

BAB I

PENDAHULUAN

1.1 Latar Belakang

Menurut Peraturan Presiden No. 72 Tahun 2021 tentang Percepatan Penurunan Stunting, stunting didefinisikan sebagai gangguan tumbuh kembang anak yang disebabkan oleh kekurangan gizi kronis dan infeksi berulang. Kondisi ini ditandai dengan tinggi badan di bawah standar yang ditetapkan oleh Kementerian Kesehatan. Stunting tidak hanya berdampak pada pertumbuhan fisik anak tetapi juga memengaruhi kognisi, kemampuan belajar, dan produktivitas jangka panjang (Kominfo, 2021).

Faktor-faktor yang menyebabkan stunting terbagi menjadi penyebab langsung dan tidak langsung. Penyebab langsung termasuk kurangnya pemberian kolostrum dan ASI eksklusif, pola makan anak yang tidak mencukupi, dan infeksi yang sering dialami anak, yang semuanya memengaruhi status gizi anak dan dapat menyebabkan stunting. Sementara itu, faktor tidak langsung mencakup keterbatasan akses dan ketersediaan pangan serta kondisi sanitasi dan lingkungan yang kurang mendukung kesehatan (Rosha et al., 2020).

WHO (Organisasi Kesehatan Dunia) mengungkapkan bahwa jumlah anak balita yang terkena stunting di dunia pada tahun 2022 adalah 22,3% atau 148,1 juta anak balita ("Level Dan Tren di Anak Kekurangan Gizi," 2023). Menurut laporan Studi Status Gizi Indonesia dari Kementerian Kesehatan, angka stunting di Indonesia mengalami penurunan dari 27,7% pada 2019 menjadi 24,4% pada 2021, dan turun lagi menjadi 21,6% pada 2022, dengan sebagian besar kasus terjadi pada anak usia 3-4 tahun (sekitar 6%). Kemudian pada tahun 2023 angka prevalensi stunting di Sumatera Utara (Sumut) berhasil turun menjadi 18,9%, atau berkurang sekitar 2,2%. Meski menurun, angka ini masih belum memenuhi standar WHO yang menetapkan target di bawah 20%. Karena itu, pemerintah berupaya menurunkan angka stunting lebih lanjut, dengan target 14% pada 2024 (Rokom, 2023).

Berdasarkan Survei Status Gizi Anak Balita Indonesia tahun 2021 (SSGI), prevalensi stunting di Sumatera Utara masih sebesar 25,8%, yang berada di bawah persentase nasional (Fachriza, 2022), Namun, pada tahun 2022, angka stunting di Sumatera Utara menurun menjadi 21,1% (Annur, 2023). Pada tahun 2023, prevalensi stunting di Sumatera Utara berhasil menurun menjadi 18,9%, mengalami penurunan sekitar 2,2% dari tahun sebelumnya, yaitu 2022. Pemerintah Provinsi Sumatera Utara telah menetapkan target untuk lebih lanjut mengurangi prevalensi stunting menjadi 18% pada tahun 2023, dengan mengalokasikan sekitar Rp346 miliar untuk intervensi gizi khusus dan sensitif, serta mengajak semua pemangku kepentingan untuk bekerja sama dalam mengatasi stunting di wilayah tersebut. Melalui kolaborasi, penurunan stunting sebesar 14% pada tahun 2024 dapat dicapai (Pencawan, 2023).

Permasalahan stunting tidak dapat dipandang remeh. Stunting memiliki dampak signifikan baik jangka pendek maupun jangka panjang. Efek jangka pendek meliputi penurunan kekuatan sistem kekebalan tubuh, peningkatan risiko berbagai penyakit, tingkat morbiditas dan mortalitas yang lebih tinggi, serta gangguan kemampuan intelektual dan kognitif pada individu yang mengalami stunting. Sementara itu, efek jangka panjang melibatkan risiko lebih tinggi terhadap penyakit degeneratif pada masa dewasa dan potensi sumber daya manusia yang terhambat (Saputri & Tumangger, 2019). Tidak hanya pada tingkat individu, dampak jangka panjangnya juga mempengaruhi pertumbuhan keseluruhan negara. Masalah stunting diperkirakan menyebabkan kerugian sebesar 2-3% dari Produk Domestik Bruto (PDB) setiap tahun di setiap negara (Galasso & Wagstaff, 2019).

Mempertimbangkan dampak yang dapat ditimbulkan oleh masalah stunting, diperlukan prediksi jangka panjang untuk memperkirakan kondisi stunting di masa depan dan mencegah penyakit terkait stunting tersebut. Prediksi penyakit dapat dilakukan dengan menggunakan Machine Learning. Prediksi yang dibangun menggunakan Machine Learning didasarkan pada data historis yang diproses menjadi pola yang dapat digunakan untuk memprediksi peristiwa di masa depan (Syauqi Haris et al., 2022).

Machine Learning adalah bagian dari kecerdasan buatan yang mampu menciptakan sistem komputer yang cerdas tanpa penentuan aturan manusia secara langsung, melainkan dengan melatih sistem untuk mengenali pola dalam kumpulan data, sehingga menghasilkan model yang dapat digunakan untuk memprediksi nilai (regresi) atau kelompok data (klasifikasi) (Retnoningsih & Pramudita, 2020). Machine learning dapat dilakukan jika terdapat data yang tersedia sebagai input untuk menganalisis dataset besar, sehingga pola-pola khusus dapat ditemukan. Dalam machine learning, terdapat data pelatihan dan data pengujian yang sudah diketahui. Data pelatihan digunakan untuk melatih algoritma, sementara data pengujian digunakan untuk mengevaluasi kinerja model machine learning yang telah dilatih ketika menghadapi data baru yang sebelumnya belum termasuk dalam data pelatihan (Syauqi Haris et al., 2022).

Supervised learning adalah salah satu teknik pembelajaran mesin yang menggunakan dataset berlabel (data pelatihan) untuk melatih mesin. Ini memungkinkan mesin mengidentifikasi label input menggunakan fitur-fitur yang dimilikinya, dan kemudian membuat prediksi atau klasifikasi. Ada beberapa algoritma yang termasuk dalam kategori *Supervised learning* antara lain Regresi Linear, K-nearest neighbor (KNN), *Support Vector Machine* (SVM), *Naive Bayes*, *Random Forest*, *Decision Tree*, dan banyak lainnya (Retnoningsih & Pramudita, 2020).

Support Vector Machine (SVM) adalah metode *supervised machine learning* non-parametrik yang digunakan untuk mengenali dan memprediksi pola dari data pelatihan (Firmansyah et al., 2019). SVM memiliki beberapa keunggulan dibandingkan metode supervised lainnya, seperti kemampuannya mengurangi kesalahan dalam data pelatihan dan dimensi VC (Vapnik-Chervonenkis), dapat menghasilkan solusi meski dengan data pelatihan terbatas, serta mudah dijelaskan dan diterapkan. Namun, kelemahan SVM terletak pada kesulitannya dalam menangani dataset yang sangat besar dan keterbatasannya untuk masalah klasifikasi multikelas. Dalam prosesnya, SVM memerlukan data pelatihan yang membantu mesin membedakan kategori berdasarkan data yang telah ditentukan. Setiap kategori dipisahkan oleh garis pemisah yang disebut hyperplane, dengan jarak

optimal dari pola-pola terdekat yang disebut support vector. Dengan menentukan hyperplane terbaik, metode ini dikenal sebagai Support Vector Machine (Ariyantoni & Rokhmana, 2020). Untuk regresi, SVM dapat dimodifikasi menjadi *Support Vector Regression* (SVR). Dalam penelitian ini, M. Syauqi *et al.* membandingkan beberapa algoritma supervised learning seperti regresi linear, SVR, dan *Random Forest regression* untuk memprediksi stunting pada balita. Hasilnya menunjukkan bahwa SVR memberikan pemodelan terbaik, karena memiliki nilai MAE dan MSE yang paling rendah (Syauqi Haris *et al.*, 2022).

Random Forest (RF) adalah salah satu algoritma *supervised machine learning* yang sering digunakan dalam masalah klasifikasi dan regresi. RF juga disebut metode ensemble, yang merupakan metode untuk meningkatkan akurasi klasifikasi dengan menggabungkan metode klasifikasi (Reinardus & Rosa, 2019). *Random Forest* memiliki beberapa keunggulan, yaitu kemampuan untuk meningkatkan akurasi ketika terdapat data yang hilang, ketahanan terhadap pencilan, dan efisiensi dalam menyimpan data. Selain itu, *Random Forest* memiliki proses seleksi fitur yang dapat mengekstraksi fitur-fitur terbaik, sehingga meningkatkan kinerja dalam model klasifikasi. Dengan seleksi fitur, *Random Forest* dapat bekerja secara efektif dengan parameter-parameter kompleks pada data besar (Supriyadi *et al.*, 2020). Pada penelitian oleh Farhanuddin dkk memprediksi anggaran biaya manajemen proyek system informasi dengan metode machine learning Multiple Linear Regression (MLR) dan *Random Forest Regression* (RFR). Hasil pengujian menunjukkan bahwa model random forest regression memberikan nilai akurasi yang lebih tinggi, mencapai 81,5%, dibandingkan dengan nilai akurasi 62,9% yang diperoleh oleh model multiple linear regression. *Random Forest regression* lebih fleksibel dan dapat menangani hubungan non-linear antara fitur dan variabel target (Farhanuddin *et al.*, 2024).

Algoritma *Decision Tree* bersifat sangat kuat, populer, berbasis logika, dan mudah dipahami. Yang menarik tentang Pohon Keputusan adalah penggunaan struktur pohon untuk mewakili aturan yang terbentuk dari hasil klasifikasi. Dalam pohon tersebut, atribut direpresentasikan oleh simpul, dan kelas direpresentasikan oleh daun. Setiap pohon memiliki akar, yang merupakan simpul di bagian atas.

(Prayoga et al., 2021). Beberapa Kelebihan dari *Decision Tree* yaitu *Decision Tree* hanya membutuhkan sedikit data cleaning untuk memperbaiki kesalahan atau inkonsistensi dibandingkan dengan algoritma lain. Selain itu, sifat hierarkis dari *Decision Tree* memudahkan analisis untuk melihat atribut mana yang paling penting (Revou, 2023). Pada penelitian oleh Kartika dkk memprediksi harga saham di era covid menggunakan metode *Multiple Linear Regression*, *Support Vector Regression*, *Decision Tree Regression* dan *K-Nearest Regression*, dari penelitian tersebut *Decision Tree Regression* memberikan hasil yang kompetitif dibandingkan metode lainnya (Maulida et al., 2020)

Penelitian ini bertujuan untuk membandingkan efektivitas tiga metode *supervised machine learning* dalam konteks prediksi prevalensi stunting berdasarkan data prevalensi dan indikator stunting di Sumatera Utara. Metode yang digunakan adalah *Support Vector Machine (SVM)*, *Decision Tree*, dan *Random Forest*, dipilih karena kemampuan mereka dalam menangani hubungan non-linear yang kompleks dalam data. *Support Vector Regression (SVR)* menggunakan kernel trick untuk memetakan data ke ruang fitur berdimensi lebih tinggi, memungkinkan model ini menangkap pola non-linear dengan akurasi tinggi (Syauqi Haris et al., 2022). *Decision Tree Regression* membagi dataset secara rekursif berdasarkan nilai fitur, membentuk pohon yang mampu merepresentasikan hubungan non-linear antar variable (Maulida et al., 2020). *Random Forest Regression* menggabungkan banyak *Decision Tree* untuk meningkatkan akurasi prediksi dan menangkap berbagai pola non-linear dalam data melalui pendekatan ensemble (Farhanuddin et al., 2024). Kombinasi ketiga metode ini memungkinkan perbandingan yang komprehensif dalam prediksi prevalensi stunting, karena masing-masing memiliki cara unik dalam menangani kerumitan data.

Teknik observasi membutuhkan perhitungan matematis untuk mengukur tingkat kesalahan prediksi. Ada banyak perhitungan yang dapat digunakan untuk menghitung kesalahan prediksi. Salah satu ukuran standar yang digunakan untuk menentukan akurasi prediksi adalah MSE dan MAPE. Dengan fokus pada prevalensi stunting, penelitian ini bertujuan tidak hanya untuk mengidentifikasi

metode terbaik tetapi juga untuk memperdalam pemahaman tentang dampak indikator stunting terhadap prevalensi stunting di Sumatera Utara.

1.2 Rumusan Masalah

Dari uraian diatas maka ditentukan rumusan masalah pada penelitian ini yaitu:

1. Bagaimana perbandingan kinerja antara metode SVM, *Decision Tree*, dan *Random Forest* dalam konteks prediksi prevalensi stunting pada provinsi Sumatera Utara?
2. Apa saja variabel yang memiliki pengaruh yang lebih signifikan terhadap prevalensi stunting?

1.3 Tujuan Penelitian

Dari uraian diatas maka ditentukan tujuan pada penelitian ini adalah sebagai berikut.

1. Mengidentifikasi algoritma yang paling akurat dan efisien di antara ketiga metode yang diuji dalam konteks data stunting.
2. Analisis Pengaruh Indikator Stunting: Memahami dan mengidentifikasi bagaimana indikator-indikator stunting yang berbeda mempengaruhi prevalensi stunting dan kinerja model yang digunakan, sehingga dapat memberikan insight untuk fokus intervensi.

1.4 Manfaat Penelitian

Penelitian ini membawa sejumlah manfaat ke berbagai pihak. Berikut adalah uraian manfaat dari berbagai aspek:

1. Manfaat bagi Penulis:
 - Kontribusi Ilmiah: Melalui penelitian ini, penulis berkontribusi pada literatur ilmiah dengan menyediakan analisis komparatif tentang efektivitas algoritma-algoritma *Machine Learning* dalam sektor kesehatan publik, khususnya dalam prediksi prevalensi stunting.
2. Manfaat bagi Pemerintah:
 - Pengambilan Keputusan yang Berbasis Data: Pemerintah dapat menggunakan hasil penelitian ini untuk mengidentifikasi daerah dengan risiko stunting tinggi dan mengalokasikan sumber daya dengan lebih efektif.

3. Manfaat bagi Akademisi dan Peneliti:

- Riset Lanjutan: Penelitian ini dapat menjadi dasar untuk studi lebih lanjut dalam prediksi dan intervensi stunting atau masalah kesehatan publik lainnya.
- Inovasi Metodologi: Analisis komparatif dari berbagai algoritma *Machine Learning* dapat memperkaya metodologi penelitian di bidang kesehatan publik, menawarkan wawasan baru dan pendekatan yang lebih efisien.

1.5 Batasan Masalah

Berdasarkan latar belakang yang telah disajikan, berikut adalah batasan masalah yang digunakan dalam penelitian ini:

1. Data Penelitian hanya menggunakan data yang tersedia dari Dinas Kesehatan dan BPS Sumatera Utara dalam periode tahun 2021-2023. Data dari sumber lain tidak dimasukkan dalam model.
2. Implementasi metode terbatas pada tiga algoritma regresi yaitu, Support Vector Regression (SVR) menggunakan kernel linear, Decision Tree Regression, dan Random Forest Regression. Untuk SVR, parameter seperti C, epsilon, dan lainnya tidak disesuaikan secara manual, melainkan menggunakan pengaturan default bawaan pustaka Python (scikit-learn).
3. Mengevaluasi hasil kinerja metode menggunakan dua metrik utama yaitu, Mean Squared Error (MSE) dan Mean Absolute Percentage Error (MAPE) untuk mengukur kinerja dan akurasi prediksi.