

PERBANDINGAN METODE SUPERVISED MACHINE LEARNING UNTUK PREDIKSI PREVALENSI STUNTING DI PROVINSI JAWA TIMUR

M Syauqi Haris^{*1}, Ahsanun Naseh Khudori², Wahyu Teja Kusuma³

^{1,2,3}Institut Teknologi, Sains, dan Kesehatan RS dr. Soepraoen Kesdam V/Brw, Malang
Email: ¹haris@itsk-soepraoen.ac.id, ²ahsanunnaseh@itsk-soepraoen.ac.id, ³wtkusuma@itsk-soepraoen.ac.id
^{*}Penulis Korespondensi

(Naskah masuk: 03 Desember 2022, diterima untuk diterbitkan: 26 Desember 2022)

Abstrak

Stunting atau kasus balita kerdil/pendek adalah salah satu masalah di bidang kesehatan yang saat ini sedang dihadapi oleh masyarakat Indonesia. Provinsi Jawa Timur memiliki nilai prevalensi *stunting* sebesar 26,8% berdasarkan integrasi data Kementerian Kesehatan dan Badan Pusat Statistik. Nilai tersebut masih tergolong tinggi karena standar minimal yang ditetapkan oleh World Health Organization (WHO) adalah sebesar 20%. Oleh karena itu, penelitian ini bertujuan untuk memberikan kontribusi dalam penyelesaian permasalahan *stunting* di Provinsi Jawa Timur dengan cara menganalisis faktor-faktor yang diprediksi bisa memengaruhi tingkat prevalensi *stunting* berdasarkan data sekunder hasil survei dari beberapa lembaga resmi dan terpercaya di bidang kesehatan yang telah dipublikasikan. *Supervised machine learning* merupakan pendekatan dalam pembuatan kecerdasan buatan (*artificial intelligence*) yang menggunakan data-data berlabel sebagai data latihnya. Pendekatan ini dirasa sangat sesuai digunakan dalam prediksi nilai prevalensi *stunting* pada suatu wilayah berdasarkan data-data lain yang relevan. Penelitian-penelitian sebelumnya tentang prediksi prevalensi *stunting* rata-rata hanya menggunakan salah satu metode *supervised machine learning* saja dan data sekunder yang digunakan hanya bersumber dari salah satu sumber survei saja. Hasil penelitian menunjukkan bahwa faktor-faktor penyebab yang memiliki korelasi tinggi terhadap nilai prevalensi *stunting* bukan hanya Berat Badan Lahir Rendah (BBLR) saja, namun juga Indeks Pembangunan Manusia, sanitasi, dan Indeks Penduduk Miskin. Selain itu, beberapa metode dalam *supervised machine learning* juga dibandingkan yaitu, *linier regression*, *support vector regression*, dan *random forest regression*. Metode *support vector regression* dalam penelitian ini memiliki nilai galat yang lebih rendah yaitu 0,91 untuk MAE dan 1,30 untuk MSE.

Kata kunci: *stunting*, nilai prevalensi *stunting*, jawa timur, machine learning, supervised learning

SUPERVISED MACHINE LEARNING METHODS COMPARISON ON STUNTING PREVALENCE PREDICTION IN EAST JAVA PROVINCE

Abstract

Stunting or the case of stunted/short toddlers is one of the problems in the health sector that is currently being faced by the people of Indonesia. East Java Province has a *stunting* prevalence value of 26.8% based on data integration from the Ministry of Health and the Central Statistics Agency. This value is still relatively high because the minimum standard set by the World Health Organization (WHO) is 20%. Therefore, this study aims to contribute to solving the *stunting* problem in East Java Province by analyzing the factors that are predicted to affect the *stunting* prevalence rate based on published secondary data from surveys from several official and trusted institutions in the health sector. *Supervised machine learning* is an approach in making artificial intelligence that uses labeled data as training data. This approach is considered very suitable to be used in predicting the value of *stunting* prevalence in an area based on other relevant data. Previous studies on predicting the prevalence of *stunting* on average only used one supervised machine learning method and the secondary data used was only sourced from one survey source. The results showed that the causative factors that have a high correlation to the prevalence of *stunting* are not only low birth weight (BBLR), but also the Human Development Index, sanitation, and the Poor Population Index. In addition, several methods in supervised machine learning are also compared, namely, linear regression, support vector regression, and random forest regression. The support vector regression method in this study has a lower error value, namely 0.91 for MAE and 1.30 for MSE.

Keywords: *stunting*, *stunting* prevalence value, east java, machine learning, supervised learning

1. PENDAHULUAN

Kejadian balita kerdil ketika lahir yang biasa disebut sebagai *stunting*, sampai saat ini masih menjadi permasalahan utama kesehatan gizi di Indonesia. Data prevalensi balita *stunting* yang dikumpulkan World Health Organization (WHO), Indonesia termasuk ke dalam negara ketiga dengan prevalensi tertinggi di Regional Asia Tenggara atau South-East Asia Regional (SEAR). Rata-rata prevalensi balita *stunting* di Indonesia tahun 2005-2017 adalah 36,4%. Kejadian balita *stunting* merupakan masalah gizi utama yang dihadapi Indonesia. Berdasarkan data Pemantauan Status Gizi (PSG), Balita *stunting* memiliki prevalensi tertinggi dibandingkan dengan masalah gizi lainnya seperti gizi kurang, kurus, dan gemuk. Prevalensi Balita *stunting* mengalami peningkatan dari tahun 2016 yaitu 27,5% menjadi 29,6% pada tahun 2017 (Kemenkes RI, 2018). Provinsi Jawa Timur sendiri memiliki nilai prevalensi *stunting* sebesar 26,8% berdasarkan integrasi data Kementerian Kesehatan dan Badan Pusat Statistik pada tahun 2019. Nilai tersebut masih tergolong tinggi karena standar minimal yang ditetapkan oleh World Health Organization (WHO) adalah sebesar 20%.

Permasalahan *stunting* tidak dapat dipandang remeh. *Stunting* Balita menimbulkan dampak yang luar biasa pada anak, selain menyebabkan terhambatnya pertumbuhan fisik dan juga kerentanan oleh penyakit, *stunting* juga mengakibatkan pada sisi perkembangan kognitif. Kejadian *stunting* pada anak pula dapat menyebabkan gangguan degeneratif ketika menyerang gangguan metabolik pada anak. Penderita *stunting* mengalami pertumbuhan dengan banyak risiko yang meliputi obesitas, penyakit jantung koroner, tekanan darah tinggi, tulang keropos, menurunnya produktivitas penderita, sehingga dampaknya bukan hanya pada sisi individu namun jangka panjangnya berdampak pada pertumbuhan suatu negara. Permasalahan *stunting* diperkirakan berdampak pada hilangnya 2-3% Produk Domestik Bruto (PDB) pada tiap negara setiap tahunnya (Galasso & Wagstaff, 2019).

Melihat dari dampak yang dapat dihasilkan oleh permasalahan *stunting* ini maka diperlukannya suatu prediksi jangka panjang dalam memperkirakan kondisi *stunting* di masa yang akan datang sehingga dapat mencegah penyakit *stunting* tersebut. Prediksi penyakit dapat dilakukan dengan menggunakan bantuan *machine learning*. Prediksi yang dibangun menggunakan *machine learning*, berdasar dari data-data masa lalu yang diolah menjadi pola-pola data yang dapat dijadikan prediksi hal yang akan datang (Kusuma, 2020).

Pembelajaran mesin (*machine learning*) adalah bagian dari kecerdasan artifisial (*artificial intelligence*) yang mampu menghasilkan sistem cerdas komputer tanpa penentuan aturan (rule) oleh

manusia secara langsung, akan tetapi sistem dibuat agar mengenali pola kumpulan data yang digunakan sebagai data latih sehingga menghasilkan sebuah model yang bisa digunakan untuk melakukan prediksi suatu nilai (regresi) atau kelompok data (klasifikasi) (Retnoningsih & Pramudita, 2020). Pembelajaran mesin mungkin untuk dilakukan jika tersedia data sebagai masukan (*input*) untuk dilakukan analisis terhadap kumpulan data yang besar sehingga bisa ditemukan pola tertentu. Dalam pembelajaran mesin dikenal data latih (*training data*) dan data uji (*testing data*), data latih untuk melatih algoritme dan data uji untuk mengetahui performa dari *machine learning* yang telah dilatih yaitu ketika menemukan data baru yang belum pernah diberikan dalam data latih (Kusuma, 2020).

2. METODE PENELITIAN

Tahapan yang harus dilakukan pada penelitian ini sebanyak 6 (enam) tahap, yaitu kajian pustaka, pengumpulan data, pra-pemrosesan (*preprocessing*) data, pembagian data, pemodelan data dan evaluasi model sebagaimana diilustrasikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1. Kajian Pustaka

Pada tahapan ini dilakukan inventarisasi informasi dari beberapa sumber sebagai penunjang penelitian ini dilakukan. Artikel ilmiah yang digunakan sebagai dasar kerangka berfikir penelitian ini meliputi artikel ilmiah seputar *stunting* pada balita, metode *machine learning*, prediksi *stunting*, penggunaan metode *support vector regression*, penggunaan metode *linier regression*, penggunaan metode *random forest regression* dan penggunaan metode *feature selection*.

2.2. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data prevalensi *stunting* di Jawa Timur dan data hasil survei di bidang kesehatan dari beberapa lembaga yaitu:

1. Studi Status Gizi Balita Indonesia (SSGBI) tahun 2019 (Kemenkes RI & BPS, 2019),

2. Survei Sosial Ekonomi Nasional (SUSENAS) tahun 2019 (Kemenkes RI & BPS, 2019),
3. Laporan Direktorat Statistik Kesejahteraan Rakyat Badan Pusat Statistik tahun 2019 (Subdirektorat Statistik Rumah Tangga, 2019),
4. Statistik Perumahan dan Pemukiman tahun 2019 (Subdirektorat Statistik Kesehatan dan Perumahan, 2019),
5. Indeks Pembangunan Manusia tahun 2019 (Subdirektorat Analisis Statistik, 2019).

Adapun data-data yang bisa dikompilasi dan digunakan sebagai atribut dalam pemodelan prediksi prevalensi *stunting* sebagai berikut:

1. Persentase Ibu Hamil K4,
2. Persentase Ibu Hamil Mendapat Tablet Tambah Darah (90 Tablet),
3. Persentase Persalinan Ditolong Nakes,
4. Persentase Ibu Nifas Mendapat Vitamin A,
5. Persentase Berat Badan Lahir Rendah (BBLR),
6. Persentase Bayi Baru Lahir Mendapat IMD,
7. Persentase Baduta diberi ASI,
8. Rata-rata Lama Pemberian ASI (bulan),
9. Rata-rata Lama Pemberian ASI dengan Makanan Pendamping,
10. Persentase Balita Imunisasi Lengkap,
11. Persentase Bayi 6-11 Bulan Vitamin A,
12. Persentase Bayi 12-59 Bulan Vitamin A,
13. Persentase Bayi 6-59 Bulan Vitamin A,
14. Persentase Cakupan Pelayanan Kesehatan Balita,
15. Rata-rata Jiwa/Rumah Tangga,
16. Persentase RT Memiliki Akses Sanitasi Layak,
17. Persentase RT Memiliki Akses Sumber Air Minum Layak,
18. Indeks Pembangunan Manusia (IPM) 2019,
19. Pengeluaran Riil per Kapita,
20. dan Indeks Penduduk Miskin.

Data yang telah dikompilasi disimpan dalam bentuk berkas Comma Separated Value (CSV) agar bisa diproses menggunakan *supervised machine learning tools* yaitu Python menggunakan Jupiter Notebook.

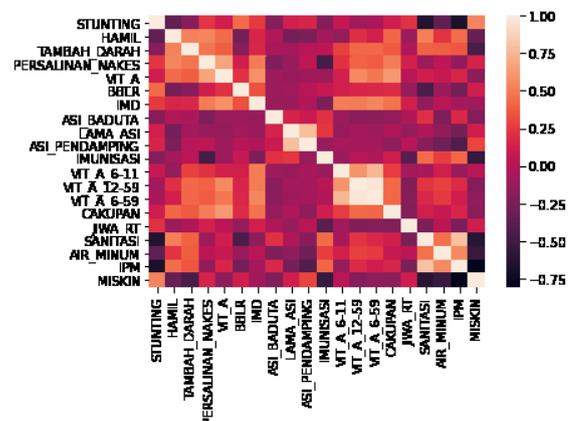
2.3. Pra-Pemrosesan Data

Pada tahapan ini, dilakukan pemrosesan data sebelum dimodelkan. *Preprocessing* data dilakukan dengan melakukan data *cleaning* atau melakukan pengecekan data, apakah terdapat data yang tidak valid dan data yang memiliki nilai kosong. Setelah data diperiksa dengan menggunakan teknik *data cleaning* dilakukan *feature selection* yang berguna untuk mengetahui fitur atau atribut yang paling dominan yang akan digunakan dalam pemodelan data.

Pada tahapan pembersihan data (*data cleaning*), dilakukan pengecekan data untuk mengetahui apakah terdapat ada dengan nilai kosong pada setiap data kabupaten. Teknik *data cleaning* merupakan suatu teknik yang digunakan dalam menangani data kosong

atau data dengan nilai yang tidak lengkap. Teknik *Data Cleaning* juga dapat digunakan dalam proses *enrichment*, artinya sebuah langkah untuk “memperkaya” data pada sebuah *dataset* (Jasmir 2016). Proses ini dilakukan dengan cara mencari ada atau tidaknya nilai yang hilang (*null*) pada *dataset* yang digunakan.

Pada tahapan seleksi fitur (*feature selection*) dilakukan pengukuran dari setiap fitur guna untuk mengetahui fitur yang dominan yang akan dipakai dalam pemodelan data. Pada penelitian ini *feature selection* dilakukan dengan cara menghitung korelasi antar fitur. Korelasi fitur ditentukan oleh nilai korelasi variabel *independent*-nya dengan variabel *dependent*-nya. Fitur yang dipilih adalah variabel dengan nilai korelasi sedang dan tinggi baik itu positif (> 0.2) maupun negatif (< -0.2). Ilustrasi korelasi fitur bisa dilihat pada *heatmap* pada Gambar 2.



Gambar 2. Heatmap Korelasi Antar Variabel

2.4. Pembagian Data

Pada tahapan selanjutnya setelah *preprocessing* data maka dilakukan proses pembagian data. Data dibagi menjadi data latih dan data uji. Data latih digunakan dalam pemodelan *machine learning* sedangkan data uji digunakan untuk mencoba hasil dari pembelajaran *machine learning*. Data dibagi dengan rasio 80% untuk data latih dan 20% digunakan sebagai data uji. Pembagian data dengan rasio 4:1 diharapkan semakin banyak data yang dilatih menggunakan *machine learning* maka akan menghasilkan akurasi sistem yang valid dengan nilai yang tinggi.

2.5. Pemodelan Data

2.5.1. Linier Regression

Linear regression atau regresi linier merupakan salah satu algoritma *machine learning*. Regresi linier digunakan untuk memodelkan data dengan mencari hubungan antara satu atau lebih variabel bebas terhadap variabel respon. Regresi linier dapat dibedakan menjadi dua, yaitu regresi linier sederhana dan regresi linier berganda. Perbedaan diantara keduanya dilihat dari jumlah variabel bebas yang dihitung hubungannya dengan variabel responnya

(Indarwati et al., 2019). Dikatakan regresi linier sederhana jika terdapat satu variabel bebas, namun jika variabel bebas terdapat lebih dari satu maka disebut regresi linier berganda.

2.5.2. Support Vector Regression (SVR)

Algoritma SVR merupakan turunan dari metode Support Vector Machine (SVM) yang sering digunakan dalam proses klasifikasi. Implementasi metode SVR diterapkan dalam permasalahan regresi. Metode SVM melakukan klasifikasi pada data yang memiliki nilai bulat, sedangkan metode SVR melakukan pemecahan kasus regresi pada hasil yang memiliki luaran bilangan riil (Purnama et al., 2020).

Konsep metode Support Vector Regression menjalankan sebuah estimasi proses regresi dengan menghitung nilai fungsi linier dengan menggunakan variabel *lagrange non-negative multiplier*. Pemecahan dari permasalahan ini secara manual didapatkan dengan menjalankan pemrograman kuadratik. Permukaan aproksimasi optimal menggunakan formulasi yang telah dimodifikasi setelah memperpanjang *Support Vector Regression* menjadi non linier (Hidayat, 2013).

2.5.3. Random Forest Regression

Random forest merupakan salah satu algoritma *supervised learning*. *Random forest* diimplementasikan untuk mengklasifikasikan sebuah kelompok dari pohon regresi, pemodelan menggunakan fungsi acak dalam proses pengacakan pohon keputusan. Setelah sejumlah besar *tree* telah di-generate, maka *tree* di-voting guna memilih kelas yang paling populer. Tahapan *voting tree* secara keseluruhan ini dijelaskan sebagai *random forest* (Ramadhan et al., 2019).

2.6. Evaluasi Model

Tahap evaluasi adalah tahap yang dilakukan untuk melihat seberapa baik performa algoritma klasifikasi yang digunakan dalam penelitian. Tolok ukur yang digunakan untuk mengukur performa tersebut adalah nilai galat (*error*) dalam Mean Absolute Error (MAE) dan Mean Squared Error (MSE).

3. PENELITIAN TERKAIT

Telah ada beberapa penelitian sebelumnya terkait prediksi *stunting* dengan metode *machine learning*. Penelitian sebelumnya, menerapkan metode *support vector regression* dalam memprediksi *stunting*. *Support vector regression* diterapkan dengan menghitung semua *kernel*. *Kernel* yang dihitung meliputi kernel linier, *polynomial* dan RBF. Model terbaik dengan menggunakan *kernel* RBF, yang memiliki nilai korelasi tertinggi dan galat terendah di tiap jenis *Z-score* yang diprediksi (Hidayat, 2013).

Pada Penelitian lain, juga melakukan prediksi *stunting* menggunakan salah satu algoritma *machine learning*. Penelitian prediksi *stunting* dilakukan dengan menggunakan bantuan algoritma *random forest*. Pada penelitian ini data *stunting* difokuskan pada bulan Januari hingga Oktober tahun 2020 di Kecamatan Pitu, Kabupaten Ngawi, Jawa Timur. Hasil yang didapatkan pada penelitian ini berupa sebuah pembelajaran dari data dan dilakukan pemodelan data dengan membagi data latih dan data uji sehingga mampu menghasilkan sebuah model yang dapat memprediksi tingkat *stunting* pada balita (Perdana et al., 2021).

Penelitian sebelumnya juga melakukan prediksi *stunting* dengan membandingkan dua metode *machine learning*. Penelitian ini berjudul "Prediksi Tingkat Risiko Kejadian *Stunting* Pada Anak Menggunakan Algoritma *K-Nearest Neighbor* dan *Random Forest*" dengan melakukan komparasi dua algoritma yaitu *k-nearest neighbour* dengan algoritma *random forest*. Penelitian ini menggunakan *feature selection* dalam menentukan atribut pada data yang akan dimodelkan menggunakan dua algoritma tersebut. Hasil dari *feature selection* didapatkan atribut yang berpengaruh atau memiliki korelasi adalah jenis kelamin anak, berat badan lahir anak, pekerjaan orang tua, tinggi ayah, pendidikan terakhir ibu, tinggi badan lahir anak, tinggi ibu, pendidikan terakhir ayah, usia, berat badan anak sekarang (Labolo et al., 2022).

Pada penelitian lain juga serupa, membahas terkait *stunting* dengan bantuan *machine learning*, namun penelitian ini berfokus pada pemetaan faktor-faktor yang mempengaruhi risiko penyakit *stunting* pada bayi lima tahun. Penelitian dengan fokus pemetaan ini mengimplementasikan sebuah metode, yaitu Geographically Weighted Regression (GWR). Tujuan dari penelitian ini adalah memahami faktor eksternal yang mempengaruhi atau menyebabkan kasus *stunting* di Indonesia. Hasil yang didapatkan pada penelitian ini dengan menghitung faktor menggunakan regresi linier adalah faktor persentase bayi yang mendapatkan imunisasi lengkap dan presentasi ibu hamil risiko kekurangan energi kronik merupakan faktor yang signifikan dalam penyebab *stunting* pada balita (Cholid et al., 2019).

Berdasarkan beberapa penelitian terdahulu yang masih belum banyak melakukan penelitian seputar *stunting* dengan memperhatikan pra pemrosesan data, dan juga belum banyak yang membandingkan beberapa metode regresi, maka pada penelitian ini prediksi *stunting* pada balita menggunakan 3 (tiga) metode yaitu *Support Vector Regression*, *Random Forest Regression* dan *Linier Regression*. Pada proses pra pemodelan data maka dilakukan *feature selection* dan *data cleaning* untuk memastikan data yang digunakan dalam pemodelan merupakan data yang valid dan memiliki nilai

atribut yang signifikan dalam memprediksi *stunting* pada balita.

4. HASIL DAN PEMBAHASAN

Proses seleksi fitur pada penelitian ini dilakukan dengan cara menghitung nilai korelasi antara data variabel yang diperoleh dari hasil pengumpulan data terhadap nilai prevalensi *stunting*. Seleksi variabel ini diharapkan mampu mengeliminasi variabel-variabel yang memiliki nilai korelasi rendah (nilai antara -0,20 sampai dengan 0,20) antara nilai variabel terhadap nilai prevalensi *stunting* sehingga bisa diperoleh faktor-faktor yang memiliki korelasi sedang dan tinggi terhadap nilai prevalensi *stunting* yang disajikan pada Tabel 1.

Tabel 1. Nilai Korelasi Variabel Terpilih terhadap Nilai Prevalensi *Stunting*

Variabel	Nilai Korelasi
Persalinan Nakes	0,24
Asi Baduta	-0,21
IPM	-0,67
Miskin	0,54
Hamil	-0,35
Tambah Darah	-0,20
BBLR	0,44
IMD	0,26
Imunisasi	0,20
Jiwa RT	0,22
Sanitasi	-0,64
Air Minum	-0,36

Penelitian komparasi algoritma *supervised learning* untuk memprediksi *stunting* pada balita menunjukkan bahwa algoritma Support Vector Regression (SVR) memiliki nilai MAE 0,93 dan MSE 1,34. Pada algoritma Random Forest Regression diperoleh hasil prediksi dengan nilai MAE 1,07 dan MSE 1,81. Adapun Linier Regression memiliki nilai 0,93 untuk MAE dan 1,34 untuk MSE sebagaimana bisa dilihat pada Tabel 2.

Tabel 2. Perbandingan Nilai Galat Prediksi Algoritma *Supervised Machine learning*

Metode	MAE	MSE
<i>Linier Regression</i>	0,93	1,34
<i>Support Vector Regression</i>	0,91	1,30
<i>Random Forest Regression</i>	1,07	1,81

5. KESIMPULAN

Data-data yang dikumpulkan dari beberapa hasil survei pada tahun yang sama yaitu SSGBI 2019, SUSENANS 2019, Statistik Kesejahteraan Rakyat 2019, Statistik Perumahan dan Pemukiman 2019, dan IPM 2019 menghasilkan 20 variabel yang diduga menjadi faktor penyebab kejadian bayi kerdil atau *stunting*. 20 variabel tersebut diseleksi menggunakan analisis korelasi terhadap nilai prevalensi *stunting*

pada tahun yang sama sehingga dapat diperoleh 12 variabel yang diduga sebagai faktor penyebab yang memengaruhi nilai prevalensi *stunting* pada suatu daerah. Berdasarkan nilai korelasi yang diperoleh, faktor utama yang memiliki pengaruh tertinggi adalah Indeks Pembangunan Manusia dengan nilai korelasi negatif yaitu -0,67. Faktor berikutnya yang juga memiliki pengaruh tinggi adalah sanitasi dengan nilai korelasi negatif yaitu -0,64 dan Indeks Penduduk Miskin dengan nilai korelasi sebesar 0,54. Hal ini merupakan hal yang baru dikarenakan rata-rata nilai prevalensi *stunting* hanya diasosiasikan dengan faktor Berat Badan Lahir Rendah (BBLR) saja.

Penelitian komparasi algoritma supervised learning untuk memprediksi *stunting* pada balita dapat disimpulkan bahwa algoritma Support Vector Regression (SVR) memiliki pemodelan yang terbaik dikarenakan memiliki nilai MAE dan MSE lebih kecil dari pada algoritma Random Forest Regression dan Linier Regression sebesar 0,91 untuk nilai MAE dan 1,30 untuk nilai MSE.

Penelitian selanjutnya diharapkan bisa menggunakan data dari seluruh kabupaten/kota yang ada di Indonesia. Dengan banyaknya data latih diharapkan model yang dihasilkan bisa lebih akurat dan lebih aplikatif untuk diterapkan sebagai dasar pengambilan kebijakan pemerintah nasional dalam menentukan langkah-langkah preventif guna menekan nilai prevalensi *stunting* di Negara Indonesia.

DAFTAR PUSTAKA

- CHOLID, F., TRISHNANTI, D., & AZIES, H. AL. 2019. Pemetaan Faktor-Faktor yang Mempengaruhi Stunting pada Balita dengan Geographically Weighted Regression (GWR) Mapping of The Factors Affecting Stunting on Toddlers with Geographically Weighted Regression (GWR). *SEMNAkes 2019 "Improving the Quality of Health Tharough Advances in Research of Health Sciences," March 2020,* 156–165. <https://doi.org/10.17605/OSF.IO/9MZU7>
- GALASSO, E., & WAGSTAFF, A. 2019. The Aggregate Income Losses from Childhood Stunting and The Returns to A Nutrition Intervention Aimed at Reducing Stunting. *Economics and Human Biology*, 34(August), 225–238. <https://doi.org/10.1016/j.ehb.2019.01.010>
- HIDAYAT, R. 2013. Sistem Prediksi Status Gizi Balita Dengan Menggunakan Support Vector Regression. *Departemen Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Pertanian Bogor.*
- INDARWATI, T., IRAWATI, T., & RIMAWATI, E. 2019. Penggunaan Metode Linear Regression untuk Prediksi Penjualan Smartphone. *Jurnal Teknologi Informasi Dan Komunikasi (TIKomSiN)*, 6(2).

- <https://doi.org/10.30646/tikomsin.v6i2.369>
KEMENKES RI. 2018. Buletin Stunting. *Kementerian Kesehatan RI*, 301(5), 1163–1178.
- KEMENKES RI, & BPS. 2019. Laporan Pelaksanaan Integrasi Susenas Maret 2019 dan SSGBI Tahun 2019. In *Badan Pusat Statistik, Jakarta - Indonesia*.
- KUSUMA, P. D. 2020. *Machine Learning Teori, Program, Dan Studi Kasus*. Deepublish.
- LABOLO, A. Y., MOODUTO, S., BODE, A., & DRAJANA, I. C. R. 2022. Penerapan Algoritma Support Vector Machine dan K-Nearest Neighbor Menggunakan Feature Selection Backward Elimination untuk Prediksi Status Penderita Stunting pada Balita. *Jurnal Tecnosienza*, 6(2), 374–388. <https://doi.org/10.51158/tecnoscienza.v6i2.713>
- PERDANA, A. Y., LATUCONSINA, R., & DINIMAHARAWATI, A. 2021. Prediksi Stunting Pada Balita Dengan Algoritma Random Forest. *ISSN: 2355-9365 e-Proceeding of Engineering: Vol.8, No.5 Oktober 2021*, 8(5), 6650–6656.
- PURNAMA, J. J., NAWAWI, H. M., ROSYIDA, S., RIDWANSYAH, & RISANDAR. 2020. Klasifikasi Mahasiswa Her Berbasis Algoritma Svm Dan Decision Classification of Her Students Based on SVM Algorithm and. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(6), 1253–1260. <https://doi.org/10.25126/jtiik.202073080>
- RAMADHAN, A., SUSETYO, B., & INDAHAWATI. 2019. Penerapan Metode Klasifikasi Random Forest dalam Mengidentifikasi Faktor Penting Penilaian Mutu Pendidikan. *Jurnal Pendidikan dan Kebudayaan*, 4(2), 169–182. <https://doi.org/10.24832/jpnk.v4i2.1327>
- RETNONINGSIH, E., & PRAMUDITA, R. 2020. Mengenal Machine Learning dengan Teknik Supervised dan Unsupervised Learning Menggunakan Python. *BINA INSANI ICT JOURNAL*, 7(2), 156. <https://doi.org/10.51211/biict.v7i2.1422>
- SUBDIREKTORAT ANALISIS STATISTIK, BPS. 2019. Indeks Pembangunan Manusia (IPM) 2019. *Berita Resmi Statistik*, 9, 1–8. <https://www.bps.go.id/pressrelease/2020/02/17/1670/indeks-pembangunan-manusia--ipm--indonesia-pada-tahun-2019-mencapai-71-92.html>
- SUBDIREKTORAT STATISTIK KESEHATAN DAN PERUMAHAN, BPS. 2019. *Statistik Perumahan dan Pemukiman 2019*. <https://www.bps.go.id/publication/2020/08/31/6a9e70d6154fde75499239e6/statistik-perumahan-dan-permukiman-2019.html>
- SUBDIREKTORAT STATISTIK RUMAH TANGGA, BPS. 2019. *Statistik Kesejahteraan Rakyat 2019*. <https://www.bps.go.id/publication/2019/11/22/1dfd4ad6cb598cd011b500f7/statistik-kesejahteraan-rakyat-2019.html>