

BAB 2

LANDASAN TEORI

Bab ini menguraikan kerangka teoretis yang menjadi landasan penelitian secara terorganisir. Teori yang dipaparkan dipilih atas dasar keterkaitan langsungnya dengan inti permasalahan, yakni pemetaan tutupan lahan berbasis citra satelit Sentinel-2 melalui algoritma *Random Forest* dalam lingkup Informatika. Cakupan pembahasan meliputi dinamika perubahan tutupan lahan dan deforestasi, prinsip penginderaan jauh, karakteristik citra satelit, spesifikasi Sentinel-2, platform Google Earth Engine, teknik klasifikasi citra, algoritma *machine learning*, indeks spektral vegetasi, serta metrik penilaian model. Di samping itu, bab ini turut memaparkan tinjauan atas penelitian sebelumnya beserta celah yang menjadi dasar kontribusi penelitian ini.

2.1 Perubahan Tutupan Lahan dan Deforestasi

Tutupan lahan (*land cover*) mencerminkan kondisi fisik permukaan bumi yang dapat diamati secara langsung, mencakup berbagai jenis seperti hutan, sawah, kawasan permukiman, badan air, maupun lahan terbuka. Konsep ini berbeda dari penggunaan lahan (*land use*) yang merujuk pada bagaimana manusia memanfaatkan suatu kawasan untuk kegiatan tertentu, seperti bercocok tanam, menjalankan industri, atau membangun hunian. Dengan kata lain, *land cover* menggambarkan kondisi fisik permukaan, sedangkan *land use* lebih berkaitan dengan fungsi dan aktivitas manusia di atasnya [1].

Perubahan tutupan lahan (*land cover change*) berlangsung secara berkelanjutan sebagai hasil dari interaksi antara aktivitas manusia dan dinamika alam. Salah satu manifestasi perubahan tutupan lahan yang paling berdampak adalah deforestasi, yaitu konversi kawasan berhutan menjadi kawasan tidak berhutan secara permanen [1]. Proses ini dapat dipicu oleh beragam faktor, mulai dari perluasan lahan pertanian, pembangunan prasarana, aktivitas industri, hingga perkembangan kawasan permukiman.

Deforestasi menimbulkan dampak nyata terhadap kualitas lingkungan dan stabilitas ekosistem. Terkikisnya tutupan hutan berujung pada penyusutan keanekaragaman hayati, intensifikasi degradasi tanah, serta melemahnya fungsi ekosistem sebagai penyerap karbon. Kondisi ini secara langsung mendorong

kenaikan konsentrasi gas rumah kaca di atmosfer dan mempercepat laju perubahan iklim secara global [2]. Oleh sebab itu, deforestasi kerap dijadikan indikator kunci dalam kajian perubahan tutupan lahan, baik pada skala regional maupun global.

Dalam perspektif Informatika dan analisis berbasis data, deforestasi dapat dipahami sebagai peralihan kelas tutupan lahan dari kategori hutan ke kategori non-hutan. Sudut pandang ini memungkinkan deforestasi dimodelkan sebagai persoalan klasifikasi biner (*binary classification*), di mana tiap satuan analisis direpresentasikan dalam bentuk data numerik yang dapat diolah secara komputasional. Kerangka klasifikasi semacam ini banyak diadopsi dalam bidang penginderaan jauh karena bersifat objektif, kuantitatif, dan mudah diotomatisasi [3].

Pemantauan deforestasi melalui metode survei lapangan konvensional kerap terkendala oleh besarnya biaya operasional, lamanya durasi pengumpulan data, serta terbatasnya cakupan wilayah yang dapat dijangkau, terutama di kawasan yang luas dan terpencil. Keterbatasan inilah yang mendorong penggunaan teknologi penginderaan jauh sebagai alternatif utama dalam menganalisis perubahan tutupan lahan. Melalui data citra satelit multitemporal, dinamika perubahan lahan dapat dipantau secara berkala sehingga pola deforestasi dapat dikenali dengan lebih efektif [11].

Dalam penelitian ini, deforestasi ditempatkan sebagai objek kajian utama yang direpresentasikan melalui pergeseran kelas pada hasil klasifikasi citra satelit. Pendekatan ini memfasilitasi integrasi antara konsep ekologis dan metode komputasi berbasis *machine learning*, sehingga deforestasi dapat dikaji secara kuantitatif dan sistematis.

2.1.1 Deforestasi Kotor dan Perubahan Tutupan Hutan Neto

Dalam analisis perubahan tutupan hutan berbasis penginderaan jauh, terdapat dua konsep yang perlu dibedakan secara eksplisit karena keduanya mengukur fenomena yang berbeda meskipun saling berkaitan, yaitu **deforestasi kotor** (*gross deforestation*) dan **perubahan tutupan hutan neto** (*net forest change*). Pemahaman yang tepat atas perbedaan keduanya sangat penting agar hasil analisis multitemporal dapat diinterpretasikan secara akurat [1].

Deforestasi kotor (*gross deforestation*) didefinisikan sebagai total luas kawasan yang beralih dari tutupan hutan menjadi non-hutan dalam rentang waktu tertentu, tanpa memperhitungkan perubahan yang terjadi dalam arah sebaliknya

[1]. Besaran ini hanya menghitung piksel atau unit area yang pada awal periode berstatus hutan kemudian menjadi non-hutan pada akhir periode ($1 \rightarrow 0$), sehingga mencerminkan kehilangan hutan secara absolut.

Forest gain adalah kebalikannya, yaitu penambahan luas area yang bertransisi dari non-hutan menjadi hutan ($0 \rightarrow 1$). Fenomena ini dapat disebabkan oleh program penghijauan, suksesi vegetasi alami, maupun artefak variasi reflektansi antar komposit citra [12].

Perubahan tutupan hutan neto (*net forest change*) diperoleh dari selisih antara *forest gain* dan deforestasi kotor:

$$\Delta_{\text{neto}} = A_{\text{forest gain}} - A_{\text{deforest kotor}} \quad (2.1)$$

Nilai $\Delta_{\text{neto}} > 0$ berarti luas hutan bertambah secara neto meskipun deforestasi kotor tetap berlangsung, sedangkan nilai negatif mengindikasikan penyusutan hutan secara neto. Tabel 2.1 merangkum perbedaan antara kedua konsep tersebut.

Tabel 2.1. Perbedaan deforestasi kotor dan perubahan tutupan hutan neto

Aspek	Deforestasi kotor	Perubahan neto
Definisi	Total luas hutan yang hilang ($1 \rightarrow 0$)	Selisih antara <i>forest gain</i> dan deforestasi kotor
Arah perubahan	Satu arah (hutan $\rightarrow$ non-hutan)	Dua arah ($1 \rightarrow 0$ dan $0 \rightarrow 1$)
Nilai mungkin	Selalu ≥ 0	Positif, negatif, atau nol
Interpretasi	Kehilangan hutan absolut	Kondisi akhir luas hutan relatif terhadap kondisi awal

Sumber: FAO (2022); Wulder dkk. (2022)

Penting dicatat bahwa deforestasi kotor dan perubahan neto positif dapat terjadi bersamaan dalam satu periode, yakni ketika *forest gain* melampaui deforestasi kotor. Dalam penelitian ini, deforestasi kotor dijadikan ukuran utama karena secara langsung mencerminkan kehilangan hutan tanpa tersamarkan oleh *forest gain* yang mungkin merupakan artefak klasifikasi [12].

2.2 Penginderaan Jauh

Penginderaan jauh adalah metode perolehan informasi mengenai suatu objek, kawasan, atau fenomena di permukaan bumi tanpa adanya kontak fisik langsung. Informasi tersebut diperoleh dengan cara merekam energi elektromagnetik yang dipantulkan maupun dipancarkan oleh objek di permukaan bumi menggunakan sensor yang ditempatkan pada wahana tertentu, seperti satelit atau pesawat [4]. Keunggulan teknologi ini terletak pada kemampuannya mengamati wilayah yang luas secara efisien sekaligus menghasilkan data yang konsisten sepanjang waktu.

Dalam konteks kajian lingkungan dan perubahan tutupan lahan, penginderaan jauh memainkan peran yang sangat strategis karena sanggup menghadirkan informasi spasial dan temporal tentang kondisi permukaan bumi. Data citra yang dihasilkan melalui proses ini dapat dimanfaatkan untuk berbagai keperluan analisis, mulai dari pemantauan vegetasi, pemetaan tutupan lahan, hingga deteksi perubahan ekosistem [11]. Seiring meningkatnya resolusi sensor satelit dan berkembangnya teknologi pengolahan data, penginderaan jauh kini menjadi salah satu tulang punggung penelitian geospasial kontemporer.

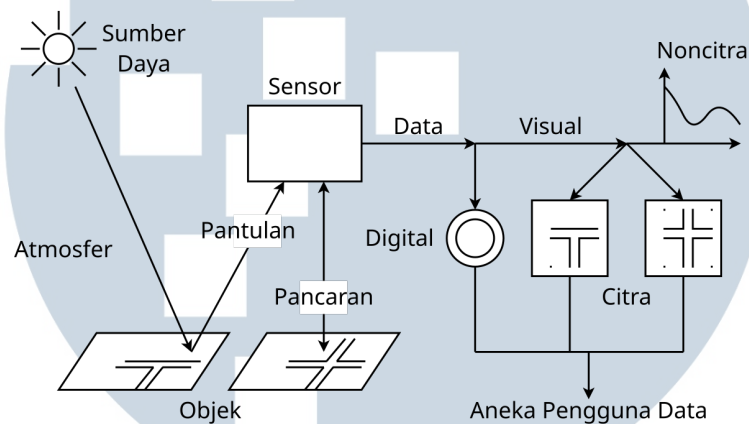
Cara kerja penginderaan jauh bertumpu pada interaksi antara energi elektromagnetik dan objek-objek di permukaan bumi. Radiasi dari sumber energi, umumnya matahari, akan menyinari objek di permukaan, lalu sebagian energinya dipantulkan kembali. Pantulan energi tersebut ditangkap oleh sensor satelit dan dikonversi menjadi data digital yang selanjutnya diolah menjadi citra [4]. Perbedaan karakteristik pantulan dari tiap-tiap objek inilah yang menjadi dasar identifikasi dan pengelompokan objek berdasarkan informasi spektralnya.

Sistem penginderaan jauh tersusun atas beberapa elemen yang bekerja secara sinergis. Elemen pertama adalah sumber energi, yang pada umumnya berasal dari radiasi matahari. Energi ini merambat melewati atmosfer sebelum akhirnya berinteraksi dengan objek di permukaan bumi. Atmosfer sendiri berperan sebagai medium yang memengaruhi perambatan energi elektromagnetik melalui proses penyerapan, hamburan, dan transmisi [3].

Energi yang telah berinteraksi dengan permukaan bumi kemudian dipantulkan menuju sensor pada platform satelit. Sensor merekam sinyal tersebut dalam format digital yang kemudian ditransmisikan ke stasiun penerima di darat. Data mentah yang diterima selanjutnya diproses menjadi produk citra yang siap dianalisis untuk berbagai keperluan, termasuk klasifikasi tutupan lahan dan

pemantauan perubahan lingkungan [11].

Dalam penelitian ini, sumber data penginderaan jauh yang digunakan adalah citra Sentinel-2 dengan kemampuan multispektral pada resolusi spasial hingga 10 meter. Citra tersebut memungkinkan analisis tutupan lahan yang lebih terperinci serta mendukung proses klasifikasi berbasis *machine learning* untuk mendeteksi perubahan tutupan hutan di wilayah kajian.



Gambar 2.1. Diagram sistem penginderaan jauh (Sumber: Sutanto, 1986)

2.3 Citra Satelit

Citra satelit adalah representasi digital permukaan bumi yang dihasilkan dari rekaman energi elektromagnetik oleh sensor yang terpasang pada wahana satelit. Sensor tersebut menangkap energi yang dipantulkan atau dipancarkan oleh objek di permukaan bumi, kemudian mengonversinya menjadi data digital yang dapat ditampilkan secara visual sebagai citra. Setiap piksel dalam citra satelit menyimpan nilai spektral tertentu yang mencerminkan sifat fisik objek di permukaan bumi [4].

Dalam ranah penginderaan jauh, citra satelit dimanfaatkan untuk mengidentifikasi dan memetakan berbagai kondisi di permukaan bumi, mulai dari jenis tutupan lahan, sebaran vegetasi, keberadaan badan air, hingga perubahan ekosistem. Kemajuan teknologi sensor telah mendorong peningkatan detail informasi yang dapat diperoleh dari citra satelit, sehingga penggunaannya merambah ke berbagai sektor seperti pemantauan lingkungan, perencanaan tata ruang, pertanian presisi, dan analisis dinamika lahan [11].

Kualitas citra satelit ditentukan oleh empat jenis resolusi yang menjadi parameter fundamental, yaitu resolusi spasial, spektral, temporal, dan radiometrik

[3]. Masing-masing resolusi menggambarkan aspek berbeda dari kemampuan sensor dalam merekam informasi.

Resolusi spasial merujuk pada luas area di permukaan bumi yang diwakili oleh satu piksel. Semakin kecil area yang direpresentasikan per piksel, semakin tinggi tingkat kedetailan citra tersebut, sehingga objek-objek di permukaan bumi dapat dikenali dengan lebih baik.

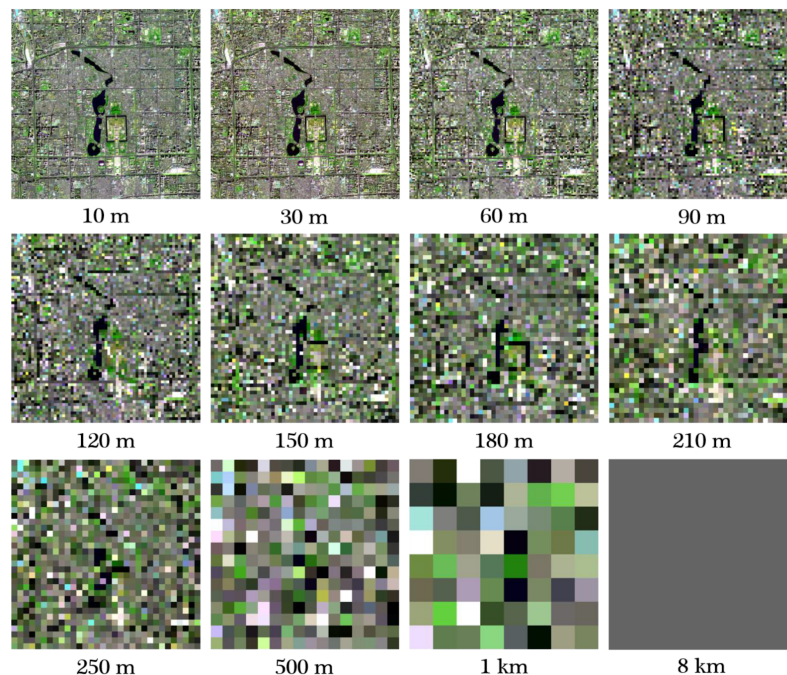
Resolusi spektral mencerminkan kemampuan sensor dalam membedakan panjang gelombang elektromagnetik. Sensor multispektral dapat merekam data pada sejumlah saluran spektral sekaligus, seperti saluran merah, hijau, biru, dan inframerah. Keragaman informasi spektral ini sangat krusial dalam penginderaan jauh karena setiap jenis objek memiliki pola pantulan yang khas.

Resolusi temporal menggambarkan seberapa sering satelit merekam kembali suatu wilayah. Semakin singkat selang waktu antar perekaman, semakin leluasa pemantauan perubahan kondisi suatu kawasan dilakukan secara berkala.

Resolusi radiometrik menunjukkan kemampuan sensor dalam membedakan gradasi intensitas energi yang diterima, biasanya dinyatakan dalam jumlah bit per piksel. Semakin tinggi bit depth-nya, semakin halus perbedaan yang dapat dideteksi, menghasilkan citra dengan kekayaan informasi yang lebih tinggi.

Dalam penelitian ini, Sentinel-2 dipilih sebagai sumber data citra utama karena menawarkan resolusi spasial hingga 10 meter pada beberapa saluran spektral dan resolusi temporal sekitar lima hari. Kombinasi karakteristik tersebut menjadikannya sangat andal untuk analisis tutupan lahan dan deteksi perubahan hutan di wilayah kajian.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A



Gambar 2.2. Perbandingan citra optik *false color* pada resolusi spasial berbeda (Sumber: Tian dkk., 2020)

2.4 Sentinel-2

Sentinel-2 adalah satelit penginderaan jauh yang dirancang dan diluncurkan oleh *European Space Agency* (ESA) dalam kerangka program Copernicus Uni Eropa. Program Copernicus dibangun dengan tujuan menyediakan data observasi bumi secara bebas dan terbuka untuk mendukung keperluan pemantauan lingkungan, pengelolaan sumber daya alam, serta pengkajian dinamika tutupan lahan [4]. Sentinel-2 secara spesifik dikembangkan untuk menghasilkan citra multispektral beresolusi tinggi yang aplikatif untuk berbagai keperluan penginderaan jauh, termasuk penilaian kondisi vegetasi, pemetaan tutupan lahan, dan pemantauan perubahan ekosistem.

Satelit Sentinel-2 beroperasi dalam konstelasi dua unit, yakni Sentinel-2A dan Sentinel-2B, yang mengorbit pada jalur yang sama dengan jeda waktu tertentu. Sinergi keduanya memungkinkan interval perekaman ulang (*revisit time*) sekitar lima hari untuk suatu wilayah. Hal ini menjadikan Sentinel-2 sangat efisien dalam memantau perubahan lingkungan secara periodik.

Sensor utama yang digunakan pada Sentinel-2 adalah *MultiSpectral Instrument* (MSI), yang mampu merekam pada 13 saluran spektral mencakup spektrum tampak (*visible*), inframerah dekat (*near infrared*), dan inframerah

gelombang pendek (*shortwave infrared*). Tiap saluran memiliki resolusi spasial yang berbeda-beda, yakni 10, 20, dan 60 meter, disesuaikan dengan fungsi masing-masing dalam menganalisis karakteristik permukaan bumi.

Saluran-saluran dengan resolusi 10 meter merupakan yang paling detail dan mencakup saluran biru, hijau, merah, serta inframerah dekat yang lazim digunakan dalam analisis vegetasi dan klasifikasi tutupan lahan. Tabel 2.2 merangkum spesifikasi saluran utama Sentinel-2.

Tabel 2.2. Beberapa band utama Sentinel-2 dan resolusi spasialnya

Band	Spektrum	Resolusi Spasial
B2	Biru (<i>Blue</i>)	10 m
B3	Hijau (<i>Green</i>)	10 m
B4	Merah (<i>Red</i>)	10 m
B8	Inframerah Dekat (<i>Near Infrared</i>)	10 m
B11	Inframerah Gelombang Pendek (<i>SWIR-1</i>)	20 m
B12	Inframerah Gelombang Pendek (<i>SWIR-2</i>)	20 m

Salah satu keunggulan kompetitif Sentinel-2 adalah aksesibilitas datanya yang sepenuhnya terbuka, dapat diunduh secara cuma-cuma oleh siapapun melalui berbagai platform seperti Copernicus Open Access Hub maupun Google Earth Engine [4]. Keterbukaan ini secara signifikan mendorong pemanfaatan Sentinel-2 dalam komunitas penelitian dan pengembangan sistem berbasis data geospasial.

Dalam penelitian ini, produk Sentinel-2 Level-2A yang telah terkoreksi atmosfer digunakan sebagai sumber data primer, sehingga nilai pikselnya merepresentasikan reflektansi permukaan (*surface reflectance*) secara konsisten antar periode. Mengingat cakupan wilayah Provinsi Banten yang luas dan kendala verifikasi lapangan secara komprehensif, penelitian ini menggunakan dataset referensi global ESA WorldCover 10m sebagai *digital ground truth*. Dataset ini memiliki resolusi spasial yang selaras dengan Sentinel-2 dan telah memperoleh validasi di tingkat internasional [9]. Data citra tersebut kemudian dijadikan basis ekstraksi fitur spektral dan indeks vegetasi sebagai masukan model klasifikasi Random Forest.

2.5 Google Earth Engine

Google Earth Engine (GEE) adalah platform berbasis komputasi awan (*cloud computing*) yang dikembangkan khusus untuk menangani analisis data geospasial dalam skala masif atau *planetary-scale*. Platform ini menggabungkan katalog arsip citra satelit multisensor yang mencakup rentang waktu puluhan tahun dengan kapasitas komputasi paralel berperforma tinggi, sehingga mampu menjalankan pemrosesan data dalam skala yang mustahil dilakukan pada perangkat lokal biasa [13].

Dari sudut pandang Informatika, GEE menyediakan antarmuka pemrograman berbasis JavaScript maupun Python API yang memungkinkan pembangunan *pipeline* pemrosesan geospasial secara otomatis dan terstruktur. Kemampuan tersebut menempatkan GEE tidak hanya sebagai instrumen analisis lingkungan, tetapi juga sebagai infrastruktur komputasi untuk riset berbasis data (*data-driven research*) di bidang penginderaan jauh. Studi-studi terkini menunjukkan bahwa penggabungan GEE dengan algoritma *machine learning* seperti *Random Forest* berhasil menghasilkan sistem klasifikasi yang efisien dan mampu diskalakan [14, 15].

GEE telah menjadi platform dominan dalam penelitian penginderaan jauh berkat sejumlah keunggulan teknis yang ditawarkannya. Pertama, tersedianya akses langsung ke arsip citra Sentinel-2, Landsat, MODIS, dan berbagai produk turunannya tanpa perlu mengunduh data secara lokal. Kedua, dukungan operasi *map-reduce* untuk komputasi paralel pada dataset geospasial berskala besar. Ketiga, integrasinya dengan algoritma *machine learning* memungkinkan pembangunan model klasifikasi yang *reproducible* dan mudah direplikasi di berbagai wilayah [12].

Dalam penelitian ini, GEE difungsikan untuk tiga keperluan utama:

1. Mengakses dan memfilter koleksi citra Sentinel-2 Level-2A untuk periode 2019, 2021, dan 2023.
2. Melakukan *preprocessing* citra meliputi penyaringan awan, pembentukan komposit median, dan koreksi geometri.
3. Mengekstraksi label kelas tutupan lahan dari dataset ESA WorldCover sebagai *digital ground truth* untuk pelatihan model.

Setelah diproses, data diekspor ke lingkungan komputasi lokal untuk tahap pemodelan dan evaluasi *Random Forest*, sehingga analisis performa model dapat dilakukan dengan lebih leluasa dan terkontrol.

2.6 Klasifikasi Citra Satelit

Klasifikasi citra satelit adalah proses pengelompokan piksel berdasarkan kemiripan karakteristik spektralnya ke dalam kategori tertentu, seperti hutan, lahan pertanian, badan air, atau kawasan terbangun. Proses ini menganalisis nilai spektral yang terekam pada setiap saluran citra untuk mengenali pola perbedaan antar objek di permukaan bumi [4].

Dalam alur pengolahan citra penginderaan jauh, tahap klasifikasi memegang peranan krusial karena berfungsi mengubah data numerik piksel menjadi informasi tematik yang bermakna. Luaran proses ini umumnya berupa peta tutupan lahan yang memperlihatkan sebaran spasial berbagai kelas objek pada suatu wilayah.

Sebagai contoh, pada citra Sentinel-2, vegetasi umumnya menunjukkan nilai reflektansi tinggi pada saluran inframerah dekat (*Near Infrared*), sementara badan air cenderung memiliki reflektansi rendah di hampir semua saluran. Perbedaan karakteristik inilah yang menjadi landasan bagi algoritma klasifikasi untuk membedakan jenis-jenis tutupan lahan.

Secara garis besar, metode klasifikasi citra terbagi menjadi dua pendekatan utama, yaitu *supervised classification* dan *unsupervised classification*.

1. Supervised Classification

Supervised classification adalah pendekatan klasifikasi yang memanfaatkan data latih (*training data*) berlabel untuk memandu proses pengelompokan. Pengguna terlebih dahulu menentukan sekumpulan piksel sampel yang mewakili setiap kelas tutupan lahan, kemudian algoritma mempelajari pola spektral dari sampel tersebut untuk mengklasifikasikan piksel lain yang memiliki kemiripan karakteristik [4].

Sebagai ilustrasi, dalam klasifikasi tutupan lahan menggunakan Sentinel-2, sampel piksel yang mewakili kelas hutan, badan air, lahan pertanian, dan permukiman ditentukan terlebih dahulu. Algoritma kemudian mempelajari ciri spektral tiap kelas dan menerapkannya untuk mengklasifikasikan seluruh piksel citra berdasarkan kemiripan polanya.

Pendekatan ini banyak diadopsi dalam penelitian penginderaan jauh karena menghasilkan klasifikasi yang lebih presisi, terutama ketika sampel latih benar-benar merepresentasikan kondisi lapangan. Selain itu, metode ini kompatibel dengan berbagai algoritma *machine learning* seperti Random Forest, *Support Vector Machine*, maupun *Neural Network* untuk meningkatkan akurasi klasifikasi [6].

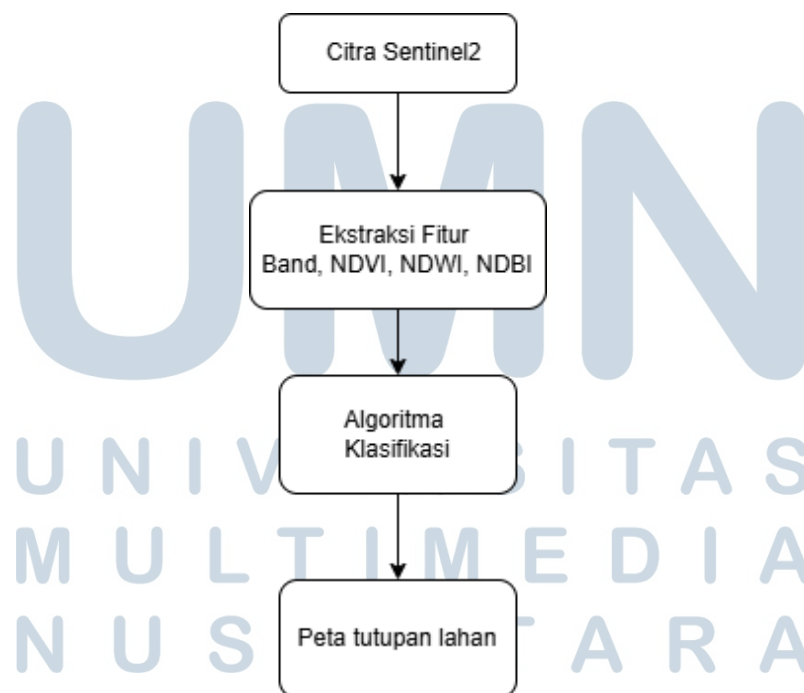
2. Unsupervised Classification

Unsupervised classification merupakan pendekatan yang tidak memerlukan data latih berlabel. Algoritma secara mandiri mengelompokkan piksel berdasarkan kesamaan nilai spektralnya, dan pengguna kemudian menginterpretasikan kelompok-kelompok yang terbentuk untuk menentukan kelas tutupan lahan yang sesuai.

Contohnya, algoritma K-Means dapat mempartisi piksel citra ke dalam sejumlah klaster berdasarkan kedekatan spektral. Setelah klaster terbentuk, pengguna mengidentifikasi masing-masing klaster sebagai kelas tutupan tertentu, seperti vegetasi, air, atau kawasan terbangun.

Metode ini lazim digunakan pada tahap eksplorasi awal ketika informasi referensi mengenai kelas objek belum tersedia. Kelemahannya, hasil pengelompokan tidak selalu mencerminkan kelas tutupan yang dikehendaki.

Dalam penelitian ini, *supervised classification* dipilih karena memungkinkan penggunaan data latih yang secara eksplisit merepresentasikan kelas hutan dan non-hutan, sehingga model dapat mengenali karakteristik spektral kedua kelas secara lebih akurat. Pendekatan ini juga mendukung implementasi algoritma *machine learning*, khususnya Random Forest, untuk menghasilkan model yang lebih *robust* dalam mendeteksi perubahan tutupan lahan.



Gambar 2.3. Ilustrasi proses klasifikasi citra satelit (Sumber: Diolah penulis)

2.7 Machine Learning dalam Penginderaan Jauh

Machine learning adalah cabang kecerdasan buatan yang berkonsentrasi pada perancangan algoritma yang dapat belajar dari data dan menghasilkan prediksi tanpa harus diprogram secara eksplisit [16]. Dalam konteks geospasial, *machine learning* banyak diterapkan untuk mengolah data citra satelit berkat kemampuannya menangani data berdimensi tinggi dan pola yang kompleks.

Dalam penginderaan jauh, tiap piksel citra dapat dipandang sebagai vektor fitur yang memuat nilai spektral dari berbagai saluran. Pada citra Sentinel-2 misalnya, setiap piksel memiliki nilai reflektansi pada saluran biru, hijau, merah, dan inframerah dekat. Dalam penelitian ini, vektor fitur tiap piksel berdimensi 9, terdiri dari enam saluran spektral Sentinel-2 (B2, B3, B4, B8, B11, B12) ditambah tiga indeks turunan (NDVI, NDWI, NDBI). Vektor tersebut menjadi masukan bagi algoritma *machine learning* untuk mengklasifikasikan tutupan lahan.

Secara formal, persoalan klasifikasi dalam *machine learning* dapat dinyatakan sebagai fungsi pemetaan:

$$f : X \rightarrow Y \quad (2.2)$$

di mana $X \in \mathbb{R}^9$ merupakan ruang fitur yang berisi nilai spektral dan indeks vegetasi dari citra Sentinel-2, sedangkan $Y \in \{0, 1\}$ merepresentasikan kelas biner tutupan lahan yang hendak diprediksi, yaitu non-hutan (0) dan hutan (1).

Dibandingkan metode statistik konvensional yang umumnya mengasumsikan distribusi data tertentu, algoritma *machine learning* lebih adaptif dalam memodelkan relasi non-linear antar fitur. Kemampuannya menangani ruang fitur berdimensi tinggi dan menghasilkan generalisasi yang baik pada data baru menjadikannya pilihan utama dalam analisis citra [6].

Beberapa algoritma *machine learning* yang umum digunakan dalam klasifikasi citra penginderaan jauh antara lain:

1. Random Forest

Random Forest merupakan algoritma *ensemble learning* yang membangun kumpulan pohon keputusan (*decision tree*) secara acak dan menggabungkan prediksinya melalui mekanisme pemungutan suara (*voting*). Metode ini dikenal handal dalam klasifikasi citra penginderaan jauh berkat kemampuannya menangani data berdimensi tinggi dan ketahanannya terhadap *noise* [5].

2. Support Vector Machine

Support Vector Machine (SVM) bekerja dengan mencari bidang pemisah

(*hyperplane*) optimal yang memisahkan kelas-kelas dalam ruang fitur. Algoritma ini banyak digunakan dalam penginderaan jauh karena performanya yang baik meski dengan jumlah sampel terbatas namun dimensi fitur tinggi [17].

3. Artificial Neural Network

Artificial Neural Network (ANN) merupakan model komputasi yang terinspirasi dari jaringan neuron biologis, tersusun atas lapisan-lapisan neuron yang saling terhubung. Model ini mampu menangkap pola yang sangat kompleks dan kerap digunakan dalam klasifikasi citra maupun analisis berbasis *deep learning* [18].

Pemilihan Random Forest dalam penelitian ini dilandasi tiga pertimbangan: kemampuannya menangani data multispektral berdimensi tinggi, risiko *overfitting* yang rendah berkat kombinasi banyak pohon keputusan, serta kemampuannya menghasilkan informasi *feature importance* yang memudahkan analisis pengaruh fitur spektral dalam proses klasifikasi [14].

2.8 Algoritma Random Forest

Random Forest adalah algoritma *ensemble learning* yang membangun sejumlah pohon keputusan (*decision tree*) secara independen melalui teknik *Bagging* dan memilih hasil akhir melalui mekanisme *majority voting* [5].

2.8.1 Parameter Random Forest

Kinerja algoritma Random Forest sangat dipengaruhi oleh konfigurasi parameternya (*hyperparameters*). Parameter-parameter utama yang digunakan dalam penelitian ini meliputi:

1. **n_estimators**: menentukan jumlah pohon keputusan yang dibangun. Semakin banyak pohon, model cenderung lebih stabil dan risiko *overfitting* berkurang, namun beban komputasi meningkat.
2. **max_depth**: menetapkan kedalaman maksimum tiap pohon. Nilai terlalu besar berpotensi menyebabkan *overfitting*, sedangkan nilai terlalu kecil dapat melemahkan kemampuan model menangkap pola dalam data.
3. **max_features**: membatasi jumlah fitur yang dipertimbangkan secara acak pada tiap pemisahan (*splitting*) simpul.

4. **random_state**: nilai tetap untuk menjamin reproduktibilitas hasil eksperimen (*reproducibility*).

Nilai optimal untuk setiap parameter dapat ditentukan melalui teknik seperti *grid search* atau *Bayesian optimization*, yang terbukti meningkatkan akurasi klasifikasi secara signifikan [19].

2.8.2 Konsep Ensemble Learning

Fondasi utama Random Forest adalah konsep *ensemble learning*, yakni strategi pembelajaran yang menggabungkan sejumlah model dasar (*base learners*) untuk menghasilkan prediksi yang lebih andal. Pada Random Forest, model dasar yang digunakan adalah *decision tree*, di mana setiap pohon dilatih secara independen pada subset data berbeda, kemudian keluarannya digabungkan untuk menghasilkan keputusan kolektif.

Apabila setiap model ke- i menghasilkan prediksi $h_i(x)$, maka prediksi akhir model *ensemble* dinyatakan sebagai:

$$H(x) = \text{majority vote}\{h_1(x), h_2(x), \dots, h_n(x)\} \quad (2.3)$$

di mana $H(x)$ adalah keputusan akhir dan $h_i(x)$ adalah prediksi pohon ke- i .

2.8.3 Decision Tree

Decision tree merupakan komponen pembentuk Random Forest sekaligus model klasifikasi yang bekerja dengan mempartisi data secara rekursif berdasarkan nilai fitur-fitur tertentu hingga diperoleh keputusan akhir berupa label kelas. Struktur *decision tree* tersusun atas tiga jenis simpul:

- **Root node**: simpul pangkal yang memuat keseluruhan data pelatihan.
- **Internal node**: simpul percabangan yang melakukan pemisahan data berdasarkan nilai suatu fitur.
- **Leaf node**: simpul ujung yang menyimpan hasil klasifikasi akhir.

Dalam proses pembangunan pohon, algoritma memilih fitur pemisah terbaik di setiap simpul. Salah satu kriteria yang umum digunakan adalah *Gini impurity*, yang dihitung sebagai:

$$Gini = 1 - \sum_{i=1}^C p_i^2 \quad (2.4)$$

dengan p_i menyatakan probabilitas kelas ke- i pada simpul tersebut dan C adalah jumlah kelas.

2.8.4 Bootstrap Sampling dan Bagging

Random Forest menerapkan teknik *bootstrap sampling* dalam kerangka metode *bagging* (*bootstrap aggregating*). Pada teknik ini, subset data pelatihan dibentuk dengan cara mengambil sampel secara acak dengan pengembalian (*sampling with replacement*).

Jika D menyatakan keseluruhan dataset pelatihan, proses bootstrap menghasilkan sejumlah subset $D_1, D_2, \dots, D_n$ yang masing-masing digunakan untuk melatih satu *decision tree*. Penggunaan subset berbeda pada setiap pohon mendorong keberagaman antar pohon dan mengurangi korelasi di antara model-model tersebut.

2.8.5 Pemilihan Fitur Secara Acak

Di samping *bootstrap sampling*, Random Forest juga menerapkan seleksi fitur secara stokastik pada setiap pemisahan simpul. Hanya sebagian kecil fitur yang dipilih secara acak untuk dijadikan kandidat dalam setiap proses *splitting*.

Mekanisme ini bertujuan memperkecil korelasi antar pohon sehingga meningkatkan kemampuan generalisasi model secara keseluruhan. Akibatnya, meskipun dilatih pada dataset yang sama, setiap pohon keputusan akan memiliki struktur yang berbeda.

2.8.6 Mekanisme Voting

Setelah seluruh pohon menghasilkan prediksinya masing-masing, Random Forest menentukan keputusan akhir melalui mekanisme *majority voting*. Setiap pohon memberikan satu suara untuk kelas tertentu, dan kelas yang memperoleh suara terbanyak ditetapkan sebagai hasil klasifikasi.

Untuk T pohon keputusan, prediksi akhir dinyatakan sebagai:

$$\hat{y} = \text{mode}\{T_1(x), T_2(x), \dots, T_T(x)\} \quad (2.5)$$

dengan $\hat{y}$ sebagai label prediksi akhir dan $T_i(x)$ sebagai prediksi pohon ke- i .

2.8.7 Kelebihan Random Forest

Random Forest memiliki sejumlah keunggulan yang menjelaskan popularitasnya dalam klasifikasi citra penginderaan jauh:

- **Tahan terhadap noise:** prediksi yang dihasilkan dari agregasi banyak pohon membuat model lebih tahan terhadap data yang mengandung gangguan.
- **Risiko overfitting rendah:** pelatihan pada subset data yang beragam membantu meningkatkan kemampuan generalisasi model.
- **Cocok untuk data berdimensi tinggi:** algoritma ini sangat sesuai untuk data multispektral dengan banyak saluran dan indeks turunan.
- **Menyediakan feature importance:** Random Forest mampu menghitung tingkat kontribusi tiap fitur terhadap proses klasifikasi [14].

Berkat keunggulan-keunggulan tersebut, Random Forest banyak diandalkan dalam penelitian penginderaan jauh untuk pemetaan tutupan lahan, pemantauan vegetasi, dan deteksi perubahan lingkungan berbasis citra satelit [6].

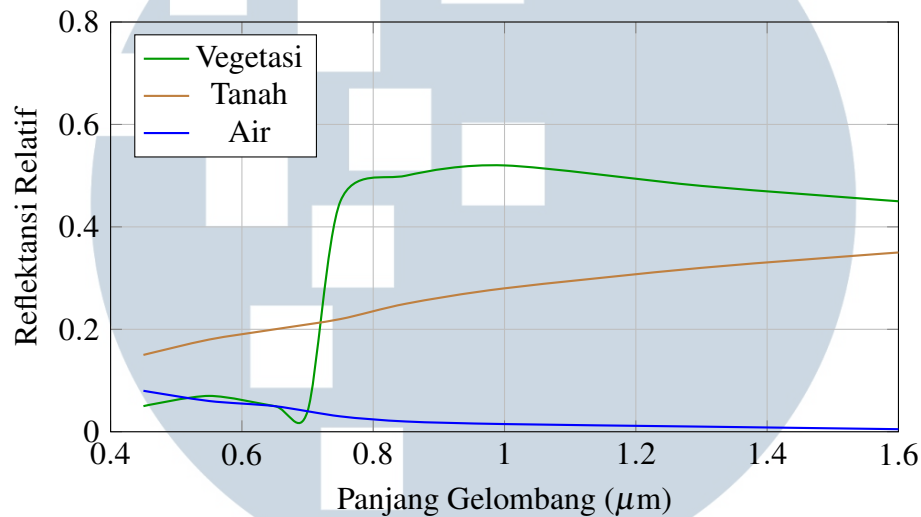
2.9 Indeks Spektral Vegetasi

Indeks spektral merupakan transformasi matematis yang mengombinasikan nilai reflektansi dari beberapa saluran spektral citra satelit guna menyoroti karakteristik spesifik permukaan bumi, seperti kondisi vegetasi, keberadaan air, atau kepadatan bangunan. Kombinasi matematis ini mampu meningkatkan kontras antar jenis objek sehingga memudahkan interpretasi dan klasifikasi citra [4].

Indeks spektral juga berperan meredam pengaruh variasi kondisi pencahayaan, bayangan, dan efek atmosfer residual, sehingga informasi yang diekstraksi dari citra menjadi lebih stabil dan konsisten antar waktu.

Pada citra multispektral seperti Sentinel-2, setiap jenis objek di permukaan bumi memperlihatkan profil reflektansi yang khas. Vegetasi misalnya, umumnya memantulkan energi tinggi pada inframerah dekat namun rendah pada saluran merah, sedangkan badan air umumnya memiliki reflektansi rendah di hampir seluruh panjang gelombang. Kekhasan profil spektral inilah yang dieksploitasi oleh indeks-indeks spektral untuk mengekstraksi informasi tertentu dari citra.

Dalam penelitian ini, tiga indeks spektral digunakan sebagai fitur tambahan: *Normalized Difference Vegetation Index* (NDVI), *Normalized Difference Water Index* (NDWI), dan *Normalized Difference Built-up Index* (NDBI), yang masing-masing membantu membedakan karakteristik vegetasi, perairan, dan kawasan terbangun.



Gambar 2.4. Kurva reflektansi spektral beberapa objek permukaan bumi (Sumber: Diolah penulis)

2.9.1 Normalized Difference Vegetation Index (NDVI)

NDVI adalah indeks vegetasi yang paling banyak digunakan dalam penginderaan jauh untuk menilai keberadaan dan kondisi vegetasi. Indeks ini memanfaatkan kontras antara reflektansi tinggi pada saluran inframerah dekat (NIR) dan reflektansi rendah pada saluran merah (Red) yang menjadi ciri khas vegetasi sehat [20].

Formulasi NDVI adalah sebagai berikut:

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (2.6)$$

Pada Sentinel-2, saluran yang digunakan adalah Band 8 (NIR) dan Band 4 (Red), dengan rentang nilai NDVI dari -1 hingga 1 . Interpretasi nilai NDVI disajikan pada Tabel 2.3.

Tabel 2.3. Interpretasi umum nilai NDVI

Rentang NDVI	Interpretasi
< 0	Air, awan, atau permukaan non-vegetasi
0 – 0,2	Tanah terbuka atau vegetasi sangat jarang
0,2 – 0,5	Vegetasi sedang
0,5 – 1,0	Vegetasi rapat dan sehat

Sumber: Tucker [20]

Kemampuan NDVI dalam mendeteksi vegetasi menjadikannya banyak dimanfaatkan dalam kajian pemantauan hutan, pertanian presisi, dan analisis dinamika tutupan lahan [8].

2.9.2 Normalized Difference Water Index (NDWI)

NDWI adalah indeks spektral yang dikembangkan untuk mendeteksi keberadaan air di permukaan bumi dengan cara memaksimalkan kontras antara badan air dan objek lain seperti vegetasi atau tanah [21].

$$NDWI = \frac{Green - NIR}{Green + NIR} \quad (2.7)$$

Pada Sentinel-2, Band 3 (Green) dan Band 8 (NIR) digunakan dalam perhitungannya. Nilai NDWI yang tinggi mengindikasikan keberadaan badan air, sedangkan nilai rendah mencerminkan daratan atau vegetasi. Indeks ini banyak diaplikasikan dalam analisis hidrologi, pemetaan badan air, dan pemantauan perubahan wilayah perairan.

2.9.3 Normalized Difference Built-up Index (NDBI)

NDBI merupakan indeks spektral yang difungsikan untuk mengidentifikasi kawasan terbangun atau perkotaan, dengan memanfaatkan perbedaan respons spektral antara saluran inframerah gelombang pendek (*Short Wave Infrared*) dan inframerah dekat [22].

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR} \quad (2.8)$$

Pada Sentinel-2, Band 11 (SWIR) dan Band 8 (NIR) digunakan dalam perhitungannya. Nilai NDBI tinggi mengindikasikan kawasan terbangun seperti permukiman dan jalan. Indeks ini sangat berguna dalam analisis ekspansi perkotaan dan perubahan penggunaan lahan.

2.9.4 Peran Indeks Spektral dalam Penelitian

Dalam penelitian ini, ketiga indeks spektral (NDVI, NDWI, NDBI) diintegrasikan sebagai fitur tambahan dalam proses klasifikasi menggunakan Random Forest. Tujuannya adalah meningkatkan kapasitas model dalam membedakan objek-objek yang karakteristik spektral band mentahnya saling tumpang tindih.

Dengan menginkorporasikan indeks spektral ke dalam vektor fitur, model tidak hanya mengandalkan nilai reflektansi mentah, tetapi juga memanfaatkan transformasi spektral yang secara lebih eksplisit merepresentasikan karakteristik vegetasi, perairan, dan kawasan terbangun. Pendekatan ini diharapkan meningkatkan akurasi klasifikasi sekaligus memperkuat kemampuan model dalam mendeteksi perubahan tutupan lahan [14].

2.10 Evaluasi Model Klasifikasi

Evaluasi model klasifikasi merupakan tahapan kritis dalam siklus pembelajaran mesin yang bertujuan mengukur sejauh mana model mampu melakukan prediksi secara tepat. Dalam penelitian klasifikasi citra penginderaan jauh, evaluasi dilakukan dengan membandingkan label prediksi model terhadap label referensi yang sesungguhnya. Instrumen evaluasi yang paling umum digunakan adalah *confusion matrix* [23].

Confusion matrix adalah tabel yang memperlihatkan distribusi hasil prediksi model relatif terhadap label sebenarnya. Pada kasus klasifikasi biner, tabel ini memuat empat elemen: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN).

Tabel 2.4. Struktur *Confusion Matrix* untuk klasifikasi dua kelas

Aktual / Prediksi	Positif	Negatif
Positif	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
Negatif	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Sumber: Diolah penulis

Makna masing-masing elemen adalah sebagai berikut:

- **True Positive (TP)**: jumlah sampel kelas positif yang berhasil diprediksi dengan benar.
- **True Negative (TN)**: jumlah sampel kelas negatif yang berhasil diprediksi dengan benar.
- **False Positive (FP)**: jumlah sampel kelas negatif yang keliru diprediksi sebagai positif.
- **False Negative (FN)**: jumlah sampel kelas positif yang keliru diprediksi sebagai negatif.

Dari keempat elemen tersebut, sejumlah metrik kinerja dapat diturunkan. Dalam penelitian ini, seluruh metrik dihitung menggunakan *weighted average* karena dataset telah diseimbangkan melalui *undersampling*. Metrik yang digunakan mencakup *accuracy*, *precision*, *recall*, *F1-score*, dan koefisien kappa.

2.10.1 Accuracy

Accuracy mengukur proporsi prediksi yang tepat dari keseluruhan data yang dievaluasi.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.9)$$

Nilai *accuracy* berkisar antara 0 dan 1, di mana nilai lebih tinggi mencerminkan kinerja model yang lebih baik.

2.10.2 Precision

Precision mengukur seberapa besar proporsi prediksi positif yang benar-benar tepat dari seluruh prediksi positif yang dihasilkan model.

$$Precision = \frac{TP}{TP + FP} \quad (2.10)$$

Nilai *precision* yang tinggi menandakan rendahnya tingkat kesalahan dalam mengidentifikasi kelas positif.

2.10.3 Recall

Recall mengukur kemampuan model dalam mendeteksi seluruh instance yang sesungguhnya termasuk kelas positif.

$$Recall = \frac{TP}{TP + FN} \quad (2.11)$$

Nilai *recall* tinggi mengindikasikan bahwa model jarang melewatkan sampel positif yang sebenarnya.

2.10.4 F1-Score

F1-score merupakan rata-rata harmonik antara *precision* dan *recall*, yang menyeimbangkan keduanya dalam satu nilai tunggal. Metrik ini sangat berguna ketika distribusi kelas tidak merata.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.12)$$

Nilai F1-score yang tinggi menandakan keseimbangan yang baik antara ketepatan dan kelengkapan deteksi.

2.10.5 Kappa Coefficient

Koefisien kappa adalah metrik yang mengukur tingkat kesepakatan antara prediksi model dan data referensi dengan memperhitungkan kemungkinan kesesuaian yang terjadi secara acak (*chance agreement*) [24].

$$\kappa = \frac{P_o - P_e}{1 - P_e} \quad (2.13)$$

di mana P_o adalah tingkat kesesuaian yang diamati dan P_e adalah tingkat kesesuaian yang diharapkan secara acak.

Nilai kappa berkisar antara -1 dan 1 . Nilai mendekati 1 menunjukkan kesepakatan yang sangat kuat antara prediksi dan referensi, sementara nilai mendekati 0 mengindikasikan kesesuaian yang tidak lebih baik dari tebakan acak [25].

Dalam penginderaan jauh, koefisien kappa lebih disukai dibandingkan akurasi biasa karena memberikan ukuran kesepakatan yang lebih andal [26]. Dalam penelitian ini, kelima metrik tersebut digunakan bersama-sama untuk menilai kinerja model Random Forest secara menyeluruh.



2.11 Kesenjangan Penelitian (*Research Gap*)

Tabel 2.5. Penelitian Terdahulu

No	Judul Penelitian	Metode	Hasil Penelitian	Celah Penelitian
1	Estimating Land Cover Change Using Sentinel-2 Satellite Imagery in the Nuraksa Forest Park Area (<i>Subarnas dkk., 2025</i>)	Supervised Classification (Maximum Likelihood), Sentinel-2	Penurunan hutan primer 669,32 ha (2016–2023); peningkatan hutan sekunder dan pertanian; akurasi 86,67%	Tidak menggunakan metode <i>machine learning</i> modern seperti Random Forest; masih menggunakan metode klasifikasi konvensional
2	Prediksi Perubahan Tutupan Lahan di Kabupaten Bogor Tahun 2026 (<i>Rohman dkk., 2025</i>)	Random Forest, Sentinel-2, Time-series Cross Validation	Peningkatan area terbangun dan penurunan lahan pertanian; akurasi 87,96%	Fokus pada prediksi masa depan; belum melakukan analisis perubahan aktual berbasis evaluasi multi-metrik yang komprehensif
3	Machine Learning Detection of Tropical Rain Forest Deforestation (<i>Iskandar & Hanafi, 2022</i>)	Random Forest, CART, SVM; Landsat dan Sentinel-2	Deforestasi 180.073 ha; akurasi terbaik pada Random Forest (OA 95%, Kappa 94%)	Tidak membahas perubahan per kelas tutupan lahan secara multi-temporal; hanya berfokus pada deteksi deforestasi biner
4	Klasifikasi Tutupan Lahan dengan NDVI dan Random Forest (<i>Marlina, 2022</i>)	NDVI, Random Forest, Sentinel-2	Akurasi klasifikasi 91,39%; NDVI terbukti efektif untuk analisis vegetasi	Tidak membahas perubahan tutupan lahan multitemporal; hanya klasifikasi satu periode waktu tanpa analisis <i>feature importance</i>
5	Utilizing Multitemporal Indices and Spectral Bands of Sentinel-2 to Enhance LULC Classification with RF and SVM (<i>Arfa & Minaei, 2024</i>)	Random Forest, SVM; indeks multitemporal Sentinel-2	Indeks multitemporal Sentinel-2 meningkatkan akurasi klasifikasi LULC secara signifikan pada kedua algoritma	Tidak membahas klasifikasi biner hutan/non-hutan untuk deteksi deforestasi; evaluasi terbatas pada <i>overall accuracy</i> tanpa <i>precision</i> , <i>recall</i> , <i>F1-score</i> , dan <i>kappa</i> ; tidak ada integrasi <i>pipeline</i> GEE-lokal yang <i>reproducible</i>
6	Forest Cover Change Detection Using Random Forest and Google Earth Engine on Sentinel-2 Multitemporal Imagery (<i>Ngo dkk., 2023</i>)	Random Forest, Google Earth Engine, Sentinel-2 multitemporal, NDVI, NDWI	Deteksi perubahan tutupan hutan di Vietnam Tengah dengan akurasi keseluruhan 89,4%; laju deforestasi 12.450 ha selama 2019–2022	Seluruh tahap pemrosesan dan pemodelan dilakukan di dalam GEE sehingga fleksibilitas evaluasi model terbatas; metrik evaluasi hanya melaporkan <i>overall accuracy</i> dan <i>kappa</i> tanpa <i>precision</i> , <i>recall</i> , dan <i>F1-score</i> per kelas; label <i>ground truth</i> bergantung pada digitasi manual yang sulit direproduksi

Meskipun penelitian sebelumnya telah mengeksplorasi penggunaan Random Forest untuk klasifikasi tutupan lahan [8] dan deteksi deforestasi [7, 27], sejumlah keterbatasan masih belum tertangani secara simultan. Pertama, mayoritas penelitian hanya melaporkan *overall accuracy* tanpa melengkapinya

dengan metrik seperti *precision*, *recall*, *F1-score*, dan koefisien kappa, sehingga gambaran performa model kurang komprehensif. Kedua, kajian multitemporal yang mendalam per kelas tutupan lahan masih jarang dilakukan, padahal informasi tersebut krusial untuk memahami pola deforestasi secara periodik. Ketiga, proses pengolahan citra dan pemodelan *machine learning* lazimnya dilakukan dalam satu platform atau terpisah tanpa alur kerja yang terstruktur, mengurangi transparansi dan kemampuan reproduksi hasil [12].

Penelitian ini hadir untuk menutup celah tersebut dengan mengintegrasikan Random Forest dalam satu *pipeline* terstruktur yang menghubungkan GEE untuk praproses citra dengan lingkungan komputasi lokal untuk pemodelan dan evaluasi. Pendekatan ini memungkinkan analisis perubahan tutupan lahan yang lebih akurat, terukur, dan dapat direproduksi, sekaligus menyajikan evaluasi model yang komprehensif melalui lima metrik statistik secara simultan [14, 12].

