

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

*Stunting* adalah gangguan pertumbuhan dan perkembangan yang dialami anak-anak karena nutrisi buruk, infeksi berulang, dan stimulasi psikososial yang tidak memadai. Efek *stunting* berdampak signifikan pada perkembangan kognitif anak, seperti penurunan kemampuan berpikir, daya ingat, pemecahan masalah, IQ yang rendah, serta perkembangan motorik yang terlambat. Anak – anak dengan *stunting* juga lebih rentan terhadap masalah kesehatan jangka panjang [1]. Berdasarkan data Survei Status Gizi Indonesia (SSGI), prevalensi balita *stunting* di Indonesia menunjukkan tren penurunan dalam beberapa tahun terakhir. Pada tahun 2019, prevalensi *stunting* tercatat sebesar 27,7%, kemudian menurun menjadi 24,4% pada tahun 2020, dan kembali turun menjadi 21,6% pada tahun 2021 [2]. Meskipun penurunan ini signifikan, angka prevalensi *stunting* di Indonesia masih berada di atas ambang batas yang ditetapkan oleh World Health Organization (WHO), yaitu 20%. Oleh karena itu *stunting* tetap menjadi salah satu masalah kesehatan prioritas nasional yang memerlukan perhatian lebih dalam upaya penanganannya. Saat ini pemerintah sedang mempertimbangkan revisi target penurunan *stunting* tahun 2024 dari 14% menjadi 18% – 19% [3].

Faktor sosioekonomi memainkan peran penting dalam prevalensi *stunting*. Intervensi yang berfokus pada peningkatan kondisi sosioekonomi serta kebijakan publik yang mendukung kesetaraan, sangat diperlukan untuk menurunkan angka *stunting* dan dampaknya dalam jangka panjang. Beberapa penelitian menemukan hubungan erat antara status sosioekonomi dan *stunting*. Penelitian oleh [4] menunjukkan bahwa anak-anak dari keluarga miskin memiliki prevalensi *stunting* sebesar 53,60% jauh lebih tinggi dibandingkan dari keluarga kaya. Penelitian oleh [5] menyebutkan bahwa tingkat pendidikan orang tua juga berpengaruh, di mana ibu tanpa pendidikan formal yang memiliki anak mencapai prevalensi *stunting* sebesar 50,1% dan ayah sebesar 49,4%. Selain itu, penelitian oleh [6] menemukan bahwa sebanyak 55% anak yang *stunting* berasal dari daerah pedesaan. Oleh karena itu, penting untuk memahami keterkaitan antara faktor sosioekonomi dan *stunting* guna mengoptimalkan upaya pencegahan dan penanganannya di Indonesia.

Dalam memprediksi *stunting*, penggunaan *machine learning* (ML) menjadi pendekatan yang efektif. *Machine learning* memainkan peran penting dalam mendukung

strategi pemerintah, terutama dalam intervensi kesehatan masyarakat untuk menurunkan prevalensi *stunting*. Dengan memberikan prediksi yang akurat dan membantu pengambilan keputusan berbasis data, *machine learning* mampu merancang intervensi yang efektif dan kontekstual [7]. Dibandingkan metode konvensional, *machine learning* lebih unggul dalam generalisasi dan pengolahan data besar, terutama dalam kasus *stunting* yang melibatkan berbagai faktor, seperti sosioekonomi. Selain fleksibilitasnya dalam menangani berbagai data, *machine learning* juga dapat memproses volume data besar secara efisien dan meningkatkan kecepatan pengambilan keputusan [8]. Berbagai model *machine learning* tersedia, termasuk *logistic regression*, *decision trees*, *random forest*, *naive bayes*, *support vector machines* dan lainnya. Namun, penelitian ini berfokus pada model *random forest* dan *support vector machine*.

Beberapa hasil penelitian yang telah dilakukan, model *random forest* dan *support vector machine* menghasilkan prediksi cukup akurat dalam memprediksi. Penelitian yang dilakukan [9] model *random forest* dan *support vector machines* lebih akurat dengan nilai akurasi 88,3%. Pada penelitian [10] model *random forest* lebih unggul dengan nilai akurasi 87,75% dari pada model *naives bayes* dan *k-nearest neighborhood*. Penelitian [11] menggunakan model *random forest* menjadi yang terbaik dengan nilai akurasi 96% dibandingkan 4 model lainnya yaitu *decision tree*, *logistic regression*, *multinomial naïve bayes*, *k-nearest neighborhood*. Penelitian [12] model *support vector machine* lebih baik dengan nilai akurasi 90% dibandingkan dengan model *naïve bayes* dan *regresi logistic*. Begitu juga dalam penelitian [13] model *support vector machine* lebih baik dengan nilai akurasi 93% dibandingkan dengan model C45. Dengan nilai akurasi yang baik menjadikan alasan menggunakan kembali model *random forest* dan *support vector machine*.

Penggunaan model *random forest* dan *support vector machine* dalam prediksi *stunting* menjadi bagian dari tahapan siklus OSEM. Tahapan OSEM (*Obtain, Scrub, Explore, Model, Interpret*) sangat penting dalam penelitian ini karena menyediakan struktur sistematis untuk pengolahan data serta pembangunan model prediksi yang andal. Tahap *Obtain* berfokus pada pengumpulan data sosioekonomi yang digunakan di dalam penelitian. Tahap *Scrub* memastikan data bersih dari kesalahan atau nilai yang hilang, untuk analisis yang lebih akurat. Di tahap *Explore*, dilakukan eksplorasi untuk memahami karakteristik dan pola data, yang membantu dalam menemukan hubungan penting antara faktor sosioekonomi dan *stunting*. Tahap *Model* membangun model prediktif, seperti *random forest* dan *support vector machine*, untuk memprediksi risiko *stunting*. Terakhir, tahap *Interpret* menafsirkan hasil model untuk mendukung pengambilan keputusan. Hasil penelitian ini diharapkan dapat

berguna bagi pemerintah dalam menurunkan prevalensi *stunting* di Indonesia secara lebih efektif dan tepat sasaran.

Berdasarkan permasalahan yang dihadapi dan Solusi yang ditawarkan, maka dalam penelitian ini akan mengangkat sebuah judul “**Analisis Prediksi *Stunting* Pada Balita Menggunakan *Machine Learning* dengan Faktor Sosioekonomi**”.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah yang akan dibahas dalam penelitian adalah sebagai berikut:

1. Bagaimana performa model *machine learning random forest* dan *support vector machine* dapat memprediksi risiko *stunting* pada balita secara akurat berdasarkan faktor sosioekonomi?
2. Faktor sosioekonomi mana yang paling signifikan mempengaruhi prediksi risiko *stunting* pada balita?

## 1.3 Tujuan

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Menganalisis performa model *machine learning random forest* dan *support vector machine* dalam memprediksi risiko *stunting* pada balita berdasarkan faktor sosioekonomi dengan matrik evaluasi *accuracy*, *precision*, *recall*, dan *F1-score*.
2. Mengidentifikasi faktor sosioekonomi yang paling signifikan dalam mempengaruhi prediksi risiko *stunting* pada balita menggunakan model *machine learning*.

## 1.4 Manfaat

Manfaat yang didapatkan dari penelitian ini adalah sebagai berikut:

1. Memberikan wawasan kepada peneliti dan praktisi mengenai performa *machine learning* dalam memprediksi *stunting* pada balita berdasarkan faktor sosioekonomi.
2. Menyediakan wawasan tentang faktor sosioekonomi yang paling signifikan dalam mempengaruhi risiko *stunting* pada balita.
3. Sebagai landasan bagi pemerintah dalam menyusun kebijakan penurunan *stunting* yang lebih tepat sasaran dengan menggunakan prediksi berbasis data yang mempertimbangkan faktor sosioekonomi.

## 1.5 Ruang Lingkup

Adapun ruang lingkup dari penelitian ini, antara lain sebagai berikut:

1. Data yang digunakan untuk memprediksi adalah *dataset Gross Domestic Product (GDP)*, Indeks Pembangunan Manusia (IPM), Persentase Penduduk Miskin (P0), Indeks Desa Membangun (IDM), Aksi Konvergensi, Upah, Prevalensi Gizi Balita dan Persentase Tinggi Balita di Indonesia dari [bps.go.id](http://bps.go.id), [layanandata.kemkes.go.id](http://layanandata.kemkes.go.id), [kemnaker.go.id](http://kemnaker.go.id), [kemendesa.go.id](http://kemendesa.go.id) dan [kemendagri.go.id](http://kemendagri.go.id) tahun 2019-2022.
2. Analisis data menggunakan bahasa pemrograman R serta *software RStudio*.
3. Untuk mendukung proses penelitian dan analisis data akan menggunakan framework OSEMN.
4. Menggunakan model *machine learning random forest* dan *support vector machine* untuk analisis prediksi klasifikasi *stunting*.
5. Proses menangani *features selection* menggunakan *correlation matrix* dan *mutual information*.
6. Matriks kinerja dari model prediksi klasifikasi akan dihitung dengan menggunakan *Accuracy, Precision, Recall* dan *F1-score*.

