

Implementasi Metode *Bagging* dan Teknik *Discretization* pada Algoritma *Machine Learning* untuk Memprediksi Status *Stunting* pada Anak Balita

Annisa Maulana Majid ^{1*}, Ismasari Nawangsih ²

^{1,2}Program Studi Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa, Kabupaten Bekasi, Provinsi Jawa Barat, Indonesia.

Corresponding Email: annisa.maulanamajid@pelitabangsa.ac.id ^{1*}

Histori Artikel:

Dikirim 10 Desember 2024; *Diterima dalam bentuk revisi* 20 Desember 2024; *Diterima* 5 Januari 2025; *Diterbitkan* 10 Januari 2025. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

Abstrak

Stunting menjadi salah satu masalah masa pertumbuhan anak balita, mengakibatkan anak balita rentan terhadap penyakit. Upaya pencegahan stunting dengan pengecekan rutin setiap satu bulan. Posyandu Sukasejati merupakan fasilitas untuk pengecekan pertumbuhan anak balita secara rutin namun pendataan diperlukan analisa stunting secara dini untuk membantu tenaga kesehatan mengurangi jumlah status stunting anak balita. Penelitian sebelumnya tentang prediksi stunting menggunakan algoritma Machine Learning Naive Bayes telah dilakukan namun tingkat akurasi masih rendah sehingga perlu teknik peningkatan akurasi untuk memberikan informasi yang akurat. Tujuan penelitian yaitu implementasi metode Bagging untuk peningkatan akurasi dan teknik discretization untuk merubah atribut kontinu menjadi kategorikal pada algoritma Machine Learning Naive Bayes dalam memprediksi stunting pada anak balita, hasil penelitian menunjukkan peningkatan accuracy, recall, dan precision menggunakan gabungan metode Bagging dan algoritma Naive Bayes yaitu accuracy sebesar 100% meningkat 5.83% dibandingkan menggunakan algoritma Naive Bayes mandiri yaitu sebesar 94.17% serta peningkatan hasil recall dan precision sebesar 28.33%.

Kata Kunci: Stunting; Bagging; Discretization; Algoritma Machine Learning; Naive Bayes.

Abstract

Stunting is one of the problems of toddler growth, making toddlers susceptible to disease. Efforts to prevent stunting with routine checks every month. Posyandu Sukasejati is a facility for routinely checking the growth of toddlers, but data collection requires early stunting analysis to help health workers reduce the number of stunting statuses in toddlers. Previous research on stunting prediction using the Naive Bayes Machine Learning algorithm has been carried out, but the level of accuracy is still low, so accuracy improvement techniques are needed to provide accurate information. The purpose of the study was to implement the Bagging method to improve accuracy and the discretization technique to change continuous attributes to categorical in the Naive Bayes Machine Learning algorithm in predicting stunting in toddlers, the results of the study showed an increase in accuracy, recall, and precision using a combination of the Bagging method and the Naive Bayes algorithm, namely accuracy of 100% increased by 5.83% compared to using the Naive Bayes algorithm alone, which was 94.17% and an increase in recall and precision results of 28.33%.

Keyword: Stunting; Bagging; Discretization; Machine Learning Algorithm; Naive Bayes.

1. Pendahuluan

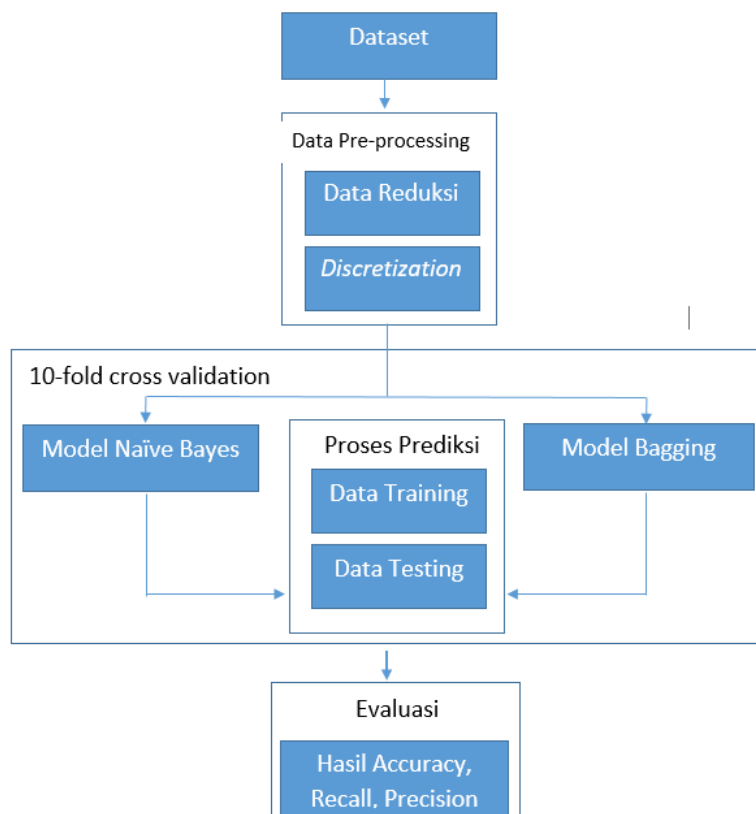
Stunting merupakan salah satu masalah gizi serius yang dapat berdampak pada terganggunya pertumbuhan dan perkembangan balita. Kondisi ini meningkatkan kerentanan balita terhadap berbagai penyakit karena kekurangan nutrisi yang signifikan selama periode penting pertumbuhan. *Stunting* disebabkan oleh ketidakseimbangan atau kekurangan asupan nutrisi, baik selama masa kehamilan maupun masa pertumbuhan balita. Masalah ini juga sering diperburuk oleh kurangnya kesadaran orang tua, khususnya ibu hamil, akan pentingnya kebutuhan nutrisi yang cukup dan seimbang bagi perkembangan janin serta balita. Pencegahan *stunting* memerlukan pendekatan komprehensif melalui edukasi dan sosialisasi kepada ibu hamil dan orang tua balita tentang pentingnya gizi yang memadai. Selain itu, upaya pencegahan ini dapat dilakukan melalui pemeriksaan rutin pertumbuhan dan perkembangan balita setiap bulan di posyandu. Berdasarkan data dari *Studi Status Gizi Indonesia* (SSGI) yang dilakukan oleh Kementerian Kesehatan Republik Indonesia pada tahun 2022, prevalensi balita yang mengalami *stunting* tercatat sebesar 21,6% (Kemenkes, 2022). Tingginya angka prevalensi ini menunjukkan bahwa masalah *stunting* masih menjadi isu yang perlu mendapatkan perhatian besar, terutama karena dampaknya yang signifikan terhadap kualitas kesehatan dan kehidupan anak-anak di masa depan. Deteksi dini melalui analisis prediktif menjadi salah satu langkah penting untuk mencegah peningkatan angka *stunting*. *Stunting* dapat diidentifikasi menggunakan parameter standar antropometri yang mencakup Berat Badan menurut Umur (BB/U) atau Tinggi Badan menurut Umur (TB/U). Di Posyandu Sukasejati, pemeriksaan rutin bulanan dilakukan untuk memantau pertumbuhan balita. Namun, proses pendataan balita yang masih dilakukan secara manual menjadi hambatan dalam upaya memprediksi status *stunting* secara dini. Hal ini mengakibatkan potensi penanganan dini menjadi kurang optimal. Untuk mengatasi permasalahan tersebut, diperlukan pendekatan berbasis teknologi, seperti penerapan algoritma *machine learning*, untuk mendukung analisis data yang lebih efisien dan akurat dalam mendeteksi status *stunting*.

Beberapa penelitian terkait prediksi *stunting* menggunakan algoritma *machine learning* telah dilakukan. Sebagai contoh, penelitian yang dilakukan oleh Alfiyah 'Ainul A. dkk menggunakan algoritma *Naïve Bayes* menunjukkan akurasi sebesar 82,14% (Ainul *et al.*, 2024). Penelitian lainnya oleh Widya Cholid Wahyudin, yang mengimplementasikan seleksi fitur *forward selection* pada algoritma *Naïve Bayes*, menghasilkan akurasi sebesar 89,67% (Cholid Wahyudin, 2020). Penelitian oleh Tikaridha Hardiani dkk dan Md. Merajul Islam dkk masing-masing menghasilkan akurasi sebesar 80% dan 76,1% menggunakan algoritma *Naïve Bayes* (Hardiani & Putri, 2024; Islam *et al.*, 2022). Walaupun hasil penelitian tersebut menunjukkan akurasi yang cukup baik, tingkat akurasi tersebut masih dapat ditingkatkan dengan penggunaan metode yang lebih efektif. Penelitian ini bertujuan untuk meningkatkan akurasi prediksi *stunting* dengan menerapkan metode Bagging dan teknik *discretization* pada algoritma *Naïve Bayes*. Metode Bagging memiliki keunggulan dalam mengurangi kesalahan prediksi dan mengatasi masalah *overfitting* pada model. Di sisi lain, teknik *discretization* digunakan untuk mengonversi atribut kontinu dalam dataset menjadi atribut kategorikal yang lebih mudah diproses oleh algoritma. Dataset yang digunakan dalam penelitian ini diperoleh dari Posyandu Sukasejati, dengan fokus pada parameter yang relevan untuk identifikasi status gizi balita. Implementasi metode dilakukan menggunakan aplikasi RapidMiner, yang memungkinkan pengolahan data, validasi model, dan evaluasi performa secara terintegrasi. Hasil penelitian ini dapat memberikan manfaat praktis dalam membantu kader posyandu dan tenaga kesehatan untuk memprediksi status *stunting* secara dini. Intervensi gizi yang diperlukan dapat dilakukan tepat waktu untuk mendukung pertumbuhan dan perkembangan balita secara optimal. Selain itu, penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan metode prediksi berbasis *machine learning* yang lebih akurat dalam bidang kesehatan masyarakat.

2. Metode Penelitian

2.1 Tahapan Penelitian

Penelitian ini dilakukan dengan mengimplementasikan metode *Bagging* dan teknik *discretization* untuk meningkatkan akurasi algoritma *Machine Learning* yaitu *Naïve Bayes* dalam memprediksi *stunting* pada anak balita di Posyandu Sukasejati menggunakan Aplikasi RapidMiner. Berikut ini merupakan tahapan dalam penelitian:



Gambar 1. Model Proses Penelitian

Tahapan dalam penelitian ini dimulai dari pengumpulan *dataset* anak balita dari Posyandu Sukasejati. Setelah pengambilan *dataset* anak balita selesai dilakukan, maka lanjut pada pengolahan data awal dengan membersihkan data-data yang *noise* atau data yang terdapat kesalahan yaitu proses *data pre-processing*. *Data Pre-processing* dilakukan untuk pembersihan data dari *missing* atau *noisy data* pada *dataset* anak balita. Pengolahan data juga dilakukan proses teknik *discretization* untuk mendapatkan data berupa kategorikal. Setelah proses data pre-processing dilakukan menghasilkan *dataset* baru untuk kembali diolah menggunakan model algoritma *Machine Learning Naïve Bayes*. *Dataset* baru juga diproses menggunakan metode *Bagging* yang dikombinasikan dengan algoritma klasifikasi *Naïve Bayes* kemudian dilakukan proses validasi menggunakan *k-fold Cross Validation* dan evaluasi. Hasil akhir sebagai evaluasi berupa *confussion matrix* yang akan menghasilkan *accuracy*, *recall*, dan *precision*.

2.2 Pengumpulan data

Pengumpulan data dalam penelitian ini menggunakan metode pengumpulan data primer dengan observasi serta kunjungan ke Posyandu Sukasejati. Sumber data yang digunakan dalam penelitian ini berasal dari kumpulan data dan laporan tentang anak balita di Posyandu Sukasejati. *Dataset* yang digunakan dalam penelitian merupakan sekumpulan *dataset* dari Posyandu Sukasejati, dengan total

data 120 data anak balita. Data tersebut terdiri dari 19 atribut dan 1 atribut label yaitu kategori status gizi sebagai berikut: NIK, Nama, Jenis Kelamin, Nama Orang Tua, Alamat, Tanggal Lahir, BB Lahir, TB Lahir, Usia Saat Ukur, Berat, Tinggi, Cara Ukur, BB/U, Zero Score BB/U, TB/U, Zero Score TB/U, BB/TB, Zero Score BB/TB, Naik Berat Badan, dan Kategori Status Gizi (Label). Dataset pada atribut kategori status gizi digunakan sebagai label dalam dataset dapat dilihat pada tabel sebagai berikut:

Tabel 1. Kategori Status Gizi

No.	Kategori Status Gizi
1	BB Tidak Naik
2	Underweigth
3	Normal
4	<i>Stunting</i>
5	Beresiko Gizi Lebih
6	Obesitas
7	Gizi Lebih

2.3 Data Pre-processing

Data *Pre-processing* merupakan proses pengolahan data yang dilakukan pada tahap awal pada dataset untuk dapat menghasilkan data dengan format yang sudah sesuai untuk proses pada langkah selanjutnya, proses ini akan menentukan kinerja model algoritma sehingga lebih optimal karena memvalidasi dan membersihkan data hilang, data tidak lengkap, dan data tidak valid (Steven Joses *et al.*, 2024). Data *Pre-processing* digunakan untuk menganalisa data yang dianggap valid agar dapat menghasilkan data yang berkualitas. Data *pre-processing* merupakan proses yang dilakukan untuk mengubah data yang mentah ke suatu bentuk format agar dapat diproses dengan mudah serta efektif. Reduksi data merupakan salah satu dari Teknik data *pre-processing* (Dessiaming *et al.*, 2022). Penelitian ini menggunakan teknik reduksi data dalam proses data *Pre-processing*. Reduksi data dilakukan dengan menghapus kolom atribut yang tidak diperlukan untuk menghasilkan dataset lebih lanjut ke suatu model *Machine Learning* (Ni Putu Viona Viandari *et al.*, 2022). Reduksi data yaitu proses menghilangkan atribut yang tidak perlu dan kurang relevan untuk hasil keputusan. Reduksi data merupakan proses menghilangkan satu atau beberapa atribut yang tidak digunakan serta ukuran penyimpanan database lebih kecil karena hanya menyertakan atribut yang efisien untuk proses mining (Situmorang *et al.*, 2022). Berdasarkan pada Peraturan Menteri Kesehatan Republik Indonesia Nomor 02 pada Tahun 2020 menyatakan bahwa suatu penilaian status gizi pada anak balita dapat dilihat dengan cara membandingkan hasil pengukuran berat badan dan panjang/tinggi badan dengan Standar Antropometri yang berdasarkan pada parameter 4 indeks, yaitu Berat Badan berdasarkan Umur (BB/U), Tinggi Badan berdasarkan Umur (TB/U), Berat Badan berdasarkan Tinggi Badan (BB/TB), dan Indeks Massa Tubuh berdasarkan Umur (IMT/U). Untuk itu penelitian ini akan mereduksi atau menentukan atribut yang relevan dalam memprediksi status *stunting* anak balita dalam dataset anak balita di Posyandu Sukasejati berdasarkan pada indeks penilaian status gizi. Hasil reduksi data dari dataset anak balita sebagai berikut, Nama, Jenis Kelamin, Tanggal Lahir, BB Lahir, TB Lahir, Usia Saat Ukur, Naik Berat Badan, Berat, Tinggi, BB/U, Zero Score BB/U, TB/U, Zero Score TB/U, BB/TB, Zero Score BB/TB, dan Kategori Status Gizi (Label).

2.4 Teknik Discretization

Teknik *Discretization* juga dilakukan pada tahap ini untuk mengolah data menjadi kategorikal. Teknik *Discretization* merupakan salah satu proses dasar dalam ilmu matematika terapan dengan cara mengubah bilangan persamaan kontinu menjadi bentuk diskrit untuk numerik analisis yang akan diimplementasi pada suatu platform digital (Yanarateş *et al.*, 2024). *Discretization* merupakan suatu teknik untuk mentransformasikan beberapa data yang besar menjadi lebih kecil untuk menjadikan kedalam skala tertentu dan terbatas (Sahu *et al.*, 2024). *Discretization* merupakan salah satu proses data

Pre-processing yang dapat mengubah data kontinu menjadi suatu nilai diskrit. *Discretization* dibagi menjadi 2 yaitu *Discretization* dengan lebar yang sama dan dengan frekuensi yang sama. Teknik *Discretization* dengan lebar yang sama yaitu membagi rentang atau range data kontinu menjadi beberapa interval namun tetap dengan nilai lebar yang sama sedangkan *Discretization* frekuensi yang sama yaitu membagi rentang atau range data kontinu menjadi beberapa interval namun tetap dengan nilai frekuensi yang sama (Putri *et al.*, 2023). Teknik *discretization* digunakan untuk proses penyederhanaan data asli agar lebih efisien. *Discretization* mentransformasikan data dari atribut kontinu menjadi atribut kategorikal, proses ini akan mempengaruhi hasil kinerja algoritma *Machine Learning* (Pertwi *et al.*, 2023). Contoh dari atribut kontinu seperti atribut kategori usia dapat dikategorikan menjadi balita, remaja, dewasa, lansia, nilai kontinu tersebut kemudian dirubah ke bentuk nominal dengan interval bin yang sudah ditentukan (Sabili & Umbara, 2024). Penelitian ini menerapkan teknik *discretization* pada dataset anak balita dengan atribut BB Lahir, TB Lahir, Berat, Tinggi, Zero Score BB/U, Zero Score TB/U, dan Zero Score BB/TB.

2.5 Naive Bayes

Naive Bayes merupakan salah satu algoritma klasifikasi sederhana dengan menghitung nilai probabilitas sebelumnya untuk hasil kelas tertentu, Teknik ini berdasarkan dari Teorema Bayes. *Naive Bayes* dapat dinyatakan fitur kelas tidak tergantung dengan fitur lain (Peretz *et al.*, 2024). *Naive Bayes* memiliki kelebihan yaitu pemahamanan yang mudah pada model klasifikasi yang dibentuk, metode klasifikasi lebih stabil dan efisien, dan data pelatihan yang dilakukan secara bertahap. Algoritma ini memenuhi kebutuhan dalam evaluasi metode *supervised learning* (Gu, 2023). *Naive Bayes* merupakan algoritma *Machine Learning* yang sederhana namun mampu bekerja dalam berbagai situasi di kehidupan nyata seperti, prediksi cuaca, penyaringan pesan spam, dan diagnosa terkait medis (Shaban *et al.*, 2021). Teknik *Naive Bayes* merupakan metode klasifikasi dengan proses perhitungan lebih cepat, mudah dipahami, dan menghasilkan akurasi yang tinggi. *Naive Bayes* merupakan sebuah metode pengelompokan statistik untuk memprediksi kemungkinan yang akan terjadi di masa depan berdasarkan masa sebelumnya (Hartono *et al.*, 2023).

2.6 Bagging

Bagging merupakan salah satu metode *Ensamble* yang melibatkan teknik pelatihan dengan sekumpulan model dengan subset data berbeda kemudian menggabungkan prediksi yang dihasilkan, sehingga dapat mengurangi masalah seperti *overfitting* (Jafari *et al.*, 2024). Algoritma *Bagging* dilakukan dengan cara meningkatkan generalisasi model yang ada, mengurangi tingkat kesalahan pada prediksi dengan cara membuat beberapa sub-data kemudian melakukan bootstrap pada sampling lalu melatih model pembelajar yang lemah (Cui *et al.*, 2024). *Bagging* digunakan untuk meningkatkan akurasi atau kualitas kluster yang lemah, dilakukan dengan cara menggabungkan beberapa kluster lemah menjadi kluster yang lebih kuat serta meningkatkan kinerja kluster secara menyeluruh (Sridhar *et al.*, 2024).

2.7 Teknik Validasi dan Evaluasi Data

Penelitian ini menggunakan *K-fold cross validation* untuk proses validasi data. *K-Fold Cross Validation* merupakan metode yang digunakan untuk mengukur tingkat keakuratan serta efisiensi dari suatu model *Machine Learning*. Cara kerja *K-Fold Cross Validation* yaitu data dibagi menjadi 2 bagian yaitu data testing dan data training, kemudian data tersebut dilatih dalam suatu model algoritma selama beberapa kali iterasi atau bagian sebanyak nilai K-subset. *K-Fold Cross Validation* diawali dengan membagi kumpulan dataset menjadi K bagian dengan nilai ukuran sama. Data pertama disebut sebagai subset pengujian, dan K-1 digunakan sebagai pelatihan model (Grzegorz Baron, 2021). *K-fold cross validation* yaitu salah satu dari metode yang digunakan untuk membagi data menjadi data training dan data testing. *K-Fold Cross Validation* dapat mengurangi nilai bias yang dihasilkan dari suatu data sampel (Ridwansyah, 2022). Penelitian ini menggunakan nilai *K-Fold Cross Validation* sebesar K=10. K dengan nilai 10 atau biasa disebut *10-fold cross validation* akan membagi data menjadi 10 bagian, 1 untuk data testing dan 9 sisanya untuk data training sebanyak 10 kali iterasi, K 10 meningkatkan efektifitas

dan efisiensi serta dapat menurunkan adanya bias (Malakouti *et al.*, 2023). Setelah proses validasi model selanjutnya proses evaluasi yang dilakukan menggunakan *confusion matrix*. *Confusion matrix* adalah suatu tabel yang menggambarkan performa hasil klasifikasi dari suatu dataset (Paraijun *et al.*, 2022). *Confusion Matrix* merupakan matrix yang digunakan untuk menguji kebenaran model prediksi dengan kondisi sebenarnya. *Confusion Matrix* merupakan metode untuk menganalisis kinerja model klasifikasi (Fatmawati *et al.*, 2023).

Tabel 2. *Confusion Matrix*

Klasifikasi Benar	Prediksi		
		Positif	Negatif
	Positif	TP (<i>True Positive</i>)	FN (<i>False Negative</i>)
Negatif	FP (<i>False Positive</i>)	TN (<i>True Negative</i>)	

Persamaan *Confusion Matrix* adalah sebagai berikut:

- 1) Nilai akurasi adalah perbandingan jumlah prediksi yang benar dengan keseluruhan data.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100\%$$

- 2) Nilai recall adalah probabilitas data yang relevan ingin dipilih antara data yang positif dengan data yang aktual positif.

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

- 3) Nilai precision adalah probabilitas data yang relevan antara data yang positif yang prediksinya benar dengan keseluruhan prediksi positif.

$$Precision = \frac{TP}{TP+FP} \times 100\%$$

3. Hasil dan Pembahasan

3.1 Hasil

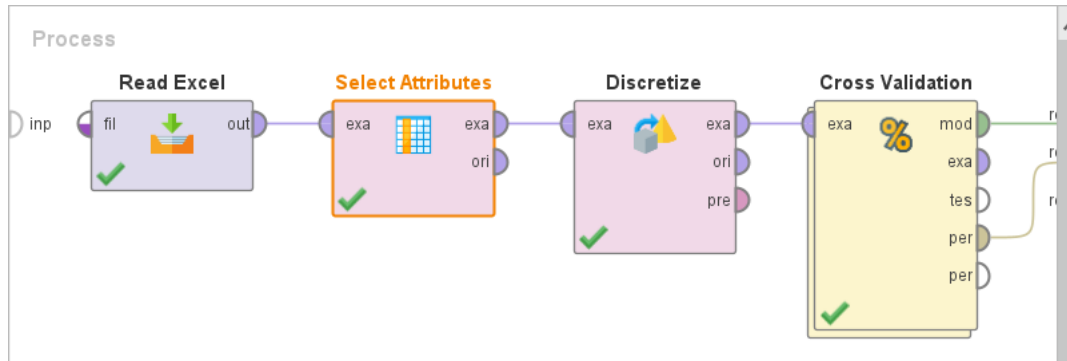
3.1.1 Pengujian Model

Penelitian ini dilakukan sebanyak 2 kali pengujian yang dimulai dari tahap import *dataset* anak balita di aplikasi RapidMiner dengan fitur Read Excel. Hasil import dataset sebagai berikut:

No	Kategori St...	NIK	Nama	JK	Tgl Lahir	BB Lahir	TB Lahir	Usia Saat U...	Berat	Tinggi
1	Normal	3216193211...	MUSA	L	Aug 11, 2...	3	50	4 Tahun - 10 ...	15	107
2	Normal	3216067008...	INAYMA DWI ...	P	Aug 30, 2...	3	50	4 Tahun - 9 B...	16	104
3	Normal	3216191809...	ATAYA ABRIZ...	L	Sep 18, 2...	3	47	4 Tahun - 9 B...	17	112
4	Normal	3216231206...	ADMADEVA A...	L	Jun 12, 2...	3	50	4 Tahun - 0 B...	14	98
5	Normal	3214120605...	RAFASYA FA...	L	May 6, 2020	3	50	4 Tahun - 1 B...	13	99
6	Normal	3216190901...	ALVIAN DON...	L	Jan 9, 2020	4	51	4 Tahun - 5 B...	14	101
7	Normal	3216190811...	MUHAMMAD ...	L	Nov 8, 2019	4	51	4 Tahun - 7 B...	16	102
8	Normal	3216235311...	DAYYAN HA...	P	Nov 13, 2...	4	51	4 Tahun - 7 B...	16	106
9	Normal	3216230305...	ARSYA FIRI...	L	May 3, 2020	3	49	4 Tahun - 1 B...	13	98
10	Normal	3216195207...	AISYAH SHA...	P	Jul 2, 2020	3	49	3 Tahun - 11 ...	15	108
11	Normal	3216236408...	SHAFANA BU...	P	Aug 24, 2...	3	48	3 Tahun - 9 B...	13	94
12	Normal	3216192411...	MUHAMMAD ...	L	Nov 24, 2...	3	50	4 Tahun - 6 B...	16	103
13	Normal	3206052101...	MUHAMMAD ...	L	Mar 21, 2...	4	54	4 Tahun - 3 B...	16	104

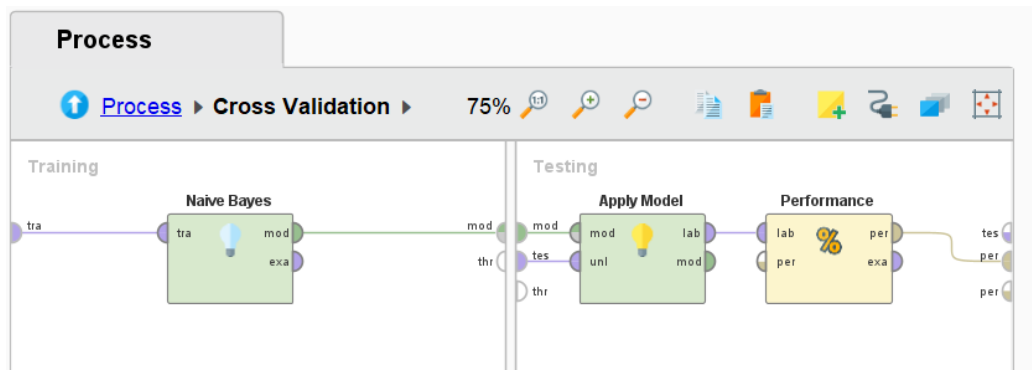
Gambar 2. Hasil Import *Dataset* Anak Balita

Selanjutnya mulai pada tahap data *pre-processing* dengan cara reduksi data dan implementasi teknik *discretization*. Pada tahap reduksi data digunakan fitur *select atribut* dan pada tahap *discretization* menggunakan fitur *Discretize by Size*. Kemudian pada tahap proses validasi digunakan fitur *cross validation*. Berikut gambar proses data *pre-processing*, teknik *discretization*, serta validasi data.



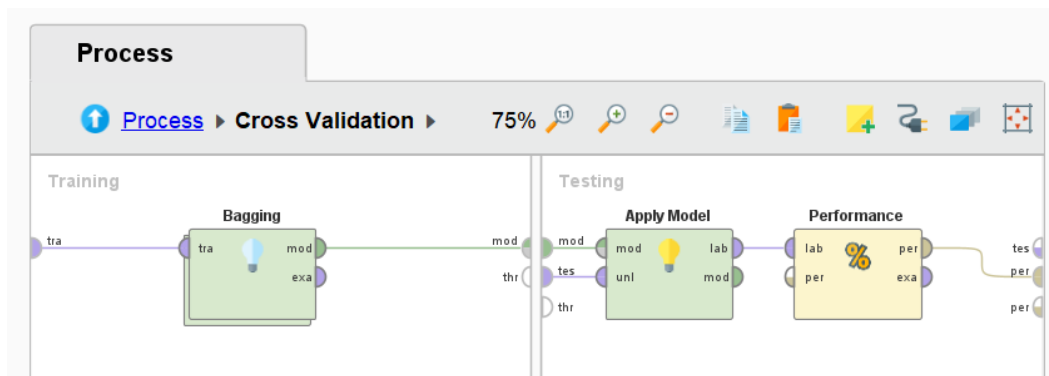
Gambar 3. Tahap Data *Pre-processing* dan Validasi

Penelitian ini dilakukan dengan pengujian pertama menggunakan model mandiri *Machine Learning* algoritma klasifikasi *Naïve Bayes*. Fitur yang digunakan yaitu *Naïve Bayes*, *Apply Model*, dan *Performance* untuk melihat hasil *accuracy*, *recall*, dan *precision*. Fitur tersebut ditambahkan di dalam *cross validation*, dapat digambarkan sebagai berikut:



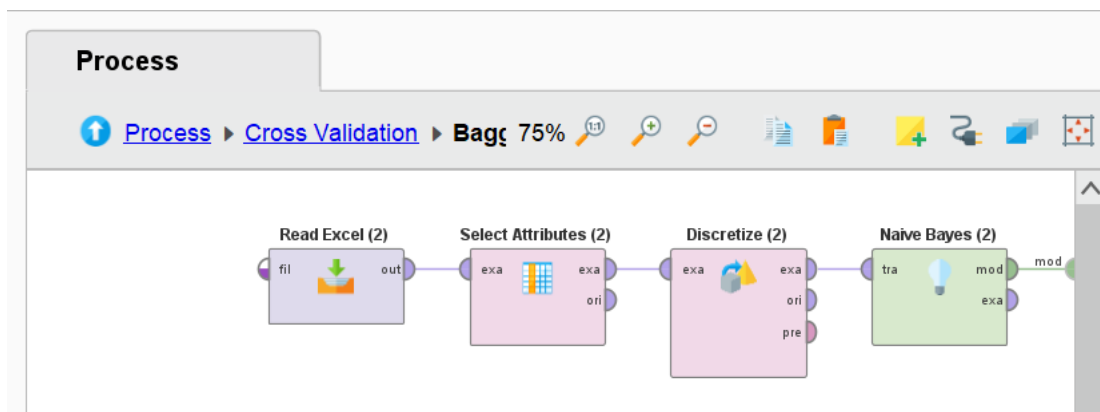
Gambar 4. Implementasi Algoritma *Naïve Bayes*

Penelitian kedua dengan implementasi metode *Bagging* yang ditambahkan di dalam *cross validation*. Gambar implementasi metode *Bagging* sebagai berikut:



Gambar 5. Implementasi Metode *Bagging*

Penelitian kedua ini menggabungkan metode *Bagging* dengan algoritma *Naïve Bayes*, sehingga fitur *Bagging* ditambahkan fitur *Read Excel*, *Select Attribute*, *Discretize*, dan *Naïve Bayes*. Gambar Implementasi *Naïve Bayes* dengan Metode *Bagging* sebagai berikut:



Gambar 6. Implementasi Metode *Bagging* dengan Algoritma *Naïve Bayes*

Hasil dari 2 pengujian menggunakan *dataset* anak balita Posyandu Sukasejati, diantaranya pengujian pertama menerapkan teknik *discretization* dan algoritma *Naïve Bayes* secara mandiri dan pengujian kedua menerapkan teknik *discretization* dengan kombinasi implementasi algoritma *Naïve Bayes* dengan metode *Bagging*. Pengujian menghasilkan kinerja yaitu *accuracy*, *recall*, serta *precision*. Berikut hasil pengujian dalam penelitian:

accuracy: 94.17% +/- 4.03% (micro average: 94.17%)

	true Normal	true Stunting	true Beresiko ...	true Gizi Lebih	true Obesitas	true Underwe...	class precision
pred. Normal	113	2	2	0	1	1	94.96%
pred. Stunting	0	0	0	0	0	0	0.00%
pred. Beresik...	0	0	0	1	0	0	0.00%
pred. Gizi Lebih	0	0	0	0	0	0	0.00%
pred. Obesitas	0	0	0	0	0	0	0.00%
pred. Underw...	0	0	0	0	0	0	0.00%
class recall	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	

Gambar 7. Hasil *Accuracy* Pengujian I

weighted_mean_recall: 16.67% +/- 0.00% (micro average: 16.67%), weights: 1, 1, 1, 1, 1, 1

	true Normal	true Stunting	true Beresiko ...	true Gizi Lebih	true Obesitas	true Underwe...	class precision
pred. Normal	113	2	2	0	1	1	94.96%
pred. Stunting	0	0	0	0	0	0	0.00%
pred. Beresik...	0	0	0	1	0	0	0.00%
pred. Gizi Lebih	0	0	0	0	0	0	0.00%
pred. Obesitas	0	0	0	0	0	0	0.00%
pred. Underw...	0	0	0	0	0	0	0.00%
class recall	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	

Gambar 8. Hasil *Recall* Pengujian I

weighted_mean_precision: 15.83% +/- 0.72% (micro average: 15.83%), weights: 1, 1, 1, 1, 1, 1

	true Normal	true Stunting	true Beresiko ...	true Gizi Lebih	true Obesitas	true Underwe...	class precision
pred. Normal	113	2	2	0	1	1	94.96%
pred. Stunting	0	0	0	0	0	0	0.00%
pred. Beresik...	0	0	0	1	0	0	0.00%
pred. Gizi Lebih	0	0	0	0	0	0	0.00%
pred. Obesitas	0	0	0	0	0	0	0.00%
pred. Underw...	0	0	0	0	0	0	0.00%
class recall	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	

Gambar 9. Hasil Precision Pengujian I

accuracy: 100.00% +/- 0.00% (micro average: 100.00%)

	true Normal	true Stunting	true Beresiko ...	true Gizi Lebih	true Obesitas	true Underwe...	class precision
pred. Normal	113	0	0	0	0	0	100.00%
pred. Stunting	0	2	0	0	0	0	100.00%
pred. Beresik...	0	0	2	0	0	0	100.00%
pred. Gizi Lebih	0	0	0	1	0	0	100.00%
pred. Obesitas	0	0	0	0	1	0	100.00%
pred. Underw...	0	0	0	0	0	1	100.00%
class recall	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	

Gambar 10. Hasil Accuracy Pengujian II

weighted_mean_recall: 28.33% +/- 8.05% (micro average: 100.00%), weights: 1, 1, 1, 1, 1, 1

	true Normal	true Stunting	true Beresiko ...	true Gizi Lebih	true Obesitas	true Underwe...	class precision
pred. Normal	113	0	0	0	0	0	100.00%
pred. Stunting	0	2	0	0	0	0	100.00%
pred. Beresik...	0	0	2	0	0	0	100.00%
pred. Gizi Lebih	0	0	0	1	0	0	100.00%
pred. Obesitas	0	0	0	0	1	0	100.00%
pred. Underw...	0	0	0	0	0	1	100.00%
class recall	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	

Gambar 11. Hasil Recall Pengujian II

weighted_mean_precision: 28.33% +/- 8.05% (micro average: 100.00%), weights: 1, 1, 1, 1, 1, 1

	true Normal	true Stunting	true Beresiko ...	true Gizi Lebih	true Obesitas	true Underwe...	class precision
pred. Normal	113	0	0	0	0	0	100.00%
pred. Stunting	0	2	0	0	0	0	100.00%
pred. Beresik...	0	0	2	0	0	0	100.00%
pred. Gizi Lebih	0	0	0	1	0	0	100.00%
pred. Obesitas	0	0	0	0	1	0	100.00%
pred. Underw...	0	0	0	0	0	1	100.00%
class recall	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	

Gambar 12. Hasil Precision Pengujian II

Berikut hasil ringkasan dari 2 pengujian dalam bentuk tabel:

Tabel 3. Hasil Uji

	Naïve Bayes	Naïve Bayes, Bagging
Accuracy	94.17%	100%
Recall	16.67%	28.33%
Precision	15.83%	28.33%

Berdasarkan dari pengujian yang telah dilakukan terjadi peningkatan *accuracy*, *recall* dan *precision* pada metode *Bagging* yang digabungkan dengan algoritma *Naïve Bayes*. *Bagging* membantu meningkatkan akurasi dan mengurangi adanya *overfitting*. Berikut adalah perbandingan hasil dari penelitian ini dengan penelitian sebelumnya.

Tabel 4. Hasil Perbandingan Penelitian

Penelitian	Algoritma	Hasil Akurasi
Penelitian ini	Naïve Bayes, Bagging, Teknik <i>Discretization</i>	100%
Ainul A. dkk	Naïve Bayes	82,14%
Tikaridha Hardiani	Naïve Bayes	80%
Md. Merajul Islam, dkk	Naïve Bayes	76.1%
Widya Cholid Wahyudin	Naïve Bayes dengan fitur <i>forward selection</i>	89.67%

Berdasarkan dari hasil perbandingan pengujian model yang telah dilakukan dengan penelitian sebelumnya, algoritma *Naïve Bayes* yang digabungkan dengan metode *Bagging* serta penanganan data dengan teknik *discretization* menghasilkan nilai akurasi lebih unggul yaitu 100%. Metode *Bagging* dapat menangani masalah *overfitting* sehingga dapat meningkatkan nilai akurasi lebih baik. Hasil dari pengujian dapat diimplementasikan untuk mempermudah tenaga kesehatan untuk mendeteksi dini adanya diagnosa *stunting* pada anak balita, sehingga dapat ditangani secara cepat dan tepat dengan memberikan informasi kebutuhan gizi agar dapat terpenuhi.

3.2 Pembahasan

Penelitian ini bertujuan untuk meningkatkan performa algoritma *Naïve Bayes* dalam memprediksi status *stunting* pada balita melalui integrasi metode *Bagging* dan teknik *discretization*. Implementasi metode dilakukan secara bertahap, dimulai dengan pengujian algoritma *Naïve Bayes* secara mandiri. Hasil pengujian awal menunjukkan akurasi sebesar 94,17%, namun *recall* dan *precision* masing-masing hanya mencapai 16,67% dan 15,83%. Nilai *recall* dan *precision* yang rendah mengindikasikan bahwa algoritma belum mampu secara optimal mengidentifikasi data positif. Hal ini disebabkan oleh atribut kontinu dalam dataset yang belum dikonversi menjadi kategori, sehingga membatasi kemampuan

algoritma untuk mendeteksi pola (Ainul *et al.*, 2024). Untuk mengatasi kekurangan tersebut, diterapkan teknik *discretization* sebagai bagian dari proses *data pre-processing*. Teknik ini mengubah atribut kontinu, seperti berat badan dan tinggi badan, menjadi atribut kategorikal dengan rentang tertentu, sehingga pola data menjadi lebih terstruktur dan mudah diinterpretasikan oleh algoritma. Penelitian sebelumnya menunjukkan bahwa teknik *discretization* mampu meningkatkan performa model klasifikasi, terutama dalam menangani atribut kontinu (Yanarates *et al.*, 2024; Putri *et al.*, 2023). Meski demikian, hasil pengujian menggunakan *Naïve Bayes* dengan teknik *discretization* masih belum sepenuhnya mengatasi masalah *overfitting* pada dataset. Langkah berikutnya adalah mengintegrasikan metode Bagging ke dalam algoritma *Naïve Bayes*. Metode Bagging bekerja dengan membagi dataset menjadi beberapa subset, melatih model secara terpisah pada setiap subset, dan menggabungkan prediksi dari masing-masing model untuk menghasilkan hasil akhir yang lebih akurat. Penggunaan metode ini terbukti efektif dalam mengurangi bias, mengatasi *overfitting*, dan meningkatkan generalisasi model (Jafari *et al.*, 2024; Cui *et al.*, 2024). Hasil pengujian menunjukkan bahwa kombinasi Bagging dan *Naïve Bayes* menghasilkan akurasi 100%, dengan peningkatan recall dan precision masing-masing menjadi 28,33%. Peningkatan performa ini mengindikasikan bahwa metode Bagging berhasil mengoptimalkan kemampuan model dalam menangkap pola data dan menghasilkan prediksi yang lebih andal.

Hasil penelitian ini dibandingkan dengan penelitian sebelumnya yang menggunakan algoritma *Naïve Bayes* untuk prediksi *stunting*. Penelitian oleh Alfiah 'Ainul A. dkk mencatat akurasi sebesar 82,14%, sementara Widya Cholid Wahyudin mencapai akurasi 89,67% dengan seleksi fitur *forward selection* (Ainul *et al.*, 2024; Cholid Wahyudin, 2020). Tikaridha Hardiani dkk dan Md. Merajul Islam masing-masing mencapai akurasi 80% dan 76,1% (Hardiani & Putri, 2024; Islam *et al.*, 2022). Dibandingkan hasil penelitian ini, kombinasi Bagging dan *Naïve Bayes* memberikan performa yang jauh lebih unggul, khususnya dalam hal akurasi. Penelitian ini memiliki implikasi signifikan dalam membantu tenaga kesehatan, khususnya di fasilitas seperti posyandu, untuk memprediksi status *stunting* secara dini. Dengan akurasi yang tinggi, model ini dapat menjadi alat bantu yang efektif untuk mendukung pengambilan keputusan dalam intervensi gizi yang tepat waktu. Meski demikian, penelitian ini memiliki beberapa keterbatasan. Ukuran dataset yang relatif kecil (120 sampel) menjadi salah satu keterbatasan yang dapat memengaruhi generalisasi model pada dataset yang lebih besar. Selain itu, penelitian ini hanya menggunakan algoritma *Naïve Bayes*, tanpa perbandingan dengan algoritma lain seperti *Decision Tree* atau *Support Vector Machines* (SVM). Penelitian di masa depan dapat mengatasi keterbatasan ini dengan menggunakan dataset yang lebih besar dan beragam, serta membandingkan performa *Naïve Bayes* dengan algoritma klasifikasi lainnya. Selain itu, pengembangan aplikasi berbasis sistem pengambilan keputusan juga dapat menjadi langkah berikutnya untuk mengimplementasikan hasil penelitian ini dalam lingkungan operasional, sehingga dapat membantu mendeteksi *stunting* secara real-time dan mempermudah intervensi yang diperlukan.

4. Kesimpulan

Penelitian ini dilakukan dengan menghasilkan tingkat akurasi tinggi dengan implementasi algoritma *Naïve Bayes* dengan metode *Bagging* serta teknik *discretization* untuk prediksi *stunting* pada anak balita di Posyandu Sukasejati. Hasil akurasi yaitu sebesar 100%, meningkat sebesar 5.83% dari pada menggunakan algoritma *Naïve Bayes* mandiri. Penelitian ini dapat digunakan untuk mempermudah para tenaga medis dalam prediksi dini *stunting* pada anak balita, sehingga dapat diambil tindakan pencegahan atau meningkatkan sosialisasi kepada orang tua untuk dapat memberikan konsumsi nutrisi yang baik dan lengkap pada anak balita. Penelitian selanjutnya yang akan datang dapat diimplementasikan menjadi aplikasi sistem pengambilan keputusan dalam prediksi status *stunting* pada anak balita. Penelitian juga dapat dilakukan dengan komparasi algoritma klasifikasi lain seperti *Decision Tree*, *K-Nearest Neighbors*, dan *Support Vector Machines* atau menambahkan metode peningkatan akurasi lain untuk mengetahui tingkat kinerja model.

5. Ucapan Terima Kasih

Kami ucapkan terima kasih kepada Direktorat Penelitian dan Pengabdian Masyarakat Universitas Pelita Bangsa yang telah mendukung secara finansial sehingga penelitian ini terlaksana dengan baik serta kami ucapkan terima kasih kepada seluruh pihak yang telah mendukung dan membantu terwujudnya penelitian ini.

6. Daftar Pustaka

- Baron, G., & Stańczyk, U. (2021). Standard vs. non-standard cross-validation: Evaluation of performance in a space with structured distribution of datapoints. *Procedia Computer Science*, *192*, 1245-1254. <https://doi.org/10.1016/j.procs.2021.08.128>.
- Cui, G., Zhang, Y., Tao, H., Man, S., & Chen, H. (2024). Optimised prediction of tunnel fire heat release rate using the ResNet18_2CLSTM model with bagging for multimodal data. *Case Studies in Thermal Engineering*, *63*, 105268.
- Dessiaming, T. Z., Anraeni, S., & Pomalingo, S. (2022). College Academic Data Analysis Using Data Visualization. *Jurnal Teknik Informatika (Jutif)*, *3*(5), 1203-1212. <https://doi.org/10.20884/1.jutif.2022.3.5.310>.
- Fatmawati, D., Trisnawati, W., Jumaryadi, Y., & Triyono, G. (2023). Klasifikasi Tingkat Kepuasan Penggunaan Layanan Teknologi Informasi Menggunakan Decision Tree. *KLIK: Kajian Ilmiah Informatika dan Komputer*, *3*(6), 1056-1062. <https://doi.org/10.30865/klik.v3i6.803>.
- Gu, Y. (2023). Exploring the application of teaching evaluation models incorporating association rules and weighted naive Bayesian algorithms. *Intelligent Systems with Applications*, *20*, 200297.
- Hartono, H., Hajjah, A., & Marlim, Y. N. (2023). PENERAPAN METODE NAÏVE BAYES CLASSIFIER UNTUK KLASIFIKASI JUDUL BERITA. *Jurnal Simantec*, *12*(1), 37-46.
- Hastuti, N. T., & Budiman, F. (2024). Optimasi Klasifikasi Stunting Balita dengan Teknik Boosting pada Decision Tree. *Edumatic: Jurnal Pendidikan Informatika*, *8*(2), 655-664. <https://doi.org/10.47709/digitech.v4i1.4481>.
- Islam, M. M., Rahman, M. J., Islam, M. M., Roy, D. C., Ahmed, N. F., Hussain, S., ... & Maniruzzaman, M. (2022). Application of machine learning based algorithm for prediction of malnutrition among women in Bangladesh. *International Journal of Cognitive Computing in Engineering*, *3*, 46-57.
- Jafari, S., Kim, J., & Byun, Y. C. (2024). Integrating ensemble learning and meta bagging techniques for temperature-specific State of Health prediction in Lithium-ion Batteries. *Energy Reports*, *12*, 2388-2407.
- Malakouti, S. M., Menhaj, M. B., & Suratgar, A. A. (2023). The usage of 10-fold cross-validation and grid search to enhance ML methods performance in solar farm power generation prediction. *Cleaner Engineering and Technology*, *15*, 100664.

- NURFA'IZAH, M. E. L. A. (2022). *KLASIFIKASI PENENTUAN GIZI STUNTING PADA BALITA MENGGUNAKAN METODE NAÏVE BAYES* (Doctoral dissertation, Universitas Islam Sultan Agung).
- Paraijun, F., Aziza, R. N., & Kuswardani, D. (2022). Implementasi Algoritma Convolutional Neural Network Dalam Mengklasifikasi Kesegaran Buah Berdasarkan Citra Buah. *Kilat*, 11(1), 1-9. <https://doi.org/10.33322/kilat.v10i2.1458>.
- Peretz, O., Koren, M., & Koren, O. (2024). Naive Bayes classifier – An ensemble procedure for recall and precision enrichment. *Engineering Applications of Artificial Intelligence*, 136, 108972. <https://doi.org/10.1016/j.engappai.2024.108972>.
- Pertiwi, D. A. A., Setyorini, P. R., Muslim, M. A., & Sugiharti, E. (2023). Implementation of discretisation and correlation-based feature selection to optimize support vector machine in diagnosis of chronic kidney disease. *Buletin Ilmiah Sarjana Teknik Elektro*, 5(2), 201–209. <https://doi.org/10.12928/biste.v5i2.7548>.
- Putri, P. A. R., Prasetyowati, S. S., & Sibaroni, Y. (2023). The performance of the equal-width and equal-frequency discretization methods on data features in classification process. *Sinkron*, 8(4), 2082–2098. <https://doi.org/10.33395/sinkron.v8i4.12730>.
- Ridwansyah, T. (2022). Implementasi text mining terhadap analisis sentimen masyarakat dunia di Twitter terhadap kota Medan menggunakan K-fold cross-validation dan Naïve Bayes classifier. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 2(5), 178–185. <https://doi.org/10.30865/klik.v2i5.362>.
- Sabili, N. L., Umbara, F. R., & Melina, M. (2024). KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN ALGORITMA CATEGORICAL BOOSTING DENGAN FAKTOR RISIKO DIABETES. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(6), 11391-11398. <https://doi.org/10.36040/jati.v8i6.11447>.
- Sahu, M., Shrivastava, A., Jhariya, D. C., Diwan, S., & Subhadarsini, J. (2024). Evaluation of correlation of physicochemical parameters and major ions present in groundwater of Raipur using discretization. *Measurement: Sensors*, 34, 101278. <https://doi.org/10.1016/j.measen.2024.101278>.
- Shaban, W. M., Rabie, A. H., Saleh, A. I., & Abo-Elsoud, M. A. (2021). Accurate detection of COVID-19 patients based on distance biased Naïve Bayes (DBNB) classification strategy. *Pattern Recognition*, 119, 108110. <https://doi.org/10.1016/j.patcog.2021.108110>.
- Situmorang, Z., Mandasari, S., Franciska, Y., Andriyani, K., & Ramadhan, P. S. (2022). Algoritma C45 dalam memprediksi minat calon mahasiswa. *Journal of Science and Social Research*, 5(1), 125. <https://doi.org/10.54314/jssr.v5i1.809>.
- Sridhar, V., Annamani, T., Renuka, M., Kumar, V. V., & Madupu, A. (2024). Bagging ensemble mean-shift Gaussian kernelized clustering-based D2D connectivity-enabled communication for 5G networks. *E-Prime - Advances in Electrical Engineering, Electronics and Energy*, 7, 100400. <https://doi.org/10.1016/j.prime.2023.100400>.

- Steven Joses, Yulvida, D., & Rochimah, S. (2024). Pendekatan metode ensemble learning untuk prakiraan cuaca menggunakan soft voting classifier. *Journal of Applied Computer Science and Technology*, 5(1), 72–80. <https://doi.org/10.52158/jacost.v5i1.741>.
- Viandari, N. P. V., Suarjaya, I. M. A. D., & Piarsa, I. N. (2022). Pemetaan Pelanggan dengan LRFM dan Two Stage Clustering untuk Memenuhi Strategi Pengelolaan. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 6(1), 130-139. <https://doi.org/10.29207/resti.v6i1.3778>.
- Wahyudin, W. C. (2020). Klasifikasi Stunting Balita Menggunakan Naive Bayes Dengan Seleksi Fitur Forward Selection. *JURNAL ILMU KOMPUTER DAN MATEMATIKA*, 1(1), 71-74.
- Yanarates, C., Zhou, Z., & Altan, A. (2024). Investigating the impact of discretization techniques on real-time digital control of DC-DC boost converters: A comprehensive analysis. *Heliyon*, 10(20). <https://doi.org/10.1016/j.heliyon.2024.e39591>.