

Aplikasi Berbasis Streamlit untuk Memprediksi Stunting pada Balita Menggunakan Metode *Random Forest* dengan Pengaruh Status Gizi dan Karakteristik Pemberian ASI Eksklusif

Osa Nastiyar Maulani
S1 Data Sains
Univeristas Telkom
Bandung, Indonesia
osanastiyar@student.telkomuniversity.a
c.id

Putri Cahyaning Wulandari
S1 Data Sains
Univeristas Telkom
Bandung, Indonesia
putricahyaning@student.telkomuniversi
ty.ac.id

Putri Eka Lestari
S1 Data Sains
Universitas Telkom
Bandung, Indonesia
putriekalestari@student.telkomuniversit
y.ac.id

Abstract— Di Indonesia, *stunting* pada masa kanak-kanak merupakan masalah kesehatan serius yang mempengaruhi pertumbuhan fisik dan kognitif. Dalam penelitian ini, kami mengembangkan aplikasi berbasis Streamlit yang menggunakan metode *random forest* untuk memprediksi risiko *stunting* pada anak kecil. Faktor-faktor seperti status gizi dan ASI eksklusif dianalisis menggunakan dataset Kaggle. Tujuan dari aplikasi ini adalah untuk mendukung intervensi dini dan memberikan perkiraan yang akurat untuk mendukung pengambilan kebijakan. Dengan menerapkan model hutan acak, aplikasi ini memberikan solusi praktis kepada dokter dan pembuat kebijakan dalam deteksi dini dan pengobatan keterbelakangan pertumbuhan.

Keywords— *stunting*, *Random Forest*, aplikasi prediksi, status gizi, ASI eksklusif

I. PENDAHULUAN

Stunting pada anak balita merupakan masalah kesehatan serius yang mempengaruhi pertumbuhan fisik dan kognitif. Di Indonesia, kemungkinan seorang anak mengalami *stunting* masih tinggi, mencapai 16% pada tahun 2023. Penelitian ini mengembangkan aplikasi berbasis Streamlit untuk memprediksi risiko *stunting* menggunakan metode *Random Forest*, dengan menganalisis faktor status gizi dan pemberian ASI eksklusif. Dataset yang digunakan diambil dari Kaggle dan mencakup variabel seperti jenis kelamin, usia, berat badan, panjang badan, dan status pemberian ASI eksklusif. Dengan *Random Forest*, aplikasi ini diharapkan memberikan prediksi yang akurat dan membantu praktisi kesehatan serta pembuat kebijakan dalam intervensi dini untuk mengurangi *stunting*.

II. STUDI LITERATUR

A. *Stunting* Pada Balita

Menurut Organisasi Kesehatan Dunia atau World Health Organization (WHO) *Stunting* adalah gangguan pertumbuhan dan perkembangan yang dialami anak akibat gizi buruk, infeksi berulang, dan stimulasi psikososial yang tidak memadai. Anak-anak

didefinisikan sebagai *stunting* jika tinggi badan terhadap usia mereka lebih dari dua standar deviasi di bawah median Standar Pertumbuhan Anak WHO. Sedangkan pengertian *stunting* menurut Kementerian Kesehatan (Kemenkes) adalah anak balita dengan nilai z-scorenya kurang dari -2.00 SD/standar deviasi (*stunted*) dan kurang dari -3.00 SD (*severely stunted*).

Pemerintah Indonesia telah mengambil berbagai langkah untuk mengatasi *stunting*, termasuk program perbaikan gizi, penyuluhan kesehatan, dan peningkatan akses ke air bersih dan sanitasi. Faktor-faktor ini termasuk faktor ekonomi serta kurangnya pengetahuan tentang asupan gizi yang cukup.

Dalam beberapa tahun terakhir, pengurangan angka *stunting* telah menjadi prioritas nasional. Berbagai inisiatif telah diluncurkan untuk mencapai target penurunan *stunting* yang signifikan. Misalnya, melalui program intervensi gizi yang ditargetkan dan kolaborasi antar lembaga pemerintah dan organisasi non-pemerintah internasional.

B. Teknologi Prediksi dalam Kesehatan

Eric Topol menjelaskan bahwa kemajuan kecerdasan buatan dan pembelajaran mesin telah membuka jalan bagi teknologi prediktif yang dapat mengubah paradigma medis dari reaktif menjadi proaktif (Topol, 2019). Menurut Johns Hopkins Medicine, teknologi prediktif digunakan untuk menganalisis data genomik dan klinis untuk memprediksi perkembangan penyakit dan mengoptimalkan intervensi untuk setiap pasien (Johns Hopkins Medicine, 2020). Foster Provost dan Tom Fawcett menambahkan bahwa teknologi prediktif dalam layanan kesehatan menggunakan teknik *teks mining* dan pembelajaran mesin untuk memprediksi kejadian kesehatan di masa depan dan mengidentifikasi pasien berisiko tinggi (Provost & Fawcett, 2013).

C. *Random Forest*

Metode *random forest* adalah pengembangan dari metode CART, yaitu dengan menerapkan metode

bootstrap aggregating (bagging) dan *random feature selection* (Breiman 2001). Dalam *random forest*, banyak pohon ditumbuhkan sehingga terbentuk hutan (*forest*), kemudian analisis dilakukan pada kumpulan pohon tersebut.

D. Machine Learning

Machine learning dapat didefinisikan sebagai aplikasi komputer dan algoritma matematika yang diadopsi dengan cara pembelajaran yang berasal dari data dan menghasilkan prediksi di masa yang akan datang (Goldberg & Holland, 1988). Adapun proses pembelajaran yang dimaksud adalah suatu usaha dalam memperoleh kecerdasan yang melalui dua tahap antara lain latihan (training) dan pengujian (testing) (Huang, Zhu, & Siew, 2006).

Bidang machine learning berkaitan dengan pertanyaan tentang bagaimana membangun program komputer agar meningkat secara otomatis dengan berdasar dari pengalaman (Mitchell, 1997).

E. Streamlit

Menurut artikel yang berasal dari web Analytics Vidhya Streamlit adalah kerangka kerja python sumber terbuka untuk membangun aplikasi web untuk Pembelajaran Mesin dan Ilmu Data. Kami dapat langsung mengembangkan aplikasi web dan menerapkannya dengan mudah menggunakan Streamlit. Streamlit memungkinkan Anda menulis aplikasi dengan cara yang sama seperti Anda menulis kode python. Streamlit memudahkan pengerjaan loop interaktif pengkodean dan melihat hasil di aplikasi web.

III. METODE PENELITIAN

A. Dataset

Pemilihan dataset dalam proyek prediksi merupakan tahapan yang sangat penting. Dataset yang dipilih tentunya harus memiliki keterkaitan terhadap prediksi yang diinginkan. Dalam hal prediksi stunting, dataset yang pilih berasal dari website Kaggle.com dengan judul "Child Growth Prediction". Dataset ini memiliki ukuran sebesar 211KB, dengan 6500 baris data dan 6 atribut yang meliputi:

1. Sex (Jenis Kelamin Bayi)
2. Age (Usia Bayi dalam satuan bulan)
3. Birth Weight (Berat Bayi saat lahir)
4. Birth Length (Panjang Bayi saat lahir)
5. Body Weight (Berat Bayi saat ini)
6. Body Length (Panjang Bayi saat ini)

Selain itu, dataset juga menyertakan informasi tentang apakah bayi tersebut mendapatkan ASI eksklusif atau tidak. Variabel target kita adalah "Stunting". Dataset ini telah dipilih karena mencakup informasi yang relevan dengan prediksi stunting, termasuk faktor-faktor seperti berat badan lahir, panjang badan lahir, dan status gizi bayi saat ini. Dengan menggunakan dataset ini, diharapkan kita dapat mengembangkan model prediksi stunting yang akurat dan berguna bagi upaya-upaya penanggulangan stunting pada anak-anak.

B. Metode Pengerjaan

Dalam Pengerjaannya kami menggunakan pendekatan *Agile*. *Agile* merupakan metodologi pengembangan perangkat lunak yang menekankan proses iterative kelincuhan dan juga peningkatan berkelanjutan. Metode ini sangat cocok untuk proyek ini karena lebih fleksibel, terutama ketika proyek memiliki tingkat ketidakpastian yang tinggi dan membutuhkan perubahan secara cepat. *Agile* memecah proses menjadi langkah atau fase pendek yang dapat disebut juga dengan sprint.

C. Metode Machine Learning

Metode *machine learning* yang digunakan dalam prediksi adalah metode Random Forest. Random Forest merupakan algoritma *machine learning* yang serbaguna dimana model ini menggabungkan beberapa model untuk menghasilkan prediksi yang lebih akurat. Lebih tepatnya, Random Forest adalah algoritma yang menggabungkan output dari beberapa decision tree. Berikut adalah tahapan penerapan Random Forest yang akan diterapkan pada Bahasa pemrograman python

- Import Library: Library yang akan digunakan dalam prediksi Random Forest ini adalah 'RandomForestClassifier' dari pustaka `sklearn.ensemble`
- Inisialisasi Model: model Random Forest diinisialisasi dengan menentukan jumlah pohon keputusan dalam ensemble.
- Pelatihan Model: Setelah model diinisialisasi, langkah selanjutnya adalah melatihnya menggunakan data latih.
- Prediksi: Setelah model dilatih, kita dapat menggunakannya untuk membuat prediksi pada data uji atau data baru.

Pemilihan metode Random Forest untuk prediksi stunting didasarkan pada beberapa keuntungan:

- Akurasi tinggi: Random Forest dapat menghasilkan prediksi yang akurat pada data yang kompleks.
- Mencegah overfitting: Metode ini efektif dalam menangani masalah overfitting pada model machine learning.
- Menangani banyak variabel: Random Forest dapat dengan baik memperhitungkan banyak faktor risiko stunting secara bersamaan (status gizi dan ASI eksklusif).

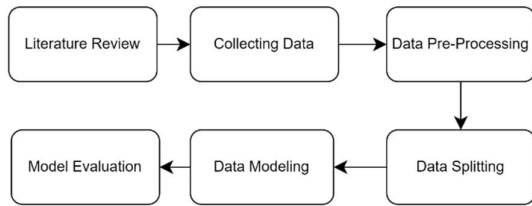
D. Pengembangan Aplikasi Web

Pengembangan aplikasi web menggunakan Streamlit menjadi pilihan yang sangat tepat dalam proyek ini. Streamlit memudahkan pembuatan aplikasi web dengan menggunakan Python, yang mana menjadi bahasa pemrograman yang telah digunakan dalam pengembangan model *Machine Learning*. Dengan Streamlit, proses integrasi antara model *Machine Learning* yang telah dibuat dengan aplikasi web menjadi lebih lancar dan efisien. Hal ini membantu dalam menyajikan prediksi stunting kepada pengguna dengan cara yang lebih intuitif dan mudah diakses. Selain itu, Streamlit juga memberikan fleksibilitas dalam menyesuaikan tampilan dan fitur-fitur tambahan yang

dapat ditambahkan ke dalam aplikasi web sesuai dengan kebutuhan proyek.

E. Tahapan Pengerjaan

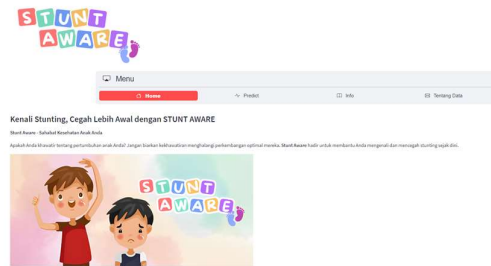
Dibawah ini merupakan gambaran dari pengerjaan proyek



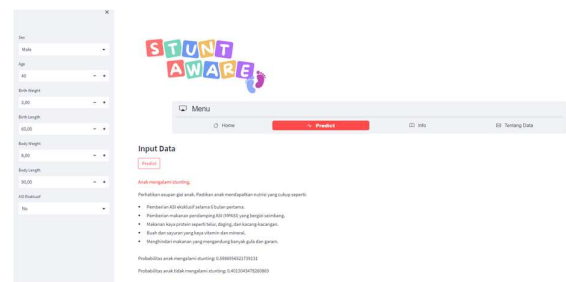
1. Pemahaman Literatur: Pada tahap ini, dilakukan studi terhadap literatur-literatur terkait prediksi stunting. Hal ini bertujuan untuk memahami konsep stunting, faktor-faktor yang mempengaruhinya, serta metode-metode yang telah digunakan dalam penelitian sebelumnya.
2. Pengumpulan dan Pencarian Dataset: Data-data terkait dengan stunting, seperti data gizi, data kesehatan ibu dan anak, serta faktor-faktor lain yang dapat mempengaruhi stunting dikumpulkan. Data tersebut dapat berasal dari berbagai sumber, seperti institusi kesehatan, organisasi internasional, atau penelitian sebelumnya.
3. Tahap Pra-Processing Data: Data yang telah terkumpul kemudian diproses untuk membersihkan, menghilangkan noise, serta menangani missing values. Selain itu, pada tahap ini juga dilakukan konversi data kategorikal menjadi data numerik, serta normalisasi atau standarisasi data numerik.
4. Pemisahan Data: Data yang telah diproses kemudian dibagi menjadi dua bagian: data latih (training data) dan data uji (test data). Pembagian ini bertujuan untuk menguji kinerja model pada data yang belum pernah dilihat sebelumnya.
5. Pemodelan Prediksi: Pada tahap ini, dilakukan pemilihan dan pelatihan model Machine Learning untuk memprediksi stunting berdasarkan data yang telah diproses. Model yang digunakan dapat beragam, salah satunya adalah Random Forest yang telah dijelaskan sebelumnya.
6. Evaluasi Model: Setelah model dilatih, dilakukan evaluasi kinerja model menggunakan data uji. Metrik evaluasi yang umum digunakan adalah akurasi, presisi, recall, dan F1-score. Evaluasi ini membantu untuk menentukan seberapa baik model dapat memprediksi stunting.
7. Integrasi dengan Aplikasi Streamlit: Terakhir, model yang telah dievaluasi kemudian diintegrasikan ke dalam aplikasi web menggunakan Streamlit. Aplikasi ini memungkinkan pengguna untuk melakukan prediksi stunting dengan mudah melalui antarmuka yang ramah pengguna dan interaktif.

IV. HASIL DAN PEMBAHASAN

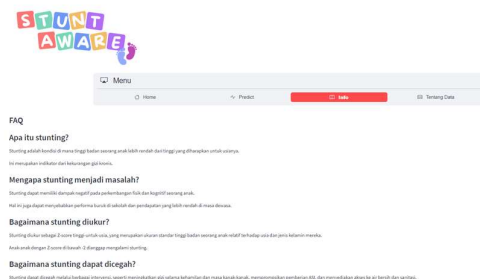
Pengembangan aplikasi prediksi stunting menggunakan metode Random Forest dan Streamlit melibatkan beberapa tahapan, termasuk analisis faktor-faktor utama stunting, pengumpulan dan preprocessing data dari Kaggle, pelatihan model, dan pengembangan aplikasi web. Model Random Forest menunjukkan tingkat akurasi yang tinggi dengan skor 98%. Aplikasi yang dibuat terdiri dari 4 menu yaitu “Home”, “Predict”, “Info”, dan “Tentang Data”. Pada menu “Home” tertera penjelasan singkat terkait dengan aplikasi yang akan digunakan dan juga tertera visualisasi sederhana terkait aplikasi



Selanjutnya pada halaman “Predict” pengguna akan melakukan prediksi dengan cara mengisi bagian kosong pada *sidebar menu* yang ada. Kemudian sistem akan mengeluarkan *output* berupa prediksi dan juga saran untuk pengguna untuk mengurangi stunting



Kemudian Pada menu “info” tertera *FAQ (Frequently Ask Question)* dari hal terkait dengan stunting



Terakhir pada menu “Tentang data” akan tertera EDA dan juga beberapa hal terkait dengan data



The screenshot shows a web application interface with a navigation menu (Home, Prediksi, Info, Stunting Data) and a 'Data Overview' section. Below the menu, there is a table titled 'Yuk kita Lihat datanya' with columns: No, Umur, Berat badan, Tinggi badan, Status gizi, ASI eksklusif, and Stunting. The table contains 10 rows of data.

No	Umur	Berat badan	Tinggi badan	Status gizi	ASI eksklusif	Stunting
1	0	3.0	55	BB	0	1 No
2	0	3.0	55	BB	0	0 No
3	1	4	65	BB	0	0 No
4	1	4	65	BB	0	0 No
5	1	4	65	BB	0	0 No
6	1	4	65	BB	0	0 No
7	1	4	65	BB	0	0 No
8	1	4	65	BB	0	0 No
9	1	4	65	BB	0	0 No
10	1	4	65	BB	0	0 No

Meskipun ada keterbatasan seperti membutuhkan data berkualitas tinggi dan koneksi internet, Streamlit memudahkan integrasi model dengan antarmuka yang mudah dipahami. Aplikasi ini membantu pembuat kebijakan dan profesional kesehatan menemukan stunting lebih awal, yang memungkinkan intervensi yang lebih baik.

V. KESIMPULAN

Aplikasi prediksi stunting yang dibuat dengan metode Random Forest berbasis Streamlit telah menunjukkan hasil yang memuaskan. Model prediksi yang dibuat sangat akurat dalam mengidentifikasi risiko stunting berdasarkan faktor penting seperti status gizi dan pemberian ASI eksklusif. Dengan antarmuka yang mudah digunakan, aplikasi ini membantu pembuat kebijakan dan praktisi kesehatan melakukan intervensi yang tepat dan deteksi dini. Tantangan seperti kualitas data dan kebutuhan untuk akses internet masih perlu diperhatikan. Untuk meningkatkan kegunaan aplikasi, penelitian lebih lanjut disarankan untuk memperluas dataset dan menyertakan fitur seperti saran intervensi gizi.

VI. REFERENSI

[1] Biznetgio.com. (2023, Aug. 11). "Mengenal Agile Development, Metode yang Cocok Diterapkan Developer" [online]. Available: <https://www.biznetgio.com/news/apa-itu-agile-development>.

[2] Atlassian.com. "The Agile Coach" [online]. Available: <https://www.atlassian.com/agile> K. Elissa, "Title of paper if known,"

[3] Revou.co. "Random Forest" [online]. Available: <https://revou.co/kosakata/random-forest>.

[4] M. Syauqi, M. Anshori, A. N. Khudori, "Predicting of Stunting Prevalence In East Java Province With Random Forest Algorithm", ResearchGate, vol. 4(1), pp. 11-13, Feb. 2023.

[5] Ilham. C. A. (2023, Des. 4). "Data Stunting di Indonesia 2022, Apakah menurun di 2023" [online]. Available: <https://tirto.id/data-stunting-di-indonesia-2022-apakah-menurun-di-2023-gSW8>.

[6] Muhtarom. (2023, June). "Stunting Dataset" [online]. Available: <https://www.kaggle.com/datasets/muhtarom/stunting>.

[7] World Health Organization, "Stunting in a nutshell," Nov. 19, 2015. [Online]. Available: <https://www.who.int/news/item/19-11-2015-stunting-in-a-nutshell>. [Accessed: Jun. 9, 2024].

[8] L. P. Salsabila and A. Fadilah, "Analisis Pengaruh Penggunaan E-Commerce terhadap Perkembangan UMKM di Indonesia," IJ-CIT (Indonesian Journal on Computer, Information Technology and Engineering), vol. 6, no. 1, pp. 1-8, 2021. [Online]. Available: <https://ejournal.bsi.ac.id/ejurnal/index.php/ijcit/article/download/7951/pdf>. [Accessed: Jun. 9, 2024].

[9] D. Setiawan, "Analisis Regresi Logistik Biner Pada Penggunaan Software Bajak dan Nonsim," Statistika, vol. 11, no. 2, pp. 1-10, 2011. [Online]. Available: <https://journal.ipb.ac.id/index.php/statistika/article/view/5443/4015>. [Accessed: Jun. 9, 2024].

[10] P. Kumar, "Build Web App Instantly for Machine Learning Using Streamlit," Analytics Vidhya, Jun. 2021. [Online]. Available: <https://www.analyticsvidhya.com/blog/2021/06/build-web-app-instantly-for-machine-learning-using-streamlit/>. [Accessed: Jun. 9, 2024].

[11] E. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nature, vol. 569, pp. 205-211, 2019. [Online]. Available: <https://www.nature.com/articles/s41586-019-1236-0>. [Accessed: Jun. 9, 2024].

[12] Johns Hopkins Medicine, "Predictive Analytics in Medicine," Johns Hopkins Medicine, 2020. [Online]. Available: <https://www.hopkinsmedicine.org/research/labs/predictive-analytics-medicine>. [Accessed: Jun. 9, 2024].

[13] F. Provost and T. Fawcett, Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking, 1st ed. Sebastopol, CA: O'Reilly Media, 2013.