

Implementation of Naïve Bayes Algorithm for Early Detection of Stunting Risk

Nita Mirantika¹, Ragel Trisudarmo², Tri Septiar Syamfithriani³
^{1,2,3} Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Kuningan,
Kabupaten Kuningan, Provinsi Jawa Barat, Indonesia
nita.mirantika@uniku.ac.id¹, ragel.trisudarmo@uniku.ac.id², tri@uniku.ac.id³

Article Info

Article history:

Received 2025-01-31
Revised 2025-03-06
Accepted 2025-03-13

Keyword:

Naive Bayes Algorithm,
Early Detection,
Stunting Risk,
Predictive Model.

ABSTRACT

This study aimed to develop an early detection model for stunting risk in children in Kuningan Regency using the Naïve Bayes algorithm. The model used 3,155 data with a division of 50% training data and 50% testing data, utilizing five predictor variables: gender, age, weight, height, and nutritional intake. The results demonstrated an accuracy of 66.8%, precision of 62.4%, and recall of 69.5%, indicating that the model performs adequately but requires further refinement to enhance predictive quality. Improvements can be achieved by incorporating additional variables, such as environmental factors, sanitation, and maternal nutritional status, as well as optimizing data preprocessing techniques. The findings provide a scientific basis for the Kuningan Regency Health Office to design targeted intervention strategies, including regular screening programs, specific nutritional interventions, and community health education. Effective implementation of these strategies requires collaborative efforts among local government, community health centers (puskesmas), integrated health posts (posyandu), and other stakeholders to ensure a holistic and sustainable approach to stunting prevention. This study highlights the potential of data-driven models in supporting evidence-based public health policies and interventions.



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

I. PENDAHULUAN

Stunting merujuk pada suatu kondisi di mana pertumbuhan fisik seorang anak terhambat. Kondisi ini ditandai dengan tinggi badan anak yang berada di bawah rata-rata untuk usia dan jenis kelaminnya serta dapat berdampak jangka panjang pada kesehatan fisik, kemampuan kognitif, dan perkembangan anak [1]. Faktor-faktor yang dapat menyebabkan *stunting* pada anak dapat terjadi selama kehamilan, persalinan, menyusui, atau masa nifas, misalnya MPASI tidak diberikan secara memadai kepada balita. Selain itu, sanitasi lingkungan yang buruk dapat membuat anak lebih rentan terhadap infeksi penyakit. Pola asuh yang buruk juga dapat menjadi salah satu penyebab *stunting*, seperti faktor ibu yang terlalu muda atau kehamilan dengan jarak yang terlalu dekat [2].

Stunting merupakan keadaan buruk pada manusia karena menghambat pertumbuhan dan perkembangan organ vital

seperti otak, jantung dan ginjal sehingga berdampak terhadap tubuh berukuran pendek. Akibat dari *stunting* pada anak berisiko besar terhadap gangguan fungsi organ, keterlambatan perkembangan psikomotor, penurunan tingkat kecerdasan (IQ) serta gangguan pada kemampuan motorik dan sensorik yang berdampak negatif dalam jangka pendek dan jangka panjang. Dampak negatif jangka panjang menyebabkan penurunan kemampuan berpikir sehingga berpengaruh terhadap prestasi akademik, melemahkan kekebalan daya tahan tubuh, mengganggu kesehatan reproduksi dan dapat berpotensi terhadap risiko penyakit kronis seperti diabetes, kanker, stroke dan gangguan jantung[3]. Bahkan kondisi ini dapat meningkatkan kemungkinan mengalami disabilitas diusia tua. Sementara itu, dampak jangka pendek meliputi gangguan kecerdasan, pertumbuhan fisik yang terhambat, serta masalah pada sistem metabolisme dan pencernaan pada anak [4].

Indonesia menghadapi potensi risiko *stunting* yang sangat besar, seperti terungkap dalam Analisis Status Gizi Indonesia. Berdasarkan data tahun 2022, angka *stunting* mencapai 24,22% di 514 kabupaten/kota seluruh Indonesia. Meskipun angkanya menurun secara signifikan, *stunting* masih dianggap sebagai masalah serius di Indonesia karena angka prevalensinya selalu di atas 20% [5]. Indonesia menargetkan penurunan *stunting* sebesar 14% di tahun 2024 [1].

Kabupaten Kuningan yang terletak di Provinsi Jawa Barat tidak luput dari risiko *stunting* pada anak. Berdasarkan data dari Dinas Kesehatan Kabupaten Kuningan, dari 70.471 balita yang ditimbang, terdapat 6.598 balita atau sekitar 9,4% balita mengalami *stunting*. Meskipun angka prevalensinya tidak terlalu tinggi, namun upaya penanggulangan dan pencegahan *stunting* harus tetap digalakan untuk mengurangi risiko, dampak jangka pendek dan jangka panjang *stunting*. Oleh karena itu, diperlukan upaya deteksi dini risiko *stunting* sehingga dapat mengidentifikasi secara cepat dan tepat anak-anak yang berisiko atau telah mengalami *stunting* sejak awal [6], [7].

Salah satu cara untuk melakukan deteksi dini *stunting* pada anak adalah dengan menerapkan *data mining* fungsi klasifikasi. *Data mining* merupakan proses menemukan pengetahuan yang menarik dari data yang berjumlah besar [8]. Aktivitas *data mining* menggambarkan sebuah proses analisis yang terjadi secara iteratif, dengan tujuan mengekstrak informasi dan *knowledge* yang akurat dan berguna untuk pengambilan keputusan [9]. Fungsi klasifikasi *data mining* merupakan metode prediktif yang belajar dari data yang ada untuk membuat model yang kemudian digunakan untuk memprediksi data baru [10]. Salah satu algoritma pada fungsi klasifikasi *data mining* adalah algoritma Naïve Bayes, yaitu pengklasifikasi probabilistik sederhana yang menghitung serangkaian probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dalam kumpulan data tertentu [11]. Klasifikasi menggunakan algoritma Naive Bayes lebih disukai karena kecepatan dan kesederhanaannya. Walaupun sederhana, hasil yang diperoleh selalu mencapai kinerja yang sama dengan algoritma lainnya [12].

Terdapat penelitian yang dilakukan Yohana Malo, dkk. mengenai penentuan status *stunting* menggunakan metode Naïve Bayes di Desa Letekonda Selatan [13]. Hasil penelitian menunjukkan bahwa keseluruhan rata-rata akurasi setiap iterasi sebesar 92%, hal ini menunjukkan bahwa model Naïve Bayes dinilai cukup baik untuk mengklasifikasikan data status *stunting* balita. Selain itu, penelitian lain yang dilakukan oleh Yogi Saputra, dkk. mengenai tingkat kesehatan dan status gizi balita di Kecamatan Baros, Kecamatan Cimahi, dan kota Cimahi [14]. Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes dapat memberikan tingkat akurasi yang tinggi, mudah dipahami dan memiliki kinerja yang baik dengan hasil pengujian akurasi sebesar 95,77%. Penelitian lain yang dilakukan oleh Endah Ratna Arumi, dkk. mengenai prediksi tingkat prevalensi gizi buruk dengan menggunakan metode Naïve Bayes di kota

Magelang [15]. Hasil penelitian menunjukkan bahwa metode Naïve Bayes berhasil memprediksi besarnya kasus balita gizi buruk di Kota Magelang dengan presentase akurasi sebesar 75%.

Selanjutnya hasil penelitian sebelumnya terkait dengan perbandingan algoritma Naïve Bayes dengan algoritma lainnya dapat diuraikan bahwa pada hasil penelitian [16] Random Forest dan CNN mencatat performa terbaik dalam tugas klasifikasi. Random Forest mencapai akurasi 99,9%, dengan metrik evaluasi seperti *Recall*, Presisi, Skor F1, dan AUC yang mendekati nilai sempurna (0,999 dan 1,00). Hal ini menunjukkan ketahanan model terhadap *overfitting* dan kemampuan generalisasi yang tinggi. CNN, meskipun memerlukan sumber daya komputasi yang lebih besar, juga mencatat performa yang sangat kompetitif dengan akurasi 99,8% dan metrik evaluasi yang serupa (0,998 untuk *Recall*, Presisi, Skor F1, dan AUC 1,00). Kedua model ini terbukti efektif dalam menangani kompleksitas data dan menghasilkan prediksi yang akurat. Di sisi lain, Naïve Bayes mencatat akurasi 99,5%, dengan *Recall*, Presisi, dan Skor F1 sebesar 0,995, serta AUC 0,997. Meskipun lebih sederhana secara komputasi, Naïve Bayes tetap menunjukkan performa yang sangat baik, menjadikannya alternatif yang efisien dan efektif, terutama dalam situasi dengan sumber daya terbatas. Sementara itu, *Decision Tree* mencatat performa terendah dengan akurasi 97,8%, *Recall*, Presisi, dan Skor F1 sebesar 0,978, serta AUC 0,969. Hasil ini mengindikasikan bahwa *Decision Tree* memiliki keterbatasan dalam menangani kompleksitas data dibandingkan dengan model lainnya.

Sedangkan pada hasil penelitian [17] Dalam konteks prediksi penyakit jantung, Naive Bayes dan Support Vector Machine (SVM) mencatat performa terbaik dengan akurasi masing-masing 83,96% dan 84,08%. Naive Bayes, yang mengadopsi pendekatan probabilistik berbasis teorema Bayes, menunjukkan efisiensi yang tinggi dalam menangani dataset besar. Sementara itu, SVM membuktikan keunggulannya dalam ruang dimensi tinggi melalui kemampuan menemukan *hyperplane* optimal untuk pemisahan kelas. Meskipun K-Nearest Neighbors (KNN) dan *Decision Tree* dikenal efektif dalam pengenalan pola dan interpretasi yang intuitif, metrik kinerja spesifik kedua algoritma ini tidak dijelaskan secara rinci dalam penelitian.

Berdasarkan uraian dari hasil penelitian sebelumnya dan pertimbangan karakteristik algoritma, Naïve Bayes dipilih sebagai metode yang lebih unggul serta menjadikannya alternatif yang efisien dan efektif, terutama dalam situasi dengan sumber daya terbatas dibandingkan dengan algoritma lain seperti *Decision Tree* atau *Random Forest* untuk deteksi risiko *stunting*. Salah satu alasan utama adalah kompleksitas model dan risiko *overfitting* yang lebih tinggi pada *Decision Tree*, terutama ketika dataset mengandung banyak fitur atau *noise*. Sementara *Random Forest*, meskipun mampu mengurangi *overfitting* melalui pendekatan *ensemble learning*, memerlukan sumber daya komputasi yang lebih besar dan proses *tuning* parameter yang lebih rumit. Hal ini menjadi kendala signifikan dalam konteks penelitian *stunting*,

di mana dataset sering kali terbatas dan implementasi dilakukan di daerah dengan infrastruktur teknologi yang minim. Di sisi lain, Naïve Bayes menawarkan efisiensi komputasi yang tinggi dan kemudahan implementasi, sehingga lebih sesuai untuk digunakan dalam kondisi sumber daya terbatas. Selain itu, Naïve Bayes secara khusus unggul dalam menangani data kategorikal, yang dominan dalam Interpretasi hasil yang mudah dipahami melalui pendekatan probabilistik juga menjadi keunggulan Naïve Bayes, memudahkan tenaga kesehatan atau peneliti dalam mengambil keputusan berbasis data. Dengan demikian, Naïve Bayes dipilih sebagai metode yang lebih praktis, efisien, dan efektif untuk deteksi risiko stunting, terutama dalam konteks penelitian yang memprioritaskan akurasi, kemudahan implementasi, dan interpretasi hasil yang cepat.

Penerapan algoritma Naïve Bayes pada penelitian-penelitian tersebut, belum menjelaskan secara rinci proses transformasi yang dilakukan dalam membangun kumpulan data akhir yang akan diproses pada tahap pemodelan. Oleh karena itu, pada penelitian ini akan dilakukan proses transformasi data dengan cara diskritisasi, yaitu untuk mengubah data yang sifat kontinu (bertipe numerik) menjadi data bertipe nominal (kategori) [18]. Selain itu, penelitian ini juga akan menggunakan aplikasi Orange yang merupakan aplikasi *data mining* berbasis visual program untuk membangun model Naïve Bayes.

Penelitian ini bertujuan untuk melakukan deteksi dini risiko *stunting* dengan memprediksi status *stunting* di Kabupaten Kuningan menggunakan algoritma Naïve Bayes dengan proses transformasi data berupa diskritisasi serta dukungan aplikasi data mining Orange. Deteksi *stunting* sejak dini menjadi salah satu hal yang sangat penting dalam meningkatkan pertumbuhan dan perkembangan anak. Melalui deteksi dini, permasalahan dapat diidentifikasi lebih awal sehingga sehingga pihak-pihak yang berkepentingan dapat melakukan pengelolaan risiko lebih awal dan membuat strategi yang lebih efektif untuk mencegah *stunting* pada anak.

II. METODE

Dalam penelitian ini, metode penelitian yang diterapkan adalah metode kuantitatif dengan tahapan penelitian menggunakan metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Metode CRISP-DM dipilih sebagai kerangka kerja standar (*de facto standard*) yang secara luas diakui sebagai pengembangan proyek dalam pengembangan proyek *data mining* dan *knowledge discovery* yang terdiri dari 6 (enam) tahapan pengembangan *data mining* [8][19]. Adapun proses tahapan pada metode CRISP-DM untuk pemahaman lebih mendalam, terhadap alur dan detail tahapan dalam metode CRISP-DM ini dapat dilihat pada Gambar 1.

Tahapan pertama pada metode CRISP-DM adalah *business understanding*, yaitu memahami tujuan dan kebutuhan dari perspektif bisnis, dan menterjemahkan pengetahuan itu ke dalam pendefinisian masalah pada *data mining*. Tahap kedua

adalah *data understanding*, yaitu dimulai dengan pengumpulan data dan dilanjutkan dengan proses untuk mendapatkan pemahaman yang mendalam dan komprehensif. Tahap ketiga adalah *data preparation*, yaitu kegiatan untuk membentuk data mentah menjadi data akhir yang akan digunakan pada tahap pemodelan. Pada tahap ini mencakup kegiatan pembersihan data dari *noise* dan *missing value* serta transformasi data. Tahap keempat adalah *modeling*, yaitu dilakukan pemilihan dan penerapan berbagai teknik pemodelan *data mining*.



Gambar 1. Tahapan Metode CRISP-DM [8]

Pada penelitian ini, pemodelan dilakukan dengan menerapkan *data mining* fungsi klasifikasi dengan algoritma Naïve Bayes. Tahap kelima adalah *evaluation*, model telah dibuat dan harus berkualitas baik dari perspektif analisis data. Pada tahap ini dilakukan evaluasi model untuk mengetahui apakah model tersebut mampu mencapai tujuan yang ditetapkan pada fase awal. Pada penelitian ini, evaluasi hasil pemodelan menggunakan nilai *accuracy*, *precision* dan *recall*. Tahap terakhir adalah *deployment* atau pengembangan, yaitu pengetahuan atau informasi yang telah diperoleh dalam hal ini adalah deteksi dini risiko *stunting* dipresentasikan dalam bentuk laporan agar lebih mudah dipahami oleh Dinas Kesehatan Kabupaten Kuningan.

Pada tahap *data preparation* untuk membangun kumpulan data akhir yang akan diproses pada tahap pemodelan mencakup kegiatan pembersihan data dan transformasi data. Pada penelitian ini, transformasi data dilakukan dengan cara diskritisasi yaitu proses untuk mengubah data bertipe numerik menjadi data bertipe kategori agar mempermudah analisis dan interpretasi [18]. Salah satu metode yang dapat digunakan adalah dengan membagi data kedalam sejumlah interval yang sama diantaranya adalah dengan ukuran nilai letak kuartil. Kuartil merupakan salah satu pengukuran data dalam statistika yang membagi data ke dalam empat bagian sama besar [20]. Rumus kuartil ditunjukkan dengan persamaan 1.

$$Q_i = \frac{i}{4} (n) \tag{1}$$

Keterangan:
 Qi = Nilai kuartil ke 1, 2, 3
 n = Jumlah data

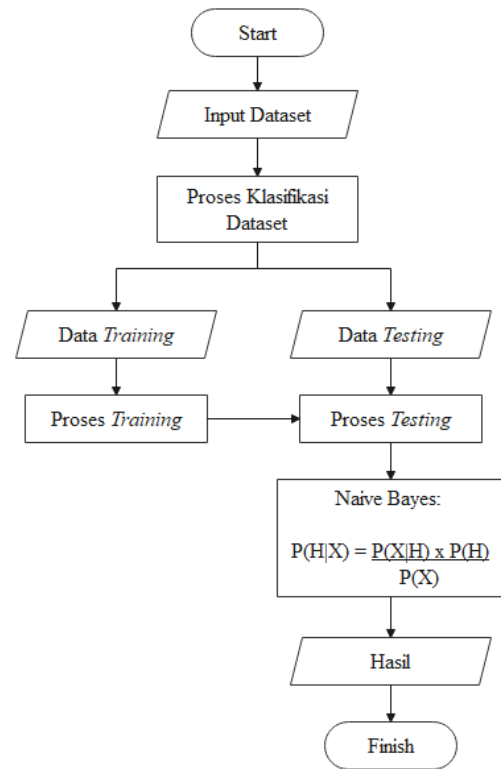
Pada tahap *modeling* dilakukan penerapan teknik pemodelan *data mining* pada fungsi klasifikasi dengan menggunakan algoritma Naïve Bayes. Algoritma Naive Bayes adalah pengklasifikasi probabilistik sederhana yang menghitung serangkaian probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dalam kumpulan data tertentu [11]. Algoritma Naïve Bayes digunakan untuk memperkirakan probabilitas setiap kelas dengan asumsi bahwa kelas-kelas tersebut independen satu sama lain. Dalam metode ini, semua atribut dianggap independen dan berkontribusi terhadap pengambilan keputusan dengan bobot yang sama. Penggunaan metode klasifikasi Naive Bayes menggunakan perhitungan probabilitas. Konsep dasar dari metode ini adalah teorema Bayes yang diterapkan dalam statistik untuk menentukan peluang pengklasifikasi Naive Bayes untuk menghitung probabilitas kelas untuk setiap kelompok yang ada dan atribut yang menentukan kelas optimal [5]. Bentuk umum rumus persamaan metode Naive Bayes ditunjukkan pada persamaan 2 [4].

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \tag{2}$$

Keterangan:
 X = Data kelas yang belum diketahui
 H = Hipotesis Data
 P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X
 P(H) = Probabilitas hipotesis H
 P(X|H) = Probabilitas X berdasarkan kondisi pada hipotesis H
 P(X) = Probabilitas X

Tahapan pada proses algoritma naïve bayes dapat dilihat pada Gambar 2 [21]. Proses algoritma Naïve Bayes dimulai dengan input dataset yang diperoleh, yang kemudian diklasifikasikan menjadi data *training* dan data *testing*. Kemudian data *training* diproses dengan menggunakan rumus Naïve Bayes, dengan menghitung probabilitas dari data *testing* sehingga menghasilkan prediksi yang akan menjadi acuan dalam pengambilan keputusan.

Evaluasi hasil pemodelan dilakukan dengan menggunakan nilai *accuracy*, *precision* dan *recall* [4][15]. *Accuracy* adalah cara mengukur kinerja dengan memeriksa kedekatan nilai prediksi dengan nilai aktual. *Precision* adalah metrik evaluasi yang mengukur ketepatan prediksi positif dari suatu model. *Precision* menunjukkan seberapa besar dari semua prediksi positif model yang benar-benar positif. Sedangkan *recall* adalah ukuran statistik yang digunakan untuk mengukur kemampuan model untuk mendeteksi semua sampel positif yang benar dari data yang tersedia.



Gambar 2. Proses Algoritma Naïve Bayes

Rumus perhitungan nilai *accuracy*, *precision* dan *recall* dapat dilihat pada persamaan 3, 4, dan 5.

$$\text{Nilai } accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

$$\text{Nilai } precision = \frac{TP}{TP+FP} \tag{4}$$

$$\text{Nilai } recall = \frac{TP}{TP+FN} \tag{5}$$

Keterangan:
 TP = Jumlah prediksi positif yang benar.
 TN = Jumlah prediksi negatif yang benar
 FP = Jumlah prediksi positif yang salah.
 FN = Jumlah prediksi negatif yang salah

III. HASIL DAN PEMBAHASAN

A. Business Understanding

Dinas Kesehatan Kabupaten Kuningan tengah berupaya untuk menanggulangi permasalahan *stunting* pada anak, yang saat ini masih menjadi salah satu tantangan besar dalam sektor kesehatan masyarakat. Untuk mengurangi risiko dan dampak jangka panjang *stunting* pada anak, maka diperlukan upaya deteksi dini risiko *stunting* sehingga dapat mengidentifikasi secara cepat dan tepat anak-anak yang berisiko atau telah mengalami *stunting* sejak awal, agar dapat diberikan intervensi yang sesuai untuk mencegah atau mengatasi

masalah tersebut sebelum kondisinya memburuk. Deteksi dini *stunting* dapat dilakukan dengan menerapkan *data mining* pada fungsi klasifikasi yang merupakan metode prediktif yang belajar dari data yang ada untuk membuat model yang kemudian digunakan untuk memprediksi data baru.

B. Data Understanding

Data yang digunakan pada penelitian ini, diambil dari data Dinas Kesehatan Kabupaten Kuningan tahun 2024 sebanyak 3.155 data balita. Variabel prediktor yang digunakan pada data tersebut yaitu variabel jenis kelamin, usia, berat badan, tinggi badan dan asupan gizi. Data penelitian dapat dilihat pada Tabel 1.

TABEL 1.
DATA PENELITIAN

No	Jenis Kelamin	Usia (bulan)	Berat Badan (kg)	Tinggi Badan (cm)	Asupan Gizi	Status <i>Stunting</i>
1	Laki-laki	57	14,2	97,4	Baik	Ya
2	Perempuan	46	12,7	98,6	Baik	Tidak
3	Laki-laki	58	14,2	96,5	Baik	Ya
...
3153	Perempuan	5	7,44	63,5	Lebih	Tidak
3154	Laki-laki	20	7,77	76	Kurang	Ya
3155	Perempuan	10	9,5	69,3	Lebih	Tidak

C. Data Preparation

Data preparation membangun kumpulan data akhir yang akan diproses pada tahap pemodelan, dalam kegiatannya mencakup pembersihan data dari *noise* dan *missing value* serta transformasi data. Pada penelitian ini, transformasi data dilakukan dengan cara diskritisasi untuk mengubah data numerik menjadi data kategori agar mempermudah analisis dan interpretasi. Data usia, berat badan dan tinggi badan yang memiliki tipe numerik ditransformasi menjadi data kategori dengan menggunakan data kuartil yang membagi data ke dalam empat bagian sama besar. Data nilai kuartil yang diperoleh dapat dilihat pada Tabel 2.

TABEL 2.
NILAI KUARTIL

Nilai Kuartil	Usia	Berat Badan	Tinggi Badan
Q1	15	9	74
Q2	30	11	86
Q3	46	14	96

Tabel 2 menunjukkan nilai kuartil yang terbentuk yaitu Q1, Q2 dan Q3. Q1 menandakan nilai yang membagi 25% data terbawah dengan 75% data sisanya. Q2 membagi data menjadi dua bagian yang sama besar yaitu 50% di bawah dan 50% di atasnya. Q3 menandakan nilai yang membagi 75% data terbawah dengan 25% data sisanya. Setelah diperoleh nilai kuartil, kemudian setiap variabel disusun menjadi empat bagian yang sama besar berdasarkan posisi nilai kuartil tersebut sehingga diperoleh nilai kategori yang dapat dilihat pada Tabel 3.

TABEL 3.
NILAI KATEGORI

Variabel	Nilai kategori	Keterangan
Usia	≤ 15	Usia dibawah 15 bulan
	16-30	Usia 16-30 bulan
	31-46	Usia 31-46 bulan
	>46	Usia diatas 46 bulan
Berat Badan	≤ 9	Berat badan dibawah 9 kg
	10-11	Berat badan 10-11 kg
	12-14	Berat badan 12-14 kg
	>14	Berat badan diatas 14 kg
Tinggi Badan	≤ 74	Tinggi badan dibawah 74 cm
	75-86	Tinggi badan 75-86 cm
	87-96	Tinggi badan 87-96 cm
	>96	Tinggi badan diatas 96 cm
Asupan Gizi	Kurang	Asupan gizi kurang
	Baik	Asupan gizi baik
	Lebih	Asupan gizi lebih

Dengan menggunakan data kategori tersebut, diperoleh data *training* yang sudah ditransformasi dan siap untuk diproses pada tahap pemodelan, dapat dilihat pada Tabel 4.

TABEL 4.
DATA TRAINING HASIL TRANSFORMASI

No	Jenis Kelamin	Usia	Berat Badan	Tinggi Badan	Asupan Gizi	Status <i>Stunting</i>
1	Laki-laki	>46	>14	>96	Baik	Ya
2	Perempuan	31-46	12-14	>96	Baik	Tidak
3	Laki-laki	>46	>14	>96	Baik	Ya
...
3153	Perempuan	≤ 15	≤ 9	≤ 74	Lebih	Tidak
3154	Laki-laki	16-30	≤ 9	75-86	Kurang	Ya
3155	Perempuan	≤ 15	≤ 9	≤ 74	Lebih	Tidak

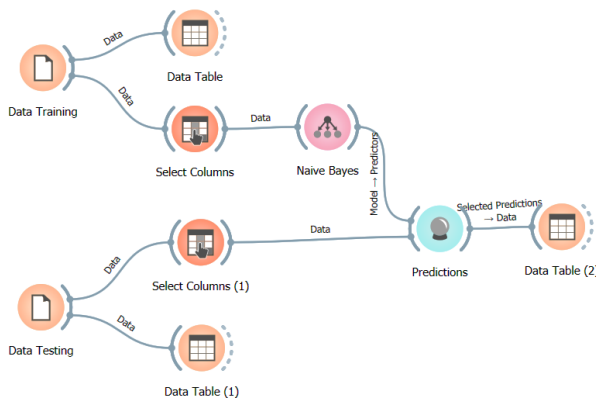
D. Modeling

Pada tahapan *modeling* dilakukan penerapan teknik data mining fungsi klasifikasi dengan algoritma Naïve Bayes. Dalam prosesnya, algoritma Naïve Bayes akan memprediksi data *testing* berdasarkan model Naïve Bayes yang telah dilatih pada data *training*. Pembagian data yang digunakan adalah 50% data *training* dan 50% data *testing*, yaitu 1.578 data *training* dan 1.577 data *testing*. Salah satu data *testing* yang digunakan pada penelitian ini ditunjukkan pada Tabel 5.

TABEL 5.
DATA TESTING PENELITIAN

Jenis Kelamin	Usia	Berat Badan	Tinggi Badan	Asupan Gizi	Status <i>Stunting</i>
Perempuan	31-46	10-11	75-86	Baik	?

Dengan dukungan menggunakan aplikasi Orange, dimana desain *modeling* pada penerapan algoritma Naïve Bayes selanjutnya dapat dilihat pada Gambar 3.



Gambar 3. Desain Model Naive Bayes

Dengan menggunakan desain model pada Gambar 3, data *testing* akan diprediksi dengan mengevaluasi kinerja model Naive Bayes yang telah dilatih pada data *training*. Selanjutnya dimana hasil prediksi data *testing* dapat dilihat pada Gambar 4.

Predictions - Orange

File View Window Help

Show probabilities for (None)

Naive Bayes	Jenis Kelamin	Usia	Berat Badan	Tinggi Badan	Asupan Gizi
1 Ya	Perempuan	31-46	10-11	75-86	Baik

Gambar 4. Hasil Prediksi Data Testing

Gambar 4 menunjukkan hasil prediksi data *testing* dan diperoleh hasil bahwa balita dengan jenis kelamin perempuan, usia 31-46 bulan, berat badan 10-11 kg, tinggi badan 75-85 dan asupan gizi baik diprediksi *stunting*. Dengan menggunakan rumus persamaan metode Naive Bayes, prediksi data *testing* dapat dilakukan dengan perhitungan sebagai berikut:

- Menghitung probabilitas label kelas
 - $P(Y=Ya) = 1.446/3.155 = 0,458$
 - $P(Y=Tidak) = 1.709/3.155 = 0,542$
- Menghitung probabilitas setiap atribut dalam data X
 - Menghitung probabilitas dengan nilai label "Ya"
 - $P(\text{Jenis Kelamin} = \text{Perempuan} | Y=Ya) = 620/1.446 = 0,429$
 - $P(\text{Usia} = 31-46 | Y=Ya) = 450/1.446 = 0,311$
 - $P(\text{Berat Badan} = 10-11 | Y=Ya) = 465/1.446 = 0,322$
 - $P(\text{Tinggi Badan} = 75-86 | Y=Ya) = 572/1.446 = 0,396$
 - $P(\text{Asupan Gizi} = \text{Baik} | Y=Ya) = 1.206/1.446 = 0,834$
 - Nilai total probabilitas $P(X|Y=Ya) = 0,429 \times 0,311 \times 0,322 \times 0,396 \times 0,834 = 0,01419$
 - Menghitung probabilitas dengan nilai label "Tidak"
 - $P(\text{Jenis Kelamin} = \text{Perempuan} | Y=Tidak) = 818/1.709 = 0,479$
 - $P(\text{Usia} = 31-46 | Y=Tidak) = 434/1.709 = 0,254$

- $P(\text{Berat Badan} = 10-11 | Y=Tidak) = 318/1.709 = 0,186$
- $P(\text{Tinggi Badan} = 75-86 | Y=Tidak) = 386/1.709 = 0,226$
- $P(\text{Asupan Gizi} = \text{Baik} | Y=Tidak) = 1.473/1.709 = 0,862$
- Nilai total probabilitas $P(X|Y=Tidak) = 0,479 \times 0,254 \times 0,186 \times 0,226 \times 0,862 = 0,00441$

- Menghitung Total probabilitas data X dengan label kelas
 - Nilai total label "Ya" = $0,01419 \times 0,458 = 0,0065$
 - Nilai total label "Tidak" = $0,00441 \times 0,542 = 0,0024$
- Membandingkan nilai total. Diperoleh bahwa nilai total untuk label "Ya" yaitu 0,0065 lebih besar daripada label "Tidak" yaitu 0,0024 sehingga nilai prediksi data *testing* adalah "Ya" yang berarti bahwa data *testing* tersebut adalah *stunting*.

E. Evaluation

Evaluasi terhadap model Naive Bayes yang terbentuk dilakukan dengan menggunakan data *testing* sebanyak 1.577 data balita di Kabupaten Kuningan. Evaluasi tersebut menggunakan nilai *accuracy*, *precision* dan *recall* yang dihitung dari *confusion matrix*. Hasil *confusion matrix* dapat dilihat pada Gambar 5.

		Predicted		Σ
		Tidak	Ya	
Actual	Tidak	552	303	855
	Ya	220	502	722
Σ		772	805	1577

Gambar 5. Confusion Matrix

Gambar 5 menunjukkan hasil *confusion matrix* dan diketahui bahwa dari 1.577 data yang diprediksi berdasarkan algoritma Naive Bayes terdapat 552 data diprediksi tidak *stunting*, 502 data diprediksi *stunting*, 303 data tidak *stunting* diprediksi sebagai *stunting*, dan 220 data *stunting* diprediksi sebagai tidak *stunting*. Perhitungan nilai *accuracy*, *precision* dan *recall* hasil *confusion matrix* yaitu sebagai berikut:

$$\text{Nilai accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{502+552}{1577} = 0,668 = 66,8\%$$

$$\text{Nilai precision} = \frac{TP}{TP+FP} = \frac{502}{502+303} = 0,624 = 62,4\%$$

$$\text{Nilai recall} = \frac{TP}{TP+FN} = \frac{502}{502+220} = 0,695 = 69,5\%$$

Perhitungan evaluasi model Naive Bayes diperoleh hasil nilai *accuracy* sebesar 66,8% yang berarti bahwa model berhasil memprediksi 66,8% dari seluruh data *testing* dengan benar, sedangkan sisanya 33,2% salah diprediksi. Nilai *precision* diperoleh sebesar 62,4% yang berarti bahwa dari semua prediksi yang dibuat oleh model sebagai positif, hanya 62,4% yang benar-benar positif, sedangkan 37,6% dari

prediksi positif ternyata salah. Nilai *recall* diperoleh sebesar 69,5% yang berarti bahwa model berhasil mendeteksi 69,5% dari semua data testing positif yang sebenarnya dalam dataset sedangkan sisanya 30,5% tidak terdeteksi dengan benar.

Hasil evaluasi ini menunjukkan bahwa model Naïve Bayes yang dikembangkan memberikan hasil yang cukup memadai, dan dapat ditingkatkan lagi dalam mengidentifikasi dan memprediksi variabel yang ditentukan, agar kualitas hasil prediksi dapat menjadi acuan yang semakin baik dalam pengambilan keputusan. Peningkatan evaluasi model dapat dilakukan dengan penambahan variabel prediktor seperti faktor lingkungan, sanitasi, status gizi ibu, dan lainnya. Selain itu, dapat pula dilakukan penambahan jumlah data *training* dan data *testing*, serta penyempurnaan teknik preprocessing data sehingga hasil prediksi dapat lebih berkualitas.

F. Deployment

Pengetahuan atau informasi yang diperoleh dalam penelitian ini adalah prediksi status *stunting* anak untuk deteksi dini risiko *stunting* yang disajikan dalam bentuk laporan agar lebih mudah dipahami oleh Dinas Kesehatan Kabupaten Kuningan. Laporan prediksi status *stunting* yang dihasilkan dapat digunakan untuk membuat rekomendasi strategi penanggulangan *stunting* di Kabupaten Kuningan, seperti program *screening* berkala, intervensi gizi spesifik, dan edukasi kesehatan masyarakat. Selain itu, perlu juga adanya kolaborasi semua *stakeholder* terkait seperti pemerintah daerah, puskesmas, dan posyandu untuk memastikan implementasi kebijakan yang holistik dan berkelanjutan, sehingga diharapkan angka *stunting* di Kabupaten Kuningan dapat berkurang.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan untuk deteksi dini risiko *stunting* pada anak di Kabupaten Kuningan dengan mengimplementasikan algoritma Naïve Bayes menggunakan 3.155 data balita dengan pembagian 50% data *training* dan 50% data *testing*, serta mempertimbangkan lima variabel prediktor (jenis kelamin, usia, berat badan, tinggi badan, dan asupan gizi), diperoleh nilai *accuracy* sebesar 66,8%, *precision* sebesar 62,4%, dan *recall* sebesar 69,5%. Hasil ini menunjukkan bahwa model yang dikembangkan telah mencapai tingkat performa yang cukup memadai, meskipun masih terdapat ruang untuk peningkatan guna memperoleh kualitas prediksi yang lebih optimal. Peningkatan akurasi model dapat dilakukan melalui penambahan variabel prediktor yang lebih relevan, seperti faktor lingkungan, sanitasi, dan status gizi ibu, serta penyempurnaan teknik preprocessing data. Selain itu, hasil penelitian ini dapat menjadi acuan ilmiah bagi Dinas Kesehatan Kabupaten Kuningan dalam merancang strategi intervensi yang lebih terarah dan efektif, seperti program *screening* berkala, intervensi gizi spesifik, dan edukasi kesehatan masyarakat. Kolaborasi multipihak antara pemerintah daerah, puskesmas, posyandu, dan *stakeholder*

terkait diperlukan untuk memastikan implementasi kebijakan yang holistik dan berkelanjutan, sehingga upaya penanggulangan *stunting* dapat dilakukan secara lebih terintegrasi dan berdampak signifikan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terimakasih kepada Lembaga Penelitian dan Pengabdian Kepada Masyarakat (LPPM) Universitas Kuningan dan Dinas Kesehatan Kabupaten Kuningan Bidang Kesehatan Masyarakat yang telah memberikan kesempatan dan dukungan hingga selesainya penelitian ini.

DAFTAR PUSTAKA

- [1] I. P. Putri, T. Tertiatavini, and N. Arminarahmah, "Analisis Perbandingan Algoritma Machine Learning untuk Prediksi Stunting pada Anak," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 257–265, 2024, doi: 10.57152/malcom.v4i1.1078.
- [2] I. Ali, D. Ade Kurnia, M. A. Pratama, and F. Al Ma'ruf, "KOPERTIP: Jurnal Ilmiah Manajemen Informatika dan Komputer Klasifikasi Status Stunting Balita Di Desa Slangit Menggunakan Metode K-Nearest Neighbor," *J. Ilm. Manaj. Inform. dan Komput.*, vol. 05, no. 03, pp. 35–38, 2021, [Online]. Available: <http://jurnal.kopertipindonesia.or.id/index.php/kopertip>.
- [3] M. R. D. Mustakim, Irwanto, R. Irawan, M. Irmawati, and B. Setyoboedi, "Impact of Stunting on Development of Children between 1–3 Years of Age," *Ethiop. J. Health Sci.*, vol. 32, pp. 569–578, 2022.
- [4] U. R. Gurning, S. F. Octavia, D. R. Andriyani, N. Nurainun, and I. Permana, "Prediksi Risiko Stunting pada Keluarga Menggunakan Naïve Bayes Classifier dan Chi-Square," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 172–180, 2024, doi: 10.57152/malcom.v4i1.1074.
- [5] W. Widhari, A. Triayudi, and R. T. K. Sari, "Implementation of Naïve Bayes and K-NN Algorithms in Diagnosing Stunting in Children," *SAGA J. ...*, vol. 2, no. 1, pp. 164–174, 2024, doi: 10.58905/SAGA.vol2i1.242.
- [6] A. D. Laksano, R. D. Wulandari, N. Amaliah, and R. W. Wisnuwardani, "Stunting among children under two years in Indonesia: Does maternal education matter?," *PLoS One*, vol. 17, 2022.
- [7] N. C. Parinduri, "Cegah Stunting Dan Gizi Buruk Pada Balita Dengan Edukasi Gizi Bagi Tumbuh Kembang Anak Di Puskesmas Batahan Kabupaten Mandailing Natal Tahun 2022," *J. Kesehat. Ilm. Indones. (Indonesian Heal. Sci. Journal)*, 2022.
- [8] S. Pratama, I. Iswandi, A. Sevitan, and T. P. Anjani, "Penerapan Data Mining Untuk Memprediksi Prestasi Akademik Mahasiswa Menggunakan Algoritma C4.5 dengan CRISP-DM," *J. Appl. Informatics Comput.*, vol. 7, no. 1, pp. 20–14, 2023, doi: 10.30871/jaic.v7i1.4998.
- [9] I. Kenedi and E. Suryana, "Penerapan Metode K-Means Clustering Dalam Pengelompokan Data Pasien Rawat Inap Peserta BPJS Di Rumah Sakit Umum Daerah Kabupaten Kaur," vol. 20, no. 2, pp. 493–500, 2024.
- [10] M. H. Sukri and Y. Handrianto, "Penerapan Algoritma C4. 5 Dalam Menentukan Prediksi Prestasi Siswa Pada SMPN 51 Jakarta," *Informatics Comput. Eng. J.*, vol. 4, no. 1, pp. 11–24, 2024.
- [11] R. Firnando, "Sistem Pakar Stunting Pada Balita Menggunakan Metode Forward Chaining dan Naive Bayes," *J. Sains Inform. Terap.*, vol. 1, no. 2, pp. 115–119, 2022.
- [12] A. P. U. Rahmayadi, U. Enri, and P. Purwantoro, "Klasifikasi

- Kinerja Asisten Laboratorium Selama Pandemi Covid-19 Menggunakan Algoritma Naïve Bayes,” *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 122–127, 2021, doi: 10.30871/jaic.v5i2.3261.
- [13] Y. Malo and A. U. Janga, “Klasifikasi Penentuan Stunting Menggunakan Metode Naïve Bayes (Studi Kasus: Desa Letekonda Selatan),” *J. Ilmu Komput. dan Sist. Inf.*, vol. 6, pp. 217–226, 2023.
- [14] Y. Saputra, N. Khoirunisa, and S. Arinal Haqq, “Classification of Health and Nutritional Status of Toddlers Using the Naïve Bayes Classification,” *CoreID J.*, vol. 1, no. 2, pp. 49–57, 2023, doi: 10.60005/coreid.v1i2.8.
- [15] E. R. Arumi, Sumarno Adi Subrata, and Anisa Rahmawati, “Implementation of Naïve bayes Method for Predictor Prevalence Level for Malnutrition Toddlers in Magelang City,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 2, pp. 201–207, 2023, doi: 10.29207/resti.v7i2.4438.
- [16] S. Sharma, K. Guleria, S. Kumar, and S. Tiwari, “Deep Learning based Model for Detection of Vitiligo Skin Disease using Pretrained Inception V3,” *Int. J. Math. Eng. Manag. Sci.*, vol. 8, no. 5, pp. 1024–1039, 2023, doi: 10.33889/IJMEMS.2023.8.5.059.
- [17] S. M. H. S. Iqbal, N. Jahan, A. S. Moni, and M. Khatun, “An Effective Analytics and Performance Measurement of Different Machine Learning Algorithms for Predicting Heart Disease,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 2, pp. 429–433, 2022, doi: 10.14569/IJACSA.2022.0130250.
- [18] N. Fajriati and B. Prasetyo, “Optimasi Algoritma Naive Bayes dengan Diskritisasi K-Means pada Diagnosis Penyakit Jantung,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 3, pp. 503–512, 2023, doi: 10.25126/jtiik.20231036510.
- [19] G. C. Lestari and S. Supatmi, “Analisis Prediksi Kelulusan Course Pada E-Learning Menggunakan Model Klasifikasi,” *J. Tata Kelola dan Kerangka Kerja Teknol. Inf.*, vol. 9, no. 2, pp. 79–85, 2023, doi: 10.34010/jtk3ti.v9i2.11319.
- [20] M. R. Maulana, A. Abdunnur, and M. R. Syahrir, “Analisis Kuartil, Desil Dan Persentil Pada Ukuran Panjang Udang Loreng (Mierspenaeopsis Sculptilis) Di Perairan Muara Ilu Kabupaten Kutai Kartanegara,” *Trop. Aquat. Sci.*, vol. 1, no. 1, pp. 10–16, 2022, doi: 10.30872/tas.v1i1.467.
- [21] N. Nurhayati aris, “Penerapan Algoritma Naive Bayes Dalam Menentukan Kelayakan Nasabah Pada Koperasi Simpan Pinjam,” *J. Comput. Inf. Syst. (J-CIS)*, vol. 1, no. 1, pp. 1–9, 2021, doi: 10.31605/jcis.v2i2.811.