

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kerentanan pangan (*food insecurity*) merupakan salah satu isu utama dalam pembangunan sosial dan ekonomi di Indonesia. Pemerintah telah menempatkan ketahanan pangan sebagai prioritas nasional, sejalan dengan Tujuan Pembangunan Berkelanjutan (*Sustainable Development Goals/SDGs*) poin kedua, yaitu mengakhiri kelaparan dan memastikan akses terhadap pangan bergizi bagi semua lapisan masyarakat. Meskipun demikian, berdasarkan laporan Badan Pusat Statistik (BPS, 2024) dan *Food and Agriculture Organization* (FAO, 2023), masih terdapat disparitas yang signifikan antarwilayah, terutama di kawasan timur Indonesia. Berdasarkan data Indeks Ketahanan Pangan (IKP) yang dirilis oleh Badan Pangan Nasional (Bapanas, 2024), rata-rata nilai IKP nasional berada pada angka 70,00, sedangkan Provinsi Nusa Tenggara Timur (NTT) hanya mencapai 66,92, yang menunjukkan bahwa tingkat kerentanan pangan di wilayah tersebut masih berada di bawah rata-rata nasional. Kondisi ini memperlihatkan bahwa *food insecurity* bukan hanya persoalan ketersediaan pangan, tetapi juga hasil dari interaksi kompleks antara faktor sosial, ekonomi, dan lingkungan.

Untuk memahami kompleksitas tersebut, pendekatan berbasis data (*data-driven approach*) menjadi sangat penting. Salah satu sumber data nasional yang dapat digunakan yaitu Survei Sosial Ekonomi Nasional (SUSENAS). Salah satu modul kuesioner di SUSENAS menyediakan informasi mendalam terkait karakteristik sosial-ekonomi rumah tangga seperti tingkat pendidikan, akses teknologi, serta variabel-variabel kesejahteraan lainnya. Melalui proses

transformasi variabel menjadi bentuk numerik dan proporsi (persentase), data ini dapat digunakan untuk klasifikasi rumah tangga yang berisiko mengalami *food insecurity*. Misalnya, proporsi anggota rumah tangga yang bekerja, tingkat pendidikan anggota, serta akses terhadap fasilitas dasar dapat menjadi indikator penting dalam menentukan status kerentanan pangan suatu rumah tangga.

Beberapa penelitian terdahulu menunjukkan bahwa pendekatan *machine learning* (ML) mampu memberikan hasil klasifikasi yang lebih akurat dibandingkan metode statistik konvensional. Khikmah *et al.* (2024) menemukan bahwa algoritma *machine learning* dapat mengklasifikasi status kerentanan pangan rumah tangga di Provinsi Jawa Barat, yaitu *Random Forest*, *Gradient Boosting*, *Rotation Forest*, dan *Extremely Randomized Tree*. Hasil serupa dikonfirmasi oleh Ramadhani *et al.* (2022), metode *Random Forest* terbukti memiliki performa terbaik dalam klasifikasi status kerentanan pangan rumah tangga menggunakan data SUSENAS 2020, dengan nilai AUC (*Area Under the Curve*) 0,718, lebih tinggi dibandingkan *Classification Tree* (AUC 0,668).

Beberapa penelitian sebelumnya telah menunjukkan efektivitas penggunaan pendekatan *machine learning* berbasis data survei dalam mengklasifikasi kerentanan pangan rumah tangga. Penelitian oleh Ramadhani *et al.* (2022) menggunakan data SUSENAS 2020 di Provinsi Aceh dan menemukan bahwa model *Random Forest* memiliki performa lebih baik dibandingkan *Classification Tree* dalam mengklasifikasikan kondisi *food insecurity*, dengan nilai AUC sebesar 0,718. Selanjutnya, Khikmah *et al.* (2024) memanfaatkan data survei rumah tangga di Jawa Barat dan membandingkan beberapa algoritma seperti *Random Forest*, *Gradient Boosting*, dan *Extremely Randomized Trees* untuk mengklasifikasi

pengalaman kerentanan pangan. Hasil studi tersebut menunjukkan bahwa metode *ensemble* memberikan akurasi dan stabilitas klasifikasi yang lebih tinggi dibanding model tunggal. Secara internasional, Qasrawi *et al.* (2024) menerapkan pendekatan *machine learning* pada data survei rumah tangga di Palestina menggunakan teknik *balancing* seperti *Synthetic Minority Over-sampling Technique* (SMOTE) dan menemukan bahwa model *machine learning* mampu klasifikasi variabel-variabel penentu kerentanan pangan secara lebih komprehensif dibanding pendekatan statistik konvensional.

Namun, model tunggal seperti Regresi Logistik atau *Decision Tree* sering kali terbatas dalam menangkap hubungan non-linear antarvariabel dan cenderung rentan terhadap *overfitting*. Untuk mengatasi hal tersebut, digunakan pendekatan *ensemble learning*, khususnya metode *Stacking Ensemble Learner* (SEL). Teknik ini pertama kali diperkenalkan oleh Ting & Witten pada tahun 1999 dan dikembangkan lebih lanjut oleh Diks & Vrugt (2010) serta Zhou (2012), yang menjelaskan bahwa *stacking* menggabungkan berbagai algoritma dasar (*base model*) untuk membentuk model yang lebih kuat dan stabil dengan cara mengurangi bias serta varians klasifikasi. Selain itu, Galar *et al.* (2011) menegaskan bahwa metode *ensemble* efektif dalam mengatasi kelemahan model individual, terutama pada permasalahan klasifikasi dengan distribusi kelas yang tidak seimbang dan risiko kesalahan deteksi pada kelas minoritas. Hal ini sejalan dengan temuan Khan *et al.* (2023) yang menyatakan bahwa penggunaan *ensemble* secara konsisten memberikan peningkatan sensitivitas model (*recall*) dan kualitas klasifikasi pada kasus *high-risk group classification*. Oleh karena itu, metode *ensemble* layak diterapkan dalam

penelitian ini untuk meningkatkan kemampuan model dalam klasifikasi rumah tangga yang rentan pangan.

Dalam penelitian ini, digunakan empat model utama sebagai *base model* yang mewakili jenis algoritma berbeda, yaitu Regresi Logistik, Random Forest, dan *eXtreme Gradient Boosting* (XGBoost). Regresi Logistik dipilih karena kemampuannya menjelaskan hubungan antara faktor sosial-ekonomi dengan probabilitas rumah tangga mengalami *food insecurity*. Random Forest digunakan karena sifatnya yang *non-parametric* dan kemampuannya mengatasi data dengan banyak variabel tanpa memerlukan asumsi distribusi tertentu (Breiman, 2001). XGBoost ditambahkan untuk menangani hubungan kompleks dan meningkatkan kemampuan generalisasi model melalui proses pembobotan iteratif yang adaptif (Chen & Guestrin, 2016). Kombinasi keempat model ini kemudian diintegrasikan dalam kerangka *Stacking Ensemble Learner* dengan Regresi Logistik sebagai *meta model* untuk mengoptimalkan hasil klasifikasi.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menguji efektivitas metode *Stacking Ensemble Learner* (SEL) dalam mengklasifikasi tingkat kerentanan pangan rumah tangga di Provinsi Nusa Tenggara Timur (NTT) dengan memanfaatkan data Survei Sosial Ekonomi Nasional (SUSENAS) Tahun 2023. Fokus utama penelitian yakni klasifikasi rumah tangga yang berada pada kondisi *food insecure* secara seakurat mungkin agar dapat klasifikasi kelompok yang membutuhkan prioritas penanganan. Oleh karena itu, penelitian ini menggunakan metrik evaluasi *F1-Score* sebagai ukuran utama performa model. Penggunaan *F1-Score* dipilih karena mampu menggambarkan keseimbangan kinerja antara *precision* dan *recall*, sehingga tidak hanya menilai kemampuan

model dalam mengklasifikasi dengan benar, tetapi juga memastikan bahwa model tidak mengabaikan kasus rumah tangga yang tergolong *food insecure*.

Metrik *F1-Score* menjadi lebih relevan dalam konteks ini karena *accuracy* saja tidak cukup mewakili kualitas klasifikasi model. *F1-Score* memberikan ukuran yang lebih komprehensif untuk menilai kemampuan model dalam klasifikasi rumah tangga yang benar-benar rentan dengan mempertimbangkan harmonisasi antara *precision* dan *recall*, mengingat kesalahan klasifikasi pada kelas *insecure* (kelas 1) berpotensi berdampak serius terhadap efektivitas perumusan kebijakan. Melalui pendekatan ini, diharapkan model klasifikasi yang dikembangkan dapat memberikan dukungan berbasis data dalam pengambilan keputusan pemerintah, khususnya untuk perencanaan intervensi sosial dan penguatan program kerentanan pangan di wilayah berisiko tinggi seperti Provinsi Nusa Tenggara Timur.

1.2 Rumusan Masalah

1. Bagaimana membangun model prediksi dengan menggunakan algoritma Regresi Logistik, Random Forest, dan XGBoost?
2. Bagaimana proses penerapan metode *Stacking Ensemble Learner* (SEL) dengan kombinasi *base model* dan *meta model* untuk meningkatkan klasifikasi status rumah tangga *food insecure* di Provinsi Nusa Tenggara Timur?
3. Bagaimana kinerja metode *Stacking Ensemble Learner* (SEL) dibandingkan dengan model tunggal dalam mengklasifikasi status rumah tangga *food insecure* berdasarkan data SUSENAS (VSEN23.K) tahun 2023?

1.3 Batasan Masalah

Penelitian ini dibatasi pada eksplorasi konfigurasi dan struktur *Stacking Ensemble Learner* (SEL) menggunakan model dasar Regresi Logistik, Random Forest, dan *eXtreme Gradient Boosting* (XGBoost). Data yang digunakan berasal dari Survei Sosial Ekonomi Nasional (SUSENAS) 2023, yang mencakup informasi sosial ekonomi rumah tangga di Provinsi Nusa Tenggara Timur (NTT). Variabel independen dibatasi pada sejumlah indikator sosial dan ekonomi rumah tangga. Variabel dependen ditetapkan berdasarkan status kerentanan pangan rumah tangga (*food security*) yang dikategorikan menjadi dua kelas yaitu tahan pangan (1) dan rentan pangan (0).

Evaluasi kinerja model difokuskan pada metrik *F1-Score* pada kelas 1 sebagai ukuran utama untuk menilai keseimbangan performa model dalam klasifikasi rumah tangga pada kelas *food insecure*. Hal ini dilakukan untuk memastikan bahwa model tidak hanya menghasilkan klasifikasi yang akurat secara keseluruhan, tetapi juga mampu klasifikasi sampel kelas yang berisiko dengan lebih efektif tanpa bias terhadap salah satu kelas. Analisis dilakukan secara *cross-sectional* tanpa mempertimbangkan aspek temporal maupun spasial di luar wilayah Provinsi Nusa Tenggara Timur.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah, terdapat tujuan yang ingin dicapai dalam penulisan Tugas Akhir ini sebagai berikut.

1. Membangun model klasifikasi menggunakan algoritma Regresi Logistik, Random Forest, dan XGBoost dalam mengklasifikasi rumah tangga berdasarkan tingkat kerentanan pangan.

2. Menerapkan metode *Stacking Ensemble Learner* (SEL) untuk menggabungkan hasil klasifikasi dari model dasar (*base model*) dengan Regresi Logistik sebagai *meta model* guna meningkatkan akurasi dan kemampuan generalisasi model.
3. Menganalisis dan membandingkan kinerja model tunggal dan model *ensemble* berdasarkan metrik evaluasi *F1-Score* untuk memperoleh model yang memiliki keseimbangan terbaik dalam klasifikasi rumah tangga pada kelas *food insecure*.