

BAB I

PENDAHULUAN

1.1 Latar Belakang

Ketidakseimbangan pembangunan antarwilayah di Indonesia masih menjadi tantangan utama yang menghambat terwujudnya pemerataan kesejahteraan. Kondisi ini menciptakan perbedaan mencolok antara wilayah yang telah berkembang dan yang masih tertinggal, terutama dalam hal infrastruktur, ketersediaan layanan publik, serta akses terhadap peluang ekonomi. Ketidakseimbangan tersebut tidak hanya memperlebar jurang sosial dan ekonomi antar masyarakat, tetapi juga memperburuk kesenjangan tingkat kesejahteraan, serta memunculkan berbagai permasalahan sosial lanjutan seperti urbanisasi yang tidak terkendali, meningkatnya angka pengangguran, dan kemiskinan struktural yang terus berlangsung.

Sutaryono *et al.* (2014) menyatakan bahwa perbedaan karakteristik fisik dan sosial ekonomi antarwilayah menyebabkan ketimpangan dalam akses terhadap sumber daya dan peluang pembangunan. Ketimpangan ini menjadi hambatan bagi terwujudnya pembangunan yang adil, merata, dan berkelanjutan. Pembangunan nasional seharusnya tidak hanya berorientasi pada pertumbuhan (*growth*), tetapi juga mencakup aspek pemerataan (*equity*), peningkatan kesejahteraan (*welfare*), dan keberlanjutan (*sustainability*).

Data statistik terkini dari BPS menunjukkan adanya ketidakseimbangan pembangunan yang nyata. Tingkat kemiskinan di daerah perdesaan Indonesia tercatat sebesar 11,79% pada September 2024, sedangkan di daerah perkotaan

hanya sebesar 7,09% (Badan Pusat Statistik, 2025). Provinsi Jawa Tengah menunjukkan pola serupa, yakni tingkat kemiskinan di perdesaan yang masih lebih tinggi dibandingkan kawasan perkotaan (Badan Pusat Statistik Provinsi Jawa Tengah, 2025). Disparitas ini mengindikasikan adanya kesenjangan struktural yang perlu ditangani secara sistematis dan menyeluruh.

Provinsi Jawa Tengah merupakan provinsi dengan jumlah penduduk besar yang tidak lepas dari problematika ketidakseimbangan pembangunan wilayah. Wilayah ini memiliki 7.810 desa yang tersebar di 29 kabupaten dan 6 kota (Badan Pusat Statistik Provinsi Jawa Tengah, 2025). Keberagaman geografis, sosial, budaya, dan ekonomi antarwilayah menciptakan variasi perkembangan desa yang mencolok.

Variasi perkembangan desa di Jawa Tengah tercermin dalam status desa berdasarkan Indeks Desa Membangun (IDM). IDM merupakan alat ukur penting untuk menilai perkembangan desa dari aspek sosial, ekonomi, dan lingkungan. Pemerintah mengklasifikasikan desa ke dalam lima kelas yang terdiri dari Sangat Tertinggal, Tertinggal, Berkembang, Maju, dan Mandiri berdasarkan IDM. Klasifikasi ini membantu pemerintah mengidentifikasi desa-desa yang memerlukan perhatian lebih melalui pembangunan infrastruktur, pengembangan kapasitas sumber daya manusia, serta program pemberdayaan ekonomi (Kementerian Desa PDTT, 2024).

Pemerintah Provinsi Jawa Tengah dapat menyusun intervensi kebijakan yang lebih tepat sasaran dan berbasis data dengan memanfaatkan IDM. Klasifikasi desa yang disusun berdasarkan indeks ini dapat digunakan untuk mengevaluasi dampak kebijakan terhadap kelompok desa dengan tingkat pembangunan yang berbeda.

Pemahaman yang akurat terhadap status desa memungkinkan pembangunan dilaksanakan secara berkelanjutan dan inklusif. Pendekatan berbasis IDM ini menjadi langkah strategis dalam mendorong pemerataan pembangunan wilayah (Kementerian Desa PDTT, 2024).

Seiring dengan meningkatnya ketersediaan data dan kebutuhan akan analisis yang lebih komprehensif, berbagai penelitian telah memanfaatkan metode berbasis data untuk mengklasifikasikan status desa. Salah satu pendekatan yang banyak digunakan dalam konteks ini adalah *machine learning*. *Machine learning* merupakan cabang dari kecerdasan buatan (*artificial intelligence*) yang memungkinkan komputer belajar dan mengambil keputusan berdasarkan data tanpa diprogram secara eksplisit (Mitchell, 1997).

Algoritma *machine learning* terbukti efektif dalam tugas klasifikasi karena kemampuannya menganalisis pola kompleks dari banyak variabel untuk mengelompokkan data ke dalam kategori yang ditentukan (Kotsiantis *et al.*, 2007). Klasifikasi merupakan proses pengelompokan objek atau entitas ke dalam kategori-kategori tertentu berdasarkan karakteristik atau atribut yang dimilikinya (Han *et al.*, 2012).

Beberapa data di dunia nyata tidak selalu memiliki jumlah yang seimbang di setiap kategori karena tingkat perkembangan yang bervariasi. Ketidakseimbangan ini dapat menyebabkan bias pada model, sehingga algoritma cenderung lebih akurat dalam mengenali kelas mayoritas, namun kurang efektif dalam mengklasifikasikan kelas minoritas (Chawla *et al.*, 2002). Untuk mengatasi hal tersebut, Batista *et al.* (2004) menjelaskan bahwa teknik seperti *Random Oversampling* dapat digunakan untuk meningkatkan representasi kelas minoritas

dengan menduplikasi sampel kelas minoritas secara acak hingga distribusi kelas menjadi lebih seimbang.

Beberapa penelitian sebelumnya telah menunjukkan bahwa *Gradient Boosting* dan *Random Forest* merupakan dua algoritma klasifikasi yang memiliki performa unggul dibandingkan metode lainnya. Omar *et al.* (2024) melakukan perbandingan antara *Logistic Regression*, *Support Vector Machine* (SVM), *Random Forest*, dan *Gradient Boosting* menggunakan data rekam medis pasien pascaoperasi jantung, dengan hasil *Gradient Boosting* mencapai akurasi tertinggi sebesar 88,66%, diikuti oleh *Random Forest* sebesar 87,39%. Zhang *et al.* (2024) juga membandingkan berbagai algoritma dalam penelitian penilaian risiko kredit dan menemukan bahwa *Gradient Boosting* menghasilkan F1-score tertinggi sebesar 91%, diikuti oleh *Random Forest* dengan F1-score sebesar 89%. Penelitian lain yang dilakukan oleh Ratnasari dan Nur'aini (2024) menemukan bahwa dengan menerapkan *Random Oversampling*, *Gradient Boosting* dan *Random Forest* menunjukkan performa paling unggul dibandingkan model lain seperti *Extra Trees*, baik dari segi akurasi maupun sensitivitas menggunakan data *Telco Customer Churn*. Temuan-temuan ini mendukung penggunaan kedua algoritma tersebut dalam penelitian ini.

Penelitian ini menggunakan pendekatan *machine learning* dengan algoritma *Random Forest* dan *Gradient Boosting* untuk mengklasifikasikan status desa, karena kedua algoritma tersebut terbukti unggul dalam penelitian sebelumnya. Sebagian besar penelitian sebelumnya menggunakan data biner, sedangkan penelitian ini menguji klasifikasi multikelas dengan empat kategori status desa.

Penelitian ini juga menerapkan teknik penanganan data tidak seimbang berupa *Random Oversampling* untuk memastikan distribusi kelas yang lebih seimbang.

Breiman (2001) mengembangkan metode *Random Forest* sebagai algoritma klasifikasi *ensemble* yang bekerja dengan membentuk banyak pohon keputusan melalui teknik *Bootstrap Aggregating (bagging)* dan pemilihan fitur secara acak. Algoritma ini menggabungkan klasifikasi dari sejumlah pohon independen yang dilatih pada *subset* data, lalu melakukan voting mayoritas untuk menghasilkan klasifikasi final.

Friedman (2001) memperkenalkan *Gradient Boosting* sebagai teknik *machine learning* yang kuat untuk membangun model prediktif. Algoritma ini melakukan proses pembelajaran secara bertahap, dengan setiap model baru berfungsi untuk memperbaiki kesalahan yang dibuat oleh model sebelumnya. *Gradient Boosting* menggabungkan beberapa pohon keputusan secara iteratif.

Penelitian di beberapa wilayah telah menerapkan metode klasifikasi status desa berbasis IDM. Sirodj *et al.* (2023) menggunakan metode hutan ordinal untuk klasifikasi status desa di Jawa Barat, sementara Latuconsina *et al.* (2024) menggunakan SVM untuk klasifikasi desa di Maluku. Beragam metode klasifikasi telah digunakan di wilayah lain, tetapi sejauh penelusuran dilakukan, belum ditemukan kajian serupa yang fokus di Provinsi Jawa Tengah. Kondisi ini mendorong peneliti melakukan penelitian berjudul “Perbandingan *Random Forest* dan *Gradient Boosting* dalam Klasifikasi Status Desa di Jawa Tengah”, untuk menguji efektivitas kedua metode dalam mengklasifikasikan status desa berbasis IDM di Jawa Tengah.

1.2 Rumusan Masalah

Latar belakang di atas mengarahkan pada rumusan masalah dari penelitian ini sebagai berikut:

1. Bagaimana penerapan algoritma *Random Forest* dalam mengklasifikasikan status desa di Jawa Tengah?
2. Bagaimana penerapan algoritma *Gradient Boosting* dalam mengklasifikasikan status desa di Jawa Tengah?
3. Bagaimana perbandingan *Random Forest* dan *Gradient Boosting* dalam klasifikasi status desa di Jawa Tengah berdasarkan metrik evaluasi?

1.3 Batasan Masalah

1. Algoritma yang digunakan dalam penelitian ini adalah *Random Forest* dan *Gradient Boosting*, dengan fokus pada analisis kinerjanya dalam mengklasifikasikan status desa di Jawa Tengah berdasarkan berbagai metrik evaluasi menggunakan bahasa pemrograman Python.
2. Model *Random Forest* dan *Gradient Boosting* digunakan dengan parameter default tanpa optimasi *hyperparameter* tambahan.
3. Lingkup *dataset* yang digunakan dalam penelitian ini terbatas pada data yang tersedia dari sumber resmi katalog.data.go.id/fi/dataset/data-indeks-desa-membangun-tahun-2024, dengan asumsi bahwa semua variabel yang digunakan sudah melalui proses validasi dan bersifat representatif.
4. Evaluasi kinerja model dilakukan menggunakan metrik akurasi, presisi, sensitivitas, F1-score, dan G-Mean untuk membandingkan performa *Random Forest* dan *Gradient Boosting*.

1.4 Tujuan Penelitian

Rumusan masalah yang telah dikemukakan mengarah pada tujuan penelitian ini, yaitu:

1. Merapkan algoritma *Random Forest* dalam mengklasifikasikan status desa di Jawa Tengah.
2. Menerapkan algoritma *Gradient Boosting* dalam mengklasifikasikan status desa di Jawa Tengah.
3. Membandingkan performa algoritma *Random Forest* dan *Gradient Boosting* dalam klasifikasi status desa di Jawa Tengah berdasarkan metrik evaluasi.