

KLASIFIKASI BERISIKO STUNTING PADA BALITA: PERBANDINGAN K-NEAREST NEIGHBOR, NAÏVE BAYES, SUPPORT VECTOR MACHINE

Ramadya Wahyu Dwinanto✉, Arif Setia Sandi A., Rian Ardianto

Universitas Harapan Bangsa, Purwokerto, Indonesia

Email: ramadyawahyudwinanto1208@gmail.com

DOI: <https://doi.org/10.46880/jmika.Vol8No2.pp264-273>

ABSTRACT

Stunting in children under five is a significant health problem that impacts child development. This study aims to develop a classification model to predict stunting risk using SVM, KNN, and Naïve Bayes algorithms. Data from the Jatilawang Health Center included 523 under-fives with variables such as age, weight, length, arm circumference, z-score, parental education, and maternal health history. Following the CRISP-DM steps, the data was processed through handling missing data, feature selection, and dividing the data into training and testing sets with a ratio of 80:20. Results showed SVM had the highest accuracy of 90%, followed by KNN 89%, and Naïve Bayes 85%. This research produces a stunting risk prediction model that is implemented in a simple website, supporting early intervention and decision-making in stunting prevention efforts.

Keyword: *Stunting, Classification, Support Vector Machine, Naïve Bayes, K-Nearest Neighbor.*

ABSTRAK

Stunting pada balita adalah masalah kesehatan signifikan yang berdampak pada perkembangan anak. Penelitian ini bertujuan mengembangkan model klasifikasi untuk memprediksi risiko stunting menggunakan algoritma SVM, KNN, dan Naïve Bayes. Data dari Puskesmas Jatilawang mencakup 523 balita dengan variabel seperti umur, berat badan, panjang badan, lingkaran lengan, z-score, pendidikan orang tua, dan riwayat kesehatan ibu. Mengikuti tahapan CRISP-DM, data diproses melalui penanganan missing data, seleksi fitur, dan pembagian data menjadi set pelatihan dan pengujian dengan rasio 80:20. Hasil menunjukkan SVM memiliki akurasi tertinggi 90%, diikuti KNN 89%, dan Naïve Bayes 85%. Penelitian ini menghasilkan model prediksi risiko stunting yang diimplementasikan dalam website sederhana, mendukung intervensi dini dan pengambilan keputusan dalam upaya penanggulangan stunting.

Kata Kunci: *Stunting, Klasifikasi, Support Vector Machine, Naïve Bayes, K-Nearest Neighbor.*

PENDAHULUAN

Penanggulangan *stunting* esensial dalam mencapai Tujuan Pembangunan Berkelanjutan *Sustainable Development Goals* (SDGs) nomor 2, "Pemberantasan Kelaparan," dan nomor 3, "Kesehatan dan Kesejahteraan." *Stunting* bukan hanya masalah kesehatan, tetapi juga mencakup hak anak, pendidikan, dan pemberdayaan ekonomi. Mencegah *stunting* melalui nutrisi yang baik pada masa awal kehidupan menciptakan dasar kuat bagi pertumbuhan dan perkembangan anak. Dengan mendukung upaya pencegahan *stunting*, kita meningkatkan kesehatan generasi masa depan dan memperkuat fondasi untuk mencapai tujuan pembangunan berkelanjutan secara menyeluruh (Kementrian Kesehatan RI, 2021).

Salah satu tolok ukur keberhasilan pencapaian tujuan pembangunan berkelanjutan (SDGs) adalah kondisi kesehatan anak umur bawah lima tahun atau balita, yang dapat dilihat dari status gizinya. Anak-anak

balita termasuk dalam kelompok rentan terhadap kondisi gizi yang kurang, dan *stunting* menjadi salah satu permasalahan kesehatan yang dapat terjadi pada masa ini. Indonesia adalah salah satu negara yang memiliki peringkat *stunting* tinggi di dunia dengan angka *stunting* 21,6% pada tahun 2022 (Kementerian Kesehatan Republik Indonesia, 2023).

Stunting sendiri merujuk pada kondisi ketidakberhasilan pertumbuhan pada anak bawah lima tahun yang disebabkan oleh kurangnya gizi kronis, yang mengakibatkan anak memiliki postur tubuh yang lebih rendah daripada seharusnya sesuai usianya (Erik et al., 2020). *Stunting* menjadi isu gizi utama yang dapat berpengaruh terhadap aspek sosial dan ekonomi dalam kehidupan bermasyarakat (Azizah et al., 2022).

Stunting sering dibahas dalam konteks kesehatan, namun integrasi dengan ilmu komputer, khususnya analisis data, diperlukan untuk pemahaman dan penanganan yang holistik. Analisis data, terutama

data mining, dapat mengidentifikasi pola dan faktor risiko kompleks melalui deskripsi, estimasi, prediksi, klasifikasi, pengklusteran, dan asosiasi (Sukmawati et al., 2018). Klasifikasi bertujuan untuk memprediksi label kelas dari data tanpa label melalui dua tahap: pembelajaran, yaitu analisis data pelatihan menggunakan algoritma klasifikasi, dan klasifikasi, yaitu memprediksi keakuratan dengan data uji (Ariyadi, 2020).

Teknologi *machine learning* menawarkan potensi yang besar dalam upaya ini, terutama dalam hal klasifikasi status stunting berdasarkan data kesehatan yang tersedia. Melalui *machine learning*, kita dapat menganalisis data secara lebih mendalam untuk mengidentifikasi pola tersembunyi serta faktor risiko yang memengaruhi stunting. Algoritma seperti *Naïve Bayes*, *K-Nearest Neighbor* (KNN), dan *Support Vector Machine* (SVM) menjadi relevan dalam konteks ini karena kemampuan mereka untuk melakukan klasifikasi yang akurat terhadap data kesehatan.

Penelitian yang telah dilakukan oleh Syahrial et al. dalam '*Stunting Classification in Children's Measurement Data Using Machine Learning Models*' menggunakan metode SVM-RBF menunjukkan hasil akurasi sebesar 78% (Syahrial et al., 2022). Penelitian lain, '*Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor dengan Feature Selection Backward Elimination*,' menunjukkan bahwa algoritma *K-Nearest Neighbor* dengan nilai $k=5$ menghasilkan rata-rata akurasi 91,90% dengan 9 atribut. Dengan penambahan *Backward Elimination*, akurasi meningkat menjadi 92,20% dengan 8 atribut (Lonang & Normawati, 2022). Penelitian lainnya dalam '*Klasifikasi Penentuan Stunting Menggunakan Metode Naïve Bayes (Studi Kasus: Desa Letekonda Selatan)*' menunjukkan bahwa metode *Naïve Bayes* mencapai akurasi rata-rata 92% dalam setiap iterasinya (Malo & Janga, 2023).

Berdasarkan dari penelitian terdahulu yang memiliki hasil akurasi yang tinggi, ketiga algoritma tersebut dipilih karena karakteristik unik yang dapat berkontribusi pada keefektifan klasifikasi *stunting*. *Naïve Bayes*, misalnya, terkenal karena kesederhanaannya dan efisiensinya dalam bekerja dengan dataset yang berukuran besar dan memiliki banyak fitur. Algoritma ini sering digunakan dalam masalah klasifikasi yang melibatkan probabilitas antar variabel (Setiawan et al., 2021).

K-Nearest Neighbor dipilih karena fleksibilitasnya dalam menangani dataset yang non-linear, serta kemampuannya untuk memberikan hasil yang presisi ketika data memiliki struktur yang cukup

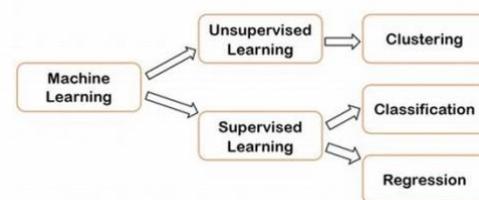
kompleks. Algoritma ini mudah diterapkan dan dapat dioptimalkan melalui metode *feature selection* untuk mengurangi dimensi data tanpa mengorbankan akurasi (Cahyanti et al., 2021).

Support Vector Machine (SVM) dipilih karena kemampuannya yang kuat dalam menangani data dengan margin yang jelas antara dua kelas. SVM sangat efektif dalam kasus-kasus di mana terdapat garis batas yang tegas antara kelompok data, seperti dalam klasifikasi status gizi anak. Algoritma ini juga mampu menghindari *overfitting*, terutama dengan kernel yang sesuai, seperti RBF (Radial Basis Function) (Nalatissifa et al., 2021).

Penelitian ini bertujuan untuk membandingkan ketiga algoritma tersebut *Naïve Bayes*, *K-Nearest Neighbor*, dan *Support Vector Machine* untuk menentukan algoritma mana yang menghasilkan akurasi tertinggi dalam klasifikasi status stunting. Diharapkan, hasil penelitian ini dapat memberikan rekomendasi praktis bagi pemangku kebijakan dalam pengambilan keputusan berbasis data untuk menekan angka stunting di Indonesia. Berdasarkan pemaparan penelitian sebelumnya terkait pada klasifikasi *stunting*, maka penelitian ini memiliki tujuan untuk memperoleh pengambilan keputusan risiko berstatus *stunting* pada anak usia bawah lima tahun menggunakan metode klasifikasi yang mampu menentukan secara lebih tepat.

KAJIAN LITERATUR

Machine Learning



Gambar 1. *Machine Learning*

Machine learning adalah bagian integral dari kecerdasan buatan yang melibatkan pembangunan model matematika berdasarkan data latih untuk membuat prediksi atau keputusan otomatis. Menggunakan algoritma dan teknik yang sesuai, *machine learning* mengenali pola kompleks dalam data, meningkatkan kinerja sistem tanpa pemrograman manual berkelanjutan. Gambar 1 menjelaskan terdapat dua jenis utama *machine learning*: *supervised learning*, yang menggunakan dataset berlabel untuk menghasilkan prediksi akurat, dan *unsupervised learning*, yang bekerja dengan dataset tanpa label untuk menemukan pola atau struktur baru. Kedua paradigma

ini penting dalam pengembangan kecerdasan buatan dan analisis data (Nisa Sofia Amriza et al., 2021).

kemudahan implementasi dalam berbagai konteks aplikatif.

Supervised Learning

Metode *supervised learning* menggunakan data berlabel untuk membentuk model yang dapat memprediksi atau mengklasifikasi data baru. Supervised learning terbagi menjadi dua masalah utama: klasifikasi, di mana variabel *output* berupa kategori seperti warna atau status kesehatan, dan regresi, di mana variabel *output* berupa nilai kontinu seperti mata uang atau berat objek (Roihan et al., 2020).

Naive Bayes

Naive Bayes classifier adalah metode klasifikasi berdasarkan *Teorema Bayes* yang mengasumsikan independensi antar fitur dalam dataset (Setiawan et al., 2021). Metode ini menggunakan probabilitas untuk memprediksi kejadian di masa depan berdasarkan informasi dari kejadian-kejadian sebelumnya, dengan asumsi bahwa setiap fitur dalam dataset bersifat independen satu sama lain.

Klasifikasi

Klasifikasi dalam analisis data adalah proses mengekstrak pola untuk meramalkan label kelas pada dataset tertentu (Abadi et al., 2019). Proses ini terdiri dari dua tahap: pembelajaran, di mana model dibentuk dengan mempelajari pola dari data pelatihan, dan pengklasifikasian, di mana model diterapkan untuk memprediksi label kelas dari data baru (Adriyendi & Melia, 2020). Klasifikasi tidak hanya meramalkan kelas tetapi juga memahami pola tersembunyi dalam data untuk membentuk model prediktif yang optimal (Syahrir et al., 2022).

Confusion matrix

Penilaian dilaksanakan untuk menilai sejauh mana kinerja model yang telah diterapkan. *Confusion matrix* dapat digunakan sebagai alat evaluasi pada tahap ini. Pendekatan ini menghasilkan sejumlah nilai yang berfungsi sebagai indikator evaluasi kinerja model, seperti *F1-skor*, akurasi, presisi, dan *recall* (Purwono et al., 2021). Tabel 1 menjelaskan perhitungan *confusion matrix*. TP adalah jumlah tepat positif, TN adalah jumlah tepat negatif, FP adalah jumlah negatif keliru, dan FN adalah jumlah positif keliru oleh model klasifikasi.

Support Vector Machine (SVM)

Support Vector Machine (SVM) diperkenalkan oleh Vapnik pada tahun 1992 sebagai teknik *machine learning* yang mengadopsi prinsip *Structural Risk Minimization* (SRM) untuk menemukan hyperplane optimal yang memisahkan dua kelas berbeda (Maulidah et al., 2021). SVM menggunakan fungsi linier dalam ruang fitur berdimensi tinggi dan dapat diterapkan pada klasifikasi dan regresi (Saputra et al., 2021). SVM berfokus pada pemisahan linier dalam klasifikasi, namun dapat menangani masalah non-linier dengan konsep kernel, mencari *hyperplane* yang memaksimalkan margin antara dua kelas untuk menciptakan model prediktif yang efisien dan akurat (Nalatissifa et al., 2021).

Tabel 1. Confusion Matrix

Kelas Aktual	Kelas Hasil Prediksi	
	+	-
+	True Positive (TP)	True Negative (TN)
-	False Positive (FP)	False Negative (FN)

Akurasi adalah persentase data uji yang diklasifikasikan dengan benar (Rarasati, 2020). Perhitungan nilai akurasi sebuah model dapat dilihat pada persamaan (1).

$$Accuracy = \frac{TP + TN}{Total\ Data} \tag{1}$$

K-Nearest Neighbor (K-NN)

Metode *K-Nearest Neighbor* (K-NN) adalah pendekatan klasifikasi yang menggunakan data terdekat berdasarkan jarak dalam dataset. Penentuan nilai k sangat memengaruhi proses pengklasifikasian, karena nilai ini menentukan jumlah tetangga terdekat yang dipertimbangkan untuk menentukan kelas observasi (Cahyanti et al., 2021). K-NN memiliki kelebihan seperti ketahanan terhadap dataset besar dan

Presisi mencerminkan jumlah data pada kategori positif yang diklasifikasikan dengan benar, dibagi oleh total data yang diklasifikasikan sebagai positif (Pratiwi et al., 2021). Presisi dapat dihitung dengan persamaan (2).

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

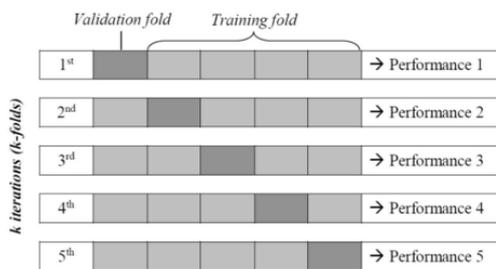
Sementara itu, *recall* menunjukkan persentase data dari kategori positif yang diklasifikasikan dengan benar oleh sistem (Religia et al., 2021). *Recall* dapat dihitung dengan formula (3).

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Skor F1 adalah nilai yang mewakili rata-rata harmonik antara presisi dan recall, yang menciptakan metrik evaluasi yang menggabungkan kedua aspek tersebut dengan mempertimbangkan keseimbangan antara keduanya (Hidayat & Astuti, 2020). *F1-Score* dicari menggunakan persamaan (4).

$$F1 - Score = \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

Cross-validation



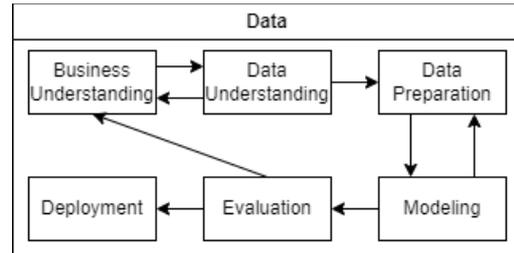
Gambar 2. Cross-validation

Cross-validation, atau estimasi rotasi, adalah teknik validasi model yang menilai kemampuan generalisasi hasil analisis ke kumpulan data independen, penting dalam *machine learning* dan statistika. Teknik umum yang digunakan adalah *k-fold cross-validation*, di mana dataset dibagi menjadi K lipatan yang berukuran sama. Setiap lipatan bergantian digunakan sebagai data uji, sementara lipatan lainnya digunakan sebagai data pelatihan. Proses ini diulang sebanyak K kali sehingga setiap lipatan digunakan sekali sebagai data uji (Azizah et al., 2022). Misalnya pada Gambar 2, dengan K=5, dataset dibagi menjadi lima bagian yang berukuran sama, model dilatih pada empat bagian dan diuji pada bagian kelima. Proses ini diulang lima kali dan hasilnya dirata-rata untuk memberikan estimasi performa model yang lebih akurat (Tembusai et al., 2021).

METODE PENELITIAN

Metode penelitian ini menggunakan CRISP-DM (*Cross-Industry Standard Process Model for Data Mining*), yang mencakup enam tahap: *Business*

Understanding, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment* (Nurkhaliza & Wijayanto, 2022). Proses ini menggambarkan alur data mining secara sistematis, seperti yang ditunjukkan pada Gambar 3.



Gambar 3. CRISP-DM

Penjelasan mengenai fase pada model CRISP-DM melibatkan beberapa langkah penting. Tahap *Business Understanding* adalah awal dari proses, di mana fokusnya adalah memahami substansi kegiatan *data mining*, menetapkan tujuan penelitian, dan menerjemahkan tujuan tersebut menjadi objektif *data mining* yang jelas. Sementara itu, pada Tahap *Data Understanding*, dilakukan pengumpulan dan analisis data untuk memahami karakteristiknya serta mengidentifikasi masalah terkait data yang akan dihadapi. Selanjutnya, Tahap *Data Preparation* melibatkan upaya mempersiapkan struktur basis data agar memfasilitasi proses mining dengan efektif (Lundén et al., 2023).

Pada Tahap *Modeling*, fokusnya beralih ke penentuan teknik, algoritma, dan parameter yang optimal untuk mencapai tujuan data mining. Tahap *Evaluation* berkaitan dengan interpretasi hasil dari pemodelan, dengan tujuan untuk menyelaraskan model dengan sasaran yang ingin dicapai. Terakhir, pada Tahap *Deployment*, laporan pengetahuan yang diperoleh dari evaluasi proses *data mining* disusun. Keseluruhan proses ini memberikan kerangka kerja yang terstruktur dan sistematis untuk memandu para profesional *data science* dalam menjalankan proyek penambangan data dengan tujuan bisnis yang jelas (Nalatissifa et al., 2021).

Penelitian ini menggunakan metode CRISP-DM dengan enam tahap:

1. *Business Understanding*: Fokus pada membandingkan kinerja tiga algoritma klasifikasi—SVM, KNN, dan *Naive Bayes Classifier* (NBC)—dalam mengidentifikasi risiko *stunting* pada balita.
2. *Data Understanding*: Data awal dari Puskesmas Jatilawang terdiri dari 1281 balita, disaring menjadi 523 data balita yang lengkap.

3. *Data Preparation*: Melibatkan penanganan missing data, label *encoding*, *feature selection*, dan pembagian data (20% untuk *testing*, 80% untuk *training*) untuk memastikan keakuratan dan keandalan analisis.
4. *Modeling*: Model klasifikasi dibangun menggunakan dataset pelatihan dengan algoritma SVM, KNN, dan NBC, kemudian diuji pada dataset uji.
5. *Evaluation*: Menggunakan *confusion matrix* untuk membandingkan akurasi ketiga model dan mengidentifikasi model terbaik.
6. *Deployment*: Mengembangkan website sederhana untuk menyajikan fungsionalitas identifikasi dan pengelolaan risiko *stunting* pada balita, berdasarkan hasil evaluasi algoritma.

Variabel penelitian ini adalah untuk memahami, mengukur, atau mengidentifikasi hubungan antara fenomena *stunting*. Dengan merinci variabel penelitian, peneliti dapat memperoleh pemahaman yang lebih akurat tentang faktor-faktor yang mempengaruhi fenomena *stunting*, yang pada gilirannya dapat membantu dalam pengembangan kebijakan, praktik, atau teori yang lebih baik. Variabel pada penelitian ini dapat dilihat pada Tabel 2.

Tabel 2. Variabel Penelitian

Variabel	Keterangan
Jenis Kelamin	Jenis Kelamin Anak (0:Perempuan, 1:Laki-laki)
Umur	Umur Bayi Bawah Lima Tahun (0-60 Bulan)
Berat Badan	Berat Badan terakhir diukur (kg)
Panjang Badan	Panjang Badan terakhir diukur (cm)
Lingkar Lengan	Lingkar Lengan terakhir diukur (cm)
<i>z-Score</i> BB/U	Nilai <i>z-Score</i> Berat Badan Sekarang dibandingkan Umur
BB/U	Berat Badan Sekarang dibandingkan Umur (0:Gizi Buruk, 1:Gizi Kurang, 2:Gizi Normal, 3:Gizi Lebih, 4:Gizi Berlebih)
<i>z-Score</i> BB/PB	Nilai <i>z-Score</i> Berat Badan Sekarang dibandingkan Panjang Badan Sekarang
BB/PB	Berat Badan Sekarang dibandingkan Panjang Badan Sekarang (0:Sangat Kurus, 1:Kurus, 2:Normal,

Variabel	Keterangan
	3:Gemuk, 4:Sangat Gemuk)
<i>z-Score</i> PB/U	Nilai <i>z-Score</i> Panjang Badan Sekarang dibandingkan Umur
PB/U	Panjang Badan Sekarang dibandingkan Umur (0:Sangat Pendek, 1:Pendek, 2:Normal, 3:Tinggi, 4:Sangat Tinggi)
Pendidikan Ayah	Pendidikan terakhir Ayah (0:Tidak Sekolah, 1:SD, 2:SMP, 3:SMA, 4:Sarjana)
Pendidikan Ibu	Pendidikan terakhir Ibu (0:Tidak Sekolah, 1:SD, 2:SMP, 3:SMA, 4:Sarjana)
Riwayat Hamil	Riwayat Ibu saat hamil (0:Tidak Ada Riwayat, 1: Ada Riwayat)
Hasil	Hasil dari Status Berisiko <i>Stunting</i> (0:Tidak Berisiko <i>Stunting</i> , 1:Berisiko <i>Stunting</i>)

HASIL DAN PEMBAHASAN

Bussines Understanding

Stunting pada anak-anak, yang disebabkan oleh kekurangan gizi selama masa pertumbuhan, memiliki dampak luas di bidang kesehatan dan ekonomi. Dari segi kesehatan, *stunting* meningkatkan risiko penyakit infeksi dan kronis serta memengaruhi perkembangan kognitif anak. Dampak ekonominya termasuk pengurangan produktivitas di masa depan dan peningkatan beban sistem kesehatan. Secara sosial, *stunting* memperdalam kesenjangan ekonomi dan pendidikan, mempengaruhi pertumbuhan ekonomi suatu negara. Oleh karena itu, penanganan *stunting* adalah masalah kesehatan masyarakat dan pengembangan manusia secara menyeluruh.

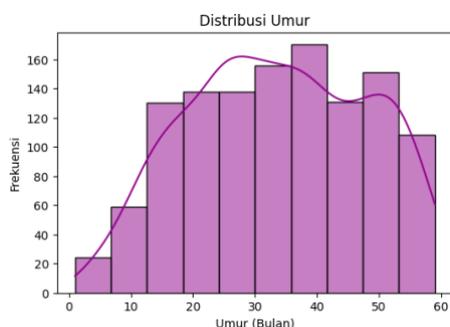
Data dari Badan Pusat Statistik Provinsi Jawa Tengah 2021 menunjukkan bahwa Banyumas memerlukan perhatian khusus dalam layanan kesehatan anak. Dengan 24.381 kelahiran bayi, Banyumas berada di peringkat kelima di Provinsi Jawa Tengah. Jumlah bayi dengan berat lahir rendah sebanyak 1.629 dan kasus gizi buruk sebanyak 2.985 menunjukkan masalah yang perlu diatasi dalam gizi dan perawatan anak. Di Kecamatan Jatilawang, Kabupaten Banyumas, data dari BPS 2022 menunjukkan 853 kelahiran bayi, dengan 246 bayi mengalami gizi buruk dan 61 bayi lahir dengan berat rendah. Data ini menyoroti kebutuhan mendesak untuk fokus pada perawatan maternal dan anak, termasuk pencegahan gizi buruk

dan intervensi dini untuk mengurangi risiko kelahiran bayi dengan berat rendah.

Penelitian ini bertujuan mengembangkan pengambilan keputusan risiko *stunting* pada anak usia di bawah lima tahun menggunakan metode klasifikasi yang lebih akurat. Dengan membandingkan tiga algoritma klasifikasi—*Naïve Bayes*, *K-Nearest Neighbor*, dan *Support Vector Machine*—penelitian ini bertujuan menentukan algoritma dengan nilai akurasi tertinggi. Hasil penelitian diharapkan memberikan manfaat signifikan dalam penerapan dan peningkatan upaya terkait, sehingga dapat menurunkan tingkat risiko *stunting* pada anak dan meningkatkan kesehatan serta kesejahteraan anak-anak di wilayah tersebut.

Data Understanding

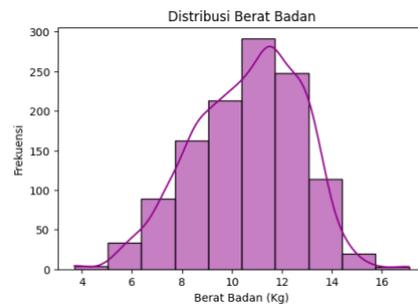
Pada penelitian ini, tahap awal yang krusial adalah melakukan *Data Understanding* dan *Exploratory Data Analysis* (EDA) untuk mendapatkan pemahaman yang mendalam tentang data yang akan digunakan. Melalui analisis ini, penulis akan menjelajahi setiap variabel dalam dataset, mengidentifikasi tipe data, dan memahami distribusi nilai. Selain itu, penulis juga akan melihat hubungan antar-variabel melalui perhitungan korelasi dan visualisasi grafik. Dalam proses ini, penulis akan mencari pola, anomali, atau kekurangan data yang perlu diperhatikan. *Data understanding* dan EDA ini penting untuk memvalidasi asumsi yang mendasari model klasifikasi yang akan dibangun. Dengan memahami karakteristik data secara menyeluruh, penulis dapat membuat keputusan yang lebih baik dalam pemilihan fitur, penanganan data yang hilang, dan pemilihan metode klasifikasi yang sesuai. Selanjutnya, penulis akan menggunakan berbagai diagram dan visualisasi data seperti *histogram*, *scatter plot*, dan *box plot* untuk mempresentasikan temuan secara jelas dan terstruktur.



Gambar 4. Distribusi Umur

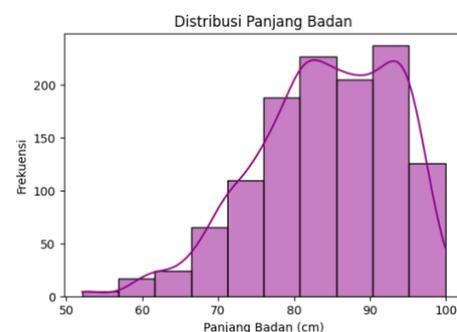
Gambar 4 menunjukkan informasi distribusi umur balita pada dataset. Umur balita pada dataset ini

dimulai dari 0 bulan hingga 60 bulan. Melihat dari diagram yang ditunjukkan memberikan informasi umur terendah pada data adalah 1 bulan sedangkan umur tertinggi adalah 59 bulan. Pada data tersebut juga menunjukkan rata-rata umur balita yang terdapat di Puskesmas Jatilawang adalah 33 bulan. Umur 40 bulan pada data menjadi jumlah balita paling banyak, seangkan umur 1 bulan adalah data paling sedikit pada dataset di Puskesmas Jatilawang.



Gambar 5. Distribusi Berat Badan

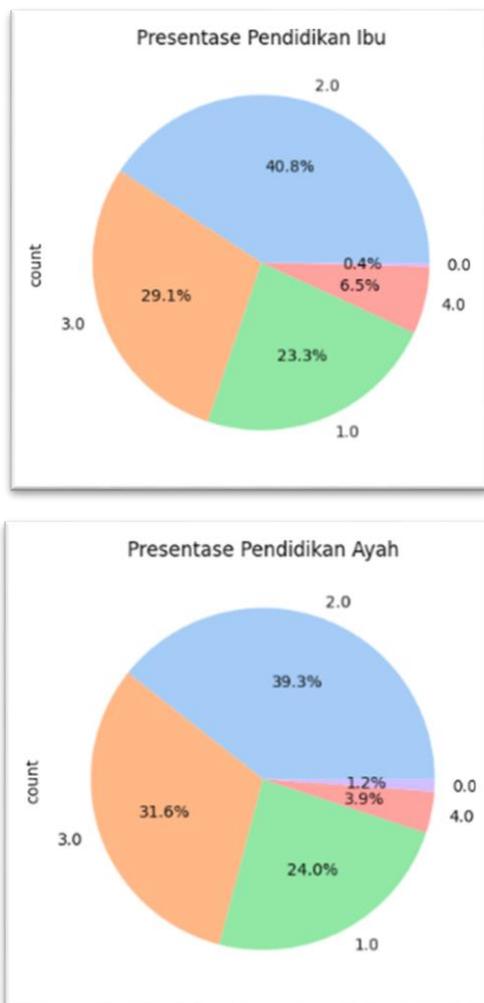
Gambar 5 menunjukkan gambar diagram distribusi berat badan balita, dapat dilihat bahwa distribusi berat badan balita cenderung mengikuti pola normal. Pada sumbu x, terdapat rentang berat badan balita yang dibagi menjadi beberapa kategori. Pada sumbu y, terdapat frekuensi atau jumlah balita yang masuk dalam setiap kategori berat badan. Dari gambar tersebut, terlihat bahwa sebagian besar balita memiliki berat badan yang berada di sekitar nilai rata-rata, ditandai dengan puncak yang lebih tinggi pada kategori tersebut. Terdapat juga beberapa balita yang memiliki berat badan di luar rentang normal, ditandai dengan ekor yang lebih panjang pada kedua ujung diagram. Informasi ini memberikan gambaran tentang sebaran berat badan balita dalam populasi yang diamati. Dengan memahami distribusi berat badan balita, kita dapat mengidentifikasi potensi masalah gizi dan memperoleh wawasan yang berguna dalam upaya meningkatkan kesehatan dan pertumbuhan balita secara keseluruhan.



Gambar 6. Distribusi Panjang Badan Balita

Pada gambar 6 menunjukkan gambar diagram distribusi panjang badan balita, terlihat bahwa persebaran panjang badan balita memiliki pola yang menunjukkan persebaran negatif atau lebih tinggi di bagian kanan. Pada sumbu x, terdapat rentang panjang badan balita yang dibagi menjadi beberapa kategori. Pada sumbu y, terdapat frekuensi atau jumlah balita yang masuk dalam setiap kategori panjang badan. Dalam gambar tersebut, terlihat bahwa sebagian besar balita memiliki panjang badan yang berada di sekitar nilai rata-rata atau di bagian kiri diagram. Namun, terdapat juga sejumlah balita yang memiliki panjang badan yang lebih tinggi, ditandai dengan ekor yang lebih panjang di bagian kanan diagram. Informasi ini memberikan gambaran tentang distribusi panjang badan balita dalam populasi yang diamati.

24,0% lulusan SD, 39,3% lulusan SMP, 31,6% lulusan SMA, dan 3,9% lulus perguruan tinggi. Sedangkan untuk pendidikan ibu, terlihat 0,4% tidak sekolah, 23,3% lulusan SD, 40,8% lulusan SMP, 29,1% lulusan SMA, dan 6,5% lulus perguruan tinggi. Dari gambar tersebut, dapat dilihat bahwa sebagian besar ayah dan ibu balita memiliki pendidikan setidaknya hingga tingkat SMP. Persentase tertinggi terletak pada kategori lulus SMP, diikuti oleh lulus SMA. Terdapat perbedaan antara tingkat pendidikan ayah dan ibu balita, dengan persentase ibu yang lulus perguruan tinggi sedikit lebih tinggi daripada ayah. Informasi ini memberikan gambaran tentang tingkat pendidikan orang tua balita dalam populasi yang diamati, yang dapat menjadi faktor penting dalam memahami konteks sosial dan pendidikan yang mempengaruhi pertumbuhan dan perkembangan balita.



Gambar 7. Pendidikan Orang Tua

Pada gambar 7 menampilkan diagram lingkaran pendidikan ayah dan ibu balita, terlihat persebaran pendidikan orang tua dalam persentase. Untuk pendidikan ayah, terdapat 1,2% yang tidak sekolah,

Data Preparation

Dalam penelitian ini, peneliti berfokus pada klasifikasi risiko *stunting* pada balita, memulai dengan tahap data preparation yang kritis. Tahap pertama adalah menangani missing data atau data yang tidak lengkap, yang penting untuk memastikan kebersihan dan keakuratan dataset. Peneliti menggunakan teknik menghapus data yang kosong karena terlalu banyak data yang kosong.

Setelah memastikan dataset lengkap, peneliti melanjutkan dengan proses *feature selection*. Tujuannya adalah memilih fitur-fitur yang paling relevan dan signifikan dalam memprediksi risiko *stunting*. Dengan menggunakan metode statistik atau algoritma tertentu, peneliti dapat mengidentifikasi fitur-fitur yang tidak penting atau *redundan*. Proses ini membantu meningkatkan efisiensi model, mengurangi dimensionalitas dataset, meningkatkan kecepatan komputasi, dan mengurangi risiko overfitting.

Langkah terakhir dalam *data preparation* adalah *data splitting*, dengan pembagian 20:80. Sebanyak 20% dari dataset digunakan sebagai data *testing*, sedangkan 80% digunakan sebagai data *training*. Pembagian ini penting untuk memastikan evaluasi model yang akurat. Data *training* digunakan untuk melatih model, sedangkan data *testing* digunakan untuk mengukur performa model dalam memprediksi risiko *stunting*. Dengan pembagian ini, peneliti dapat memastikan bahwa model yang dihasilkan memiliki kemampuan generalisasi yang baik dan memberikan prediksi yang akurat pada data baru.

Modeling

SVM (*Support Vector Machine*)

Tabel 3. Cross-validation SVM

Metric	CV1	CV2	CV3	CV4	CV5	AVG	AVG ACC
Acc	0.97	0.88	0.83	0.91	0.89	0.90	
Prec	0.96	1	0.87	1	0.86	0.94	0.90
Rec	0.93	0.58	0.48	0.68	0.71	0.68	
F1	0.94	0.73	0.62	0.81	0.78	0.78	

Pada penelitian ini, model SVM (*Support Vector Machine*) digunakan untuk memprediksi risiko *stunting* pada balita. Model ini dipilih karena kemampuannya yang baik dalam memisahkan data dengan margin yang maksimal. Tabel menunjukkan hasil dari *k-fold cross validation* dengan nilai $k=10$ menunjukkan bahwa model SVM memiliki akurasi yang konsisten dan cukup tinggi. Rata-rata nilai akurasi yang dicapai adalah 0.90, dengan nilai *precision*, *recall*, dan *f1-score* masing-masing sebesar 0.95, 0.69, dan 0.79.

Naïve Bayes

Tabel 4. Cross-Validation Naïve Bayes

Metric	CV1	CV2	CV3	CV4	CV5	AVG	AVG ACC
Acc	0.90	0.85	0.81	0.88	0.80	0.85	
Prec	0.75	0.76	0.70	0.86	0.59	0.74	0.85
Rec	0.96	0.68	0.58	0.68	0.92	0.77	
F1	0.84	0.72	0.64	0.76	0.72	0.74	

Model *Naive Bayes* juga diaplikasikan dalam penelitian ini untuk memprediksi risiko *stunting* pada balita. Model ini terkenal dengan pendekatannya yang sederhana namun efektif, terutama dalam menangani dataset yang berukuran besar dan beragam. Hasil *k-fold cross validation* dengan $k=5$ menunjukkan bahwa model *Naive Bayes* memiliki akurasi yang memadai, dengan rata-rata nilai akurasi sebesar 0.85. Nilai *precision*, *recall*, dan *f1-score* yang dihasilkan oleh model ini adalah 0.74, 0.77, dan 0.74 secara berturut-turut.

K-Nearest Neighbors (KNN)

Tabel 5. Cross-Validation KNN

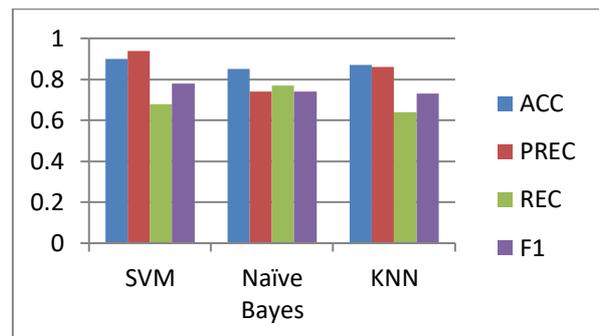
Metric	CV1	CV2	CV3	CV4	CV5	AVG	AVG ACC
Acc	0.92	0.89	0.81	0.91	0.79	0.87	
Prec	0.68	1	0.77	1	0.65	0.86	0.87
Rec	0.86	0.62	0.48	0.68	0.53	0.64	
F1	0.86	0.76	0.59	0.81	0.58	0.73	

Dalam penelitian ini, model *K-Nearest Neighbors (KNN)* digunakan untuk memprediksi risiko *stunting* pada balita dengan mempertimbangkan kedekatan antar data. KNN dipilih karena keunggulannya dalam menangani data *non-linear* dan sederhana dalam implementasinya. Hasil dari *k-fold*

cross validation dengan nilai $k=5$ menunjukkan bahwa model KNN memiliki performa yang baik, dengan rata-rata nilai akurasi sebesar 0.87. Selain itu, nilai *precision*, *recall*, dan *f1-score* yang dicapai adalah 0.86, 0.64, dan 0.73.

Evaluation

Berdasarkan hasil *modeling* tiga model yang digunakan untuk memprediksi risiko *stunting* pada balita, evaluasi dari SVM menunjukkan performa terbaik dengan akurasi rata-rata 0.90, *precision* 0.95, *recall* 0.69, dan *f1-score* 0.79. Model ini menunjukkan kemampuannya dalam memisahkan data dengan margin yang maksimal. Namun, *recall* yang relatif lebih rendah menunjukkan adanya potensi kasus positif yang terlewatkan. Model *Naive Bayes* juga menunjukkan performa yang cukup baik dengan akurasi rata-rata 0.85, *precision* 0.74, *recall* 0.77, dan *f1-score* 0.74, meskipun terdapat beberapa variabilitas dalam hasil lipatan yang menunjukkan ketidakkonsistenan dalam subset data tertentu. Model KNN, dengan akurasi rata-rata 0.87, *precision* 0.86, *recall* 0.64, dan *f1-score* 0.73, menunjukkan kemampuan yang baik dalam menangani variasi data, namun juga memiliki kekurangan pada aspek *recall*.



Gambar 8. Diagram Evaluasi

Stabilitas performa model SVM terlihat dari distribusi nilai evaluasi pada setiap lipatan, menunjukkan konsistensi yang baik dalam prediksi. *Naive Bayes*, meskipun sederhana, efektif dalam menangani dataset yang beragam tetapi memerlukan perbaikan untuk meningkatkan konsistensinya. KNN, dengan kemampuan menangani data *non-linear*, menunjukkan variabilitas hasil yang menandakan perlunya *tuning* parameter lebih lanjut. Berdasarkan hasil evaluasi ini, model SVM dapat dianggap sebagai model yang paling andal untuk mendeteksi risiko *stunting* pada balita, mendukung upaya pencegahan yang lebih efektif. Namun, setiap model memiliki kelebihan yang dapat diintegrasikan untuk

menciptakan pendekatan yang lebih komprehensif dalam pencegahan *stunting*.

Deployment

Setelah melalui evaluasi menyeluruh, model SVM terbukti menjadi model terbaik dalam memprediksi risiko *stunting* pada balita dibandingkan dengan KNN dan *Naive Bayes*. Dengan mempertimbangkan hasil akurasi, *precision*, *recall*, dan *f1-score*, model SVM dipilih untuk di-deploy agar dapat digunakan dalam aplikasi nyata. Untuk tujuan ini, peneliti menggunakan *Flask*, sebuah *framework* web ringan yang memungkinkan pembuatan aplikasi web berbasis *Python* dengan cepat dan mudah. Model SVM diintegrasikan ke dalam aplikasi *Flask*, yang menyediakan antarmuka web yang dapat diakses oleh pengguna untuk memprediksi risiko *stunting* berdasarkan input data yang relevan.

Gambar 9. Web Prediksi Risiko *Stunting* pada Balita

Aplikasi web yang dihasilkan memiliki tampilan yang intuitif dan *user-friendly*. Pengguna dapat memasukkan data balita seperti jenis kelamin, umur, berat badan, panjang badan, lingkar lengan, serta informasi tambahan terkait pendidikan orang tua dan kondisi lingkungan. Setelah data dimasukkan, aplikasi akan mengirimkan data tersebut ke model SVM untuk diproses dan mengembalikan prediksi risiko *stunting*. Hasil prediksi ditampilkan secara langsung pada halaman web, memberikan informasi yang jelas dan mudah dipahami. Implementasi ini diharapkan dapat membantu tenaga kesehatan dan pembuat kebijakan dalam mengambil keputusan yang lebih tepat dan cepat dalam upaya pencegahan *stunting*. Gambar 9 di atas ini menunjukkan tampilan antarmuka web yang sudah jadi, menggambarkan kemudahan penggunaan dan efektivitas sistem yang telah dikembangkan.

KESIMPULAN

Penelitian ini bertujuan untuk mengkomparasi tiga algoritma klasifikasi yaitu *Naive Bayes*, *K-Nearest Neighbor* (KNN), dan *Support Vector Machine* (SVM) dalam mengklasifikasikan status berisiko *stunting* pada anak usia bawah lima tahun. Hasil evaluasi menunjukkan bahwa SVM memiliki performa terbaik dibandingkan dengan *Naive Bayes* dan KNN. SVM mencapai rata-rata akurasi 0.90, dengan *precision* 0.95, *recall* 0.69, dan *f1-score* 0.79. Performa yang konsisten dan akurasi yang tinggi menjadikan SVM sebagai model yang paling andal untuk mendeteksi risiko *stunting* pada balita.

Selain itu, implementasi model SVM dalam bentuk aplikasi web menggunakan *Flask* telah dilakukan sebagai langkah lanjutan dari penelitian ini. Aplikasi web tersebut memungkinkan pengguna untuk memasukkan data balita dan mendapatkan prediksi risiko *stunting* secara langsung. Dengan antarmuka yang *user-friendly*, aplikasi ini diharapkan dapat membantu tenaga kesehatan dalam mengidentifikasi balita yang berisiko *stunting* dengan cepat dan akurat. Penggunaan teknologi ini tidak hanya meningkatkan efisiensi kerja, tetapi juga menyediakan alat bantu yang sangat diperlukan dalam upaya pencegahan *stunting*.

DAFTAR PUSTAKA

- Abadi, M. R., Defiyanti, S., & Sulistiowati, N. (2019). Penerapan Algoritma C4.5 Untuk Klasifikasi Keberhasilan Pengiriman Barang. 8(1), 36–43.
- Adriyendi, A., & Melia, Y. (2020). Klasifikasi Menggunakan Naïve Bayes Dan K-Nearest Neighbor Pada Manajemen Layanan Teknologi Informasi. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 2(2), 99–107. <https://doi.org/10.47233/jteksis.v2i2.121>
- Ariyadi. (2020). Klasifikasi Spesies Lebah Berbasis Data Citra Dengan Metode Support Vector Machine. *Jurnal Inovasi Penelitian*, 1(3), 599–597.
- Azizah, N., Nastia, & Sadat, A. (2022). Strategi Dinas Kesehatan Dalam Menekan Laju Penderita *Stunting* Di Kabupaten Buton Selatan. 2(12), 4145–4152.
- Cahyanti, F. L. D., Gata, W., & Sarasati, F. (2021). Implementasi Algoritma Naïve Bayes dan K-Nearest Neighbor Dalam Menentukan Tingkat Keberhasilan Immunotherapy Untuk Pengobatan Penyakit Kanker Kulit. *Jurnal Ilmiah Universitas Batanghari Jambi*, 21(1), 259. <https://doi.org/10.33087/jiubj.v21i1.1189>
- Erik, Rohman, A., Rosyana, A., Rianti, A., Muhaemi, E., Yuni, E. E., Fauziah, F., Rojuli, Abdi, Y., & Huda, N. (2020). *Stunting* pada anak usia dini (Study Kasus di Desa Mirat Kec Lewimunding

- Majalengka). *Jurnal Pengabdian Masyarakat*, 2(1), 24–36.
- Hidayat, R., & Astuti, T. (2020). Diagnosis Preeklamsia pada Ibu Hamil Berdasarkan Algoritme K- Nearest Neighbour. *Jurnal Ilmu-Ilmu Informatika Dan Manajemen STMIK*, 14(2), 106–116.
- Kementerian Kesehatan Republik Indonesia. (2023). *Menteri Kesehatan Republik Indonesia Hasil Survei Status Gizi Indonesia (SSGI) 2022*. 1–7.
- Kementrian Kesehatan RI. (2021). Tujuan Pembangunan Berkelanjutan (TPB)/SDGs. *Kementrian Kesehatan RI*, 6.
- Lonang, S., & Normawati, D. (2022). Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor Dengan Feature Selection Backward Elimination. *Jurnal Media Informatika Budidarma*, 6(1), 49. <https://doi.org/10.30865/mib.v6i1.3312>
- Lundén, N., Bekar, E. T., Skoogh, A., & Bokrantz, J. (2023). Domain Knowledge in CRISP-DM: An Application Case in Manufacturing. *IFAC-PapersOnLine*, 56(2), 7603–7608. <https://doi.org/10.1016/j.ifacol.2023.10.1156>
- Malo, Y., & Janga, A. U. (2023). *Klasifikasi Penentuan Stunting Menggunakan Metode Naïve Bayes (Studi Kasus : Desa Letekonda Selatan)*. 6, 217–226.
- Maulidah, N., Supriyadi, R., Utami, D. Y., Hasan, F. N., Fauzi, A., & Christian, A. (2021). Prediksi Penyakit Diabetes Melitus Menggunakan Metode Support Vector Machine dan Naive Bayes. *Indonesian Journal on Software Engineering (IJSE)*, 7(1), 63–68. <https://doi.org/10.31294/ijse.v7i1.10279>
- Nalatissifa, H., Gata, W., Diantika, S., & Nisa, K. (2021). Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja. *Jurnal Informatika Universitas Pamulang*, 5(4), 578. <https://doi.org/10.32493/informatika.v5i4.7575>
- Nisa Sofia Amriza, R., Supriyadi, D., Jl Panjaitan No, P. DI, Purwokerto Selatan, K., Banyumas, K., & Tengah, J. (2021). Komparasi Metode Machine Learning dan Deep Learning untuk Deteksi Emosi pada Text di Sosial Media. *Jurnal JUPITER*, 13(2), 130–139.
- Nurkhaliza, A. A., & Wijayanto, A. W. (2022). Perbandingan Algoritma Klasifikasi Support Vector Machine dan Random Forest pada Prediksi Status Indeks Mitigasi dan Kesiapsiagaan Bencana (IMKB) Satuan Kerja BPS di Indonesia Tahun 2020. *Maret*, 7(1), 54–59.
- Pratiwi, B. P., Handayani, A. S., & Sarjana, S. (2021). Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi Wsn Menggunakan Confusion Matrix. *Jurnal Informatika Upgris*, 6(2), 66–75. <https://doi.org/10.26877/jiu.v6i2.6552>
- Purwono, Wirasto, A., & Nisa, K. (2021). *Komparasi Algoritma Machine Learning Untuk Klasifikasi Kelompok Obat Comparison of Machine Learning Algorithms for*. 11(2), 196–207.
- Rarasati, D. B. (2020). A Grouping of Song-Lyric Themes Using K-Means Clustering. *JISA(Jurnal Informatika Dan Sains)*, 3(2), 38–41. <https://doi.org/10.31326/jisa.v3i2.658>
- Religia, Y. R., Pranoto, G. T., & Suwancita, I. M. (2021). Analysis of the Use of Particle Swarm Optimization on Naïve Bayes for Classification of Credit Bank Applications. *JISA(Jurnal Informatika Dan Sains)*, 4(2), 133–137. <https://doi.org/10.31326/jisa.v4i2.946>
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 75–82. <https://doi.org/10.31294/ijcit.v5i1.7951>
- Saputra, D., Irmayani, W., Purwaningtiyas, D., & Sidauruk, J. (2021). A Comparative Analysis of C4.5 Classification Algorithm, Naïve Bayes and Support Vector Machine Based on Particle Swarm Optimization (PSO) for Heart Disease Prediction. *International Journal of Advances in Data and Information Systems*, 2(2), 84–95. <https://doi.org/10.25008/ijadis.v2i2.1221>
- Setiawan, E. I., Johanes, S., Hermawan, A. T., & Yamasari, Y. (2021). Deteksi Validitas Berita pada Media Sosial Twitter dengan Algoritma Naive Bayes. *Journal of Intelligent System and Computation*, 3(2), 55–60. <https://doi.org/10.52985/insyst.v3i2.164>
- Sukmawati, Hendrayati, Chaerunimah, & Nurhumairah. (2018). Keterkaitan status gizi ibu hamil serta BBLR dengan stunting pada anak. *Media Gizi Pangan*, 25, 18–24.
- Syahrial, S., Ilham, R., Asikin, Z. F., & Nurdin, St. S. I. (2022). Stunting Classification in Children’s Measurement Data Using Machine Learning Models. *Journal La Multiapp*, 3(2), 52–60. <https://doi.org/10.37899/journallamultiapp.v3i2.614>
- Syahrir, S., Bayulianto, S., Kusumawati, V. F. P. I., Teguh, Prasetyo, M., Amrullah, A., Anggono, A., & Riza Ibnu Adam, M. S. (2022). *Klasifikasi Penyakit Paru-Paru Dengan Algoritma K-NN (K-Nearest Neighbor) Berdasarkan Fitur Tekstur Glcm. 1*.
- Tembusai, Z. R., Mawengkang, H., & Zarlis, M. (2021). *K-Nearest Neighbor with K-Fold Cross Validation and Analytic Hierarchy Process on Data Classification*. 2(1), 1–8. <https://doi.org/10.25008/ijadis.v2i1.1204>