

BAB 1 PENDAHULUAN

1.1 Latar Belakang

Padi merupakan komoditas pangan utama bagi sebagian besar penduduk Indonesia. Dengan produktivitas rata-rata nasional mencapai 5,1 ton per hektar dan total produksi gabah kering giling lebih dari 54 juta ton pada tahun 2023, beras menjadi sumber kalori dominan yang menopang ketahanan pangan seluruh lapisan masyarakat [1]. Pemerintah menempatkan beras dalam daftar bahan pokok strategis yang harus selalu tersedia, sehingga setiap pergeseran angka produksi tahunan dapat memengaruhi keputusan impor, alokasi pupuk bersubsidi, dan harga eceran. Sayangnya, informasi mengenai produktivitas padi sering kali baru tersedia beberapa bulan setelah musim panen berakhir karena pengumpulan data lapangan melalui metode Kerangka Sampel Area membutuhkan ribuan petugas dan tahap verifikasi yang panjang [1, 2]. Akibatnya, pengambil kebijakan kerap dihadapkan pada situasi ketika data terbaru baru muncul setelah periode pengambilan keputusan terlewati. Target prediksi dinyatakan secara langsung sebagai produktivitas padi dalam satuan ton per hektare, bukan total produksi gabah dalam satuan ton. Kebutuhan informasi yang lebih cepat juga semakin penting karena variasi iklim, khususnya kejadian kekeringan, dapat memengaruhi kondisi pertanaman padi dan perlu dideteksi sebelum dampaknya tercermin pada statistik akhir musim panen [3].

Machine learning digunakan untuk prediksi hasil tanaman [4]. Tinjauan sistematis juga mencatat pemanfaatan *deep learning* dan penginderaan jauh [5]. Data multi-sumber digunakan untuk prediksi produktivitas padi di Tiongkok [6]. Citra satelit, iklim, tanah, dan riwayat panen digunakan untuk prediksi produktivitas gandum di Amerika Serikat [7]. Sentinel-2 digunakan untuk prediksi produktivitas padi di Nepal [8] dan Pakistan [9]. Penginderaan jauh juga diasimilasikan dengan model pertumbuhan tanaman [10]. Di Indonesia, studi Sentinel mencakup Jawa Barat [11] dan Kabupaten Batang [12]. Kajian nasional berfokus pada model iklim, hidrologi, dan tanaman [13]. Belum ditemukan integrasi delapan sumber data untuk seluruh provinsi menggunakan model gabungan yang dapat diinterpretasikan. Selain itu, sebagian besar studi terdahulu masih berhenti pada evaluasi model, sehingga belum banyak menyediakan media eksplorasi interaktif yang membantu pengguna ahli non-pemrogram memahami hasil prediksi,

kesalahan model, dan kontribusi fitur secara lebih mudah. Google Earth Engine menyediakan infrastruktur komputasi geospasial berskala besar yang mendukung pengolahan koleksi citra multi-temporal tanpa harus mengunduh seluruh data mentah ke komputer lokal [14, 15]. Kemampuan tersebut relevan untuk menyusun fitur yang konsisten pada cakupan provinsi dan periode pengamatan yang panjang.

Kesenjangan tersebut ditangani melalui dua luaran yang saling terhubung. Bagian pertama adalah pembangunan model gabungan (*ensemble stacking*) yang menyatukan beberapa algoritma *machine learning* berbasis pohon dengan masukan fitur biofisik multi-sumber dari Google Earth Engine. Data yang digunakan mencakup wilayah provinsi di Indonesia pada periode 2018–2025 dan melalui proses penyaringan kualitas sebelum digunakan pada tahap pelatihan dan evaluasi model. Stacking digunakan untuk mempelajari kombinasi keluaran beberapa model dasar sehingga informasi prediktifnya dapat saling melengkapi [16, 17]. Model ini divalidasi melalui studi ablasi bertahap dan dilengkapi analisis kontribusi fitur menggunakan metode SHAP (*SHapley Additive exPlanations*) untuk menjelaskan atribusi keluaran model tanpa menafsirkannya sebagai hubungan sebab-akibat [18, 19]. Bagian kedua adalah pengembangan aplikasi berbasis Streamlit yang diimplementasikan dalam satu berkas skrip Python dan dapat dijalankan secara lokal di laptop pengguna. Aplikasi ini diposisikan sebagai media demonstrasi dan eksplorasi hasil model, bukan sebagai sistem produksi atau layanan prediksi operasional. Melalui aplikasi tersebut, pengguna ahli non-pemrogram, seperti akademisi, dosen, maupun praktisi pertanian, dapat melihat evaluasi model, analisis prediksi per provinsi, peringkat produktivitas antarprovinsi, studi ablasi, data lengkap prediksi, dan visualisasi kontribusi fitur secara interaktif tanpa perlu menjalankan kode pemodelan secara manual.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah sebagai berikut.

1. Bagaimana implementasi dan performa model *ensemble stacking* berbasis fitur biofisik multi-sumber citra satelit dan iklim untuk memprediksi produktivitas padi tingkat provinsi di Indonesia?
2. Bagaimana pengaruh perbedaan target, fitur, dan skema validasi pada enam konfigurasi eksperimen terhadap performa model, serta fitur biofisik apa yang

memberikan kontribusi utama terhadap prediksi Random Forest berdasarkan analisis SHAP?

3. Bagaimana hasil model *ensemble stacking* disajikan melalui aplikasi Streamlit lokal sebagai media demonstrasi dan eksplorasi?

1.3 Batasan Masalah

Ruang lingkup berikut membatasi ketiga rumusan masalah sebelum tujuan penelitian ditetapkan. Batasan penelitian dinyatakan secara operasional sebagai berikut.

1. Target dan satuan dibatasi pada produktivitas padi dalam ton per hektare. Total produksi gabah dalam ton tidak digunakan sebagai target model utama.
2. Unit analisis dan pelaporan dibatasi pada provinsi–tahun di Indonesia. Empat provinsi pemekaran Papua yang dibentuk pada tahun 2022 belum diikutsertakan karena belum tersedia pada lapisan batas administrasi yang digunakan. Rincian jumlah observasi hasil ekstraksi dan observasi efektif dijelaskan pada Bab 3 dan Bab 4.
3. Periode data dibatasi pada tahun 2018–2025 atau delapan periode pengamatan tahunan. Data sebelum tahun 2018 tidak digunakan karena BPS mengubah metodologi pengumpulan data luas panen pada awal periode tersebut.
4. Sumber data dibatasi pada statistik BPS dan dataset terbuka yang tersedia melalui Google Earth Engine dalam kategori optik, SAR, indeks vegetasi, suhu permukaan lahan, curah hujan, reanalisis iklim, topografi, dan tutupan lahan. Citra komersial berbayar tidak digunakan.
5. Penggunaan citra satelit dibatasi sebagai sumber pengukuran raster multitemporal dari Sentinel-1, Sentinel-2, dan MODIS. Citra merupakan kisi piksel yang menyimpan nilai pantulan spektral, hamburan balik radar, atau suhu permukaan, bukan foto yang ditampilkan kepada model. Piksel disaring, dibatasi dengan masker lahan, dibentuk menjadi komposit, dan diagregasi menjadi fitur numerik per provinsi–tahun. Model menerima tabel hasil ekstraksi tersebut, bukan matriks gambar untuk klasifikasi berbasis jaringan saraf konvolusional.

6. Model dibatasi pada gabungan empat algoritma berbasis pohon dengan satu model linier teregulasi sebagai penggabung (*meta-learner*). Eksplorasi terhadap model *deep learning* (jaringan saraf tiruan berlapis dalam) tidak dilakukan karena karakteristik data tabular agregasi provinsi-tahun dengan jumlah sampel yang relatif terbatas cenderung lebih optimal dimodelkan menggunakan algoritma berbasis pohon dibandingkan jaringan saraf tiruan berlapis dalam tanpa mekanisme *transfer learning*.
7. Evaluasi dibatasi pada enam konfigurasi eksperimen, termasuk *temporal holdout*, LOPO, dan *rolling-origin*. Analisis SHAP menjelaskan atribusi fitur pada Random Forest dan tidak digunakan sebagai bukti hubungan sebab-akibat atau dekomposisi lengkap prediksi final *stacking*.
8. Aplikasi dibatasi sebagai media demonstrasi lokal yang memuat hasil terhitung sebelumnya (*pre-computed*). Aplikasi tidak melakukan ekstraksi citra atau pelatihan ulang saat dijalankan dan bukan sistem prediksi *real-time* untuk area baru. Fitur pengelolaan banyak pengguna sekaligus, optimasi performa untuk skala besar, dan penerapan ke server publik tidak termasuk dalam lingkup skripsi ini.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah tersebut, tujuan penelitian ini adalah sebagai berikut.

1. Mengimplementasikan dan mengevaluasi performa model *ensemble stacking* berbasis fitur biofisik multi-sumber citra satelit dan iklim untuk memprediksi produktivitas padi tingkat provinsi di Indonesia.
2. Menganalisis pengaruh perbedaan target, fitur, dan skema validasi pada enam konfigurasi eksperimen terhadap performa model serta mengidentifikasi fitur biofisik yang memberikan kontribusi utama terhadap prediksi Random Forest berdasarkan analisis SHAP.
3. Menyajikan hasil model *ensemble stacking* melalui aplikasi Streamlit lokal sebagai media demonstrasi dan eksplorasi.

1.5 Manfaat Penelitian

Manfaat penelitian dirumuskan dalam tiga kelompok berikut.

1. Manfaat akademis berupa rancangan *pipeline* prediksi produktivitas padi tingkat provinsi yang menggabungkan statistik BPS, fitur hasil penginderaan jauh, data iklim, model *ensemble stacking*, studi ablasi, dan interpretasi SHAP. Rancangan tersebut dapat menjadi rujukan untuk mengevaluasi hubungan antara fitur biofisik dan produktivitas tanaman pangan di Indonesia.
2. Manfaat praktis berupa model dan aplikasi Streamlit lokal yang membantu akademisi, dosen, praktisi pertanian, dan pengambil kebijakan mengeksplorasi prediksi provinsi–tahun, perbandingan performa eksperimen, kesalahan prediksi, dan kontribusi fitur tanpa menjalankan kode pemodelan secara manual.
3. Manfaat teknis berupa kode sumber, dataset hasil pemrosesan, model terlatih, dan aplikasi Streamlit lokal yang dapat dikembangkan ulang untuk komoditas, periode, atau wilayah lain. Artefak tersebut menyediakan dasar implementasi yang dapat ditelusuri dari pengolahan data sampai penyajian hasil model.

1.6 Sistematika Penulisan

Penulisan skripsi ini terdiri atas lima bab dengan susunan sebagai berikut.

1. Bab 1 Pendahuluan memuat latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.
2. Bab 2 Landasan Teori menyajikan dasar teori yang mencakup fisiologi padi, penginderaan jauh dengan Google Earth Engine, formulasi indeks vegetasi dan parameter iklim, dasar matematis setiap algoritma yang digunakan, metode SHAP, serta tinjauan penelitian terdahulu.
3. Bab 3 Metodologi Penelitian menguraikan *pipeline machine learning* dari ekstraksi fitur hingga pelatihan model, serta rancangan aplikasi Streamlit sebagai media pemuatan artefak, visualisasi hasil prediksi, dan interpretasi fitur.

4. Bab 4 Hasil dan Pembahasan memaparkan hasil studi ablas, analisis SHAP, perbandingan prediksi terhadap data aktual BPS, identifikasi pola kesalahan per wilayah, dan demonstrasi aplikasi Streamlit lokal.
5. Bab 5 Simpulan dan Saran berisi kesimpulan yang menjawab tujuan penelitian serta saran untuk penelitian lanjutan.

