>.
- [5] H. Zhang, J. Mao, H. -Z. Qi, H. -Z. Xie, C. Shen, C. -T. Liu & L. Ding, L., Developing novel computational prediction models for assessing chemical-induced neurotoxicity using naïve Bayes classifier technique. *Food and Chemical Toxicology*. vol. 143, 2020. <https://doi.org/10.1016/j.fct.2020.111513>.
- [6] H. Chen, S. Hu, R. Hua, X. Zhao., Improved naïve Bayes classification algorithm for traffic risk management., *Eurasip Journal on Advances in Signal Processing*, (1), art. no. 30, 2021. DOI: 10.1186/s13634-021-00742-6.
- [7] P. J. B. Pajila, B. G. Sheena, A. Gayathri, J. Aswini, M. Nalini & R. S. Subramanian., A comprehensive survey on naïve bayes algorithm: advantages, limitations and applications. *Proceedings of the 4th International Conference on Smart Electronics and Communication, ICOSEC 2023*, pp. 1228-1234, 2023. DOI: 10.1109/ICOSEC58147.2023.10276274
- [8] I. Wickramasinghe & H. Kalutaraage, Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation., *Soft Computing*, 25 (3), pp. 2277-2293, 2021. DOI: 10.1007/s00500-020-05297-6
- [9] R. Blanquero, E. Carrizosa, P. Ramirez-Cobo & M. R. Sillero-Denamiel, Variable selection for Naive Bayes classification, *Computers and Operations Research*, 135, art. no. 105456, 2021. DOI: 10.1016/j.cor.2021.105456
- [10] M. Ismail, N. Hassan, S. S. Baffaish., Comparative analysis of naïve bayesian techniques in health-related for classification task, *Journal of Soft Computing and Data Mining*, vol. 1 (2), pp. 1-10, 2020. DOI: 10.30880/jscdm.2020.01.02.001
- [11] M. Artur., Review the performance of the bernoulli naïve bayes classifier in intrusion detection systems using recursive feature elimination with cross-validated selection of the best number of features, *Procedia Computer Science*, 190, pp. 564 – 570, 2021. DOI: 10.1016/j.procs.2021.06.066
- [12] A. O. Salau, T. A. Assegie, A. T. Akindadelo, J. N. Eneh., Evaluation of Bernoulli Naive Bayes model for detection of distributed denial of service attacks, *Bulletin of Electrical Engineering and Informatics*, 12 (2), pp. 1203 – 1208, 2023. DOI: 10.11591/eei.v12i2.4020
- [13] V. Ragu & P. J. Jayarin., Detecting flooding attacks in Distributed Denial of Service Attacks using Deep Neural Network compared with Naive Bayes, *AIP Conference Proceedings*, 3252 (1), art. no. 020176, 2025. DOI: 10.1063/5.0258744
- [14] K. Özgüna, A. Tosun, M. T. Sandikkaya., A recommender system to Detect Distributed Denial of Service Attacks with network and transport layer features., *International Conference on Information Systems Security and Privacy*, 1, pp. 390 – 397. 2024. DOI: 10.5220/0012350100003648

- [15] V. S. M. Reddy & T. Poovizhi., A novel method for enhancing accuracy in mining twitter data using Naive Bayes over Logistic Regression., 2022 International Conference on Business Analytics for Technology and Security, ICBATS 2022. DOI: 10.1109/ICBATS54253.2022.9759006
- [16] P. R. Dilliswar & L. R. Parvathy., Prediction of air pollution level in particular region area using Logistic Regression and Naive Bayes., Advances in Parallel Computing, (41), pp. 618 – 624, 2022. DOI: 10.3233/APC220088