

**PREDIKSI PERUBAHAN HUTAN *MANGROVE* BERDASARKAN  
DATA TUTUPAN LAHAN YANG TELAH TERSEGMENTASI  
MENGUNAKAN ALGORITMA *RANDOM FOREST*  
DI KABUPATEN MALANG**

**Oleh:**

**Danish Hasna Ayudya**

**1202213290**



**PROGRAM STUDI STRATA 1 SISTEM INFORMASI  
FAKULTAS REKAYASA INDUSTRI  
UNIVERSITAS TELKOM  
2025**

## ABSTRAK

Hutan *mangrove* memiliki peran yang sangat penting bagi ekosistem di sekitarnya. Namun, konversi lahan, aktivitas manusia, serta perubahan iklim telah menyebabkan deforestasi yang signifikan, termasuk di wilayah Pantai Selatan, Kabupaten Malang. Pemantauan perubahan tutupan lahan menjadi sangat penting untuk mengidentifikasi ancaman ini dan merancang strategi konservasi yang lebih efektif. Namun, keterbatasan akses terhadap data penginderaan jauh, seperti citra satelit, menjadi kendala dalam analisis yang lebih mendalam. Oleh karena itu, penelitian ini memanfaatkan algoritma *Random Forest* untuk memprediksi perubahan tutupan hutan *mangrove* berdasarkan peta tutupan lahan yang telah tersegmentasi dengan fitur yang terbatas melalui pendekatan rekayasa fitur. Proses rekayasa fitur dilakukan dengan mengekstraksi informasi spasial dan temporal dari fitur dasar untuk memperkaya representasi data. Dataset terdiri dari 16 titik pantai dengan data tahunan dari 2018 hingga 2023. Data diuji dalam tiga skenario pembagian data berdasarkan titik pantai yaitu 12:4, 13:3, dan 14:2, dan diklasifikasikan menjadi lima kelas, yaitu perubahan hutan, hutan *mangrove*, perairan, permukiman, dan tambak. Hasil penelitian menunjukkan bahwa skenario 12 area latihan dan 4 area uji memberikan performa terbaik, dengan *accuracy* 99,10%, serta *macro-average precision* 94%, *recall* 97%, dan *f1-score* 95%. Model berhasil mengidentifikasi hutan *mangrove* dengan sangat baik, tetapi menghadapi tantangan dalam mendeteksi penurunan hutan. Maka dari itu, penelitian ini merekomendasikan beberapa langkah untuk peningkatan akurasi prediksi di masa depan yaitu dengan penggabungan data spasial dari sumber lain seperti data satelit atau indeks vegetasi serta eksplorasi model *deep learning* untuk membandingkan pendekatan yang lebih efektif dalam mengklasifikasikan perubahan tutupan lahan.

Kata kunci—*hutan mangrove, data tutupan lahan, rekayasa fitur, Random Forest*

## ABSTRACT

*Mangrove forests play a crucial role in maintaining coastal ecosystem stability. However, land conversion, human activities, and climate change have led to significant deforestation, including in the southern coastal region of Malang Regency. Monitoring land cover changes is essential for identifying these threats and designing more effective conservation strategies. However, limited access to remote sensing data, such as satellite imagery, poses challenges for conducting in-depth analyses. To address this issue, this study utilizes the Random Forest algorithm to predict mangrove forest cover changes based on pre-segmented land cover maps with restricted features, employing a feature engineering approach to enhance data representation. Feature engineering is applied by extracting spatial and temporal information from base features to enrich the dataset. The dataset comprises 16 coastal areas with annual data from 2018 to 2023. Data is tested using three data-splitting scenarios based on coastal locations (12:4, 13:3, and 14:2) and classified into five class: forest degradation, mangrove forest, water bodies, settlements, and aquaculture ponds. The results indicate that the 12 training areas and 4 testing areas scenario achieves the best performance, with an accuracy of 99.10%, macro-average precision of 94%, recall of 97%, and f1-score of 95%. The model successfully identifies mangrove forests with high accuracy, yet it encounters challenges in detecting forest loss. To improve prediction accuracy in future studies, this research recommends integrating additional spatial data from sources such as satellite imagery or vegetation indices and exploring deep learning models to compare alternative approaches for more effective land cover change classification.*

*Keyword—mangrove forest, land cover data, feature engineering, Random Forest*

## HALAMAN PENGESAHAN

Tugas Akhir dengan judul:

**PREDIKSI PERUBAHAN HUTAN *MANGROVE* BERDASARKAN  
DATA TUTUPAN LAHAN YANG TELAH TERSEGMENTASI  
MENGUNAKAN ALGORITMA *RANDOM FOREST*  
DI KABUPATEN MALANG**

Telah disetujui dan disahkan pada Sidang Tugas Akhir

Program Studi Strata 1 Sistem Informasi

Fakultas Rekayasa Industri Universitas Telkom

**Oleh:**

**DANISH HASNA AYUDYA**

**1202213290**

Bandung, 29 Januari 2025

Disetujui oleh,

Pembimbing 1,



Dr. Sinung Suakanto, S.T., M.T.  
NIP: 22820012

Pembimbing 2,



Faqih Hamami, S.Kom., M.T.  
NIP: 23890014

## LEMBAR PERNYATAAN ORSINALITAS



Nama : Danish Hasna Ayudya  
NIM : 1202213290  
Alamat : Vila Pamulang, Jalan Amarta VI Blok  
DB VI No. 6  
No. Tlp : 087882379442

Menyatakan bahwa Tugas Akhir ini merupakan karya orisinal saya sendiri. Atas pernyataan ini, saya siap menanggung risiko atau sanksi yang dijatuhkan kepada saya apabila kemudian ditemukan adanya pelanggaran terhadap kejujuran akademik atau etika keilmuan dalam karya ini, atau ditemukan bukti yang menunjukkan ketidakaslian karya ini.

Bandung, 07 Januari 2025



Danish Hasna Ayudya

## **KATA PENGANTAR**

Puji dan syukur saya panjatkan ke hadirat Allah SWT atas limpahan rahmat, hidayah, dan kekuatan-Nya sehingga saya dapat menyelesaikan penulisan Tugas Akhir ini dengan baik. Tugas Akhir ini merupakan salah satu syarat kelulusan program studi dan menjadi bukti perjalanan akademik selama menempuh pendidikan di perguruan tinggi. Proses penyusunan Tugas Akhir ini tidak terlepas dari bantuan dan dukungan dari berbagai pihak yang telah memberikan dorongan, arahan, dan doa selama penyusunan berlangsung.

Dalam kesempatan ini, saya ingin mengucapkan rasa terima kasih yang sebesar-besarnya kepada:

1. Ayah dan Bunda, yang selalu memberikan dukungan, cinta, dan doa yang tidak pernah henti dalam setiap langkah perjalanan hidup Saya.
2. Bapak Dr. Sinung Suakanto, S.T., M.T. dan Bapak Faqih Hamami, S. Kom., M.T, selaku dosen pembimbing Tugas Akhir Saya yang telah memberikan bimbingan, arahan, serta saran berharga selama proses penyusunan Tugas Akhir ini.
3. Aisyah Rahma Rabbania dan Farisa Fikri Mumtaza, teman-teman yang menemani dalam masa perkuliahan dan banyak membantu dalam penyusunan Tugas Akhir ini.
4. Teman-teman lain yang senantiasa mendoakan, meskipun tidak selalu berada di dekat saya, doa dan dukungan dari kejauhan yang mereka berikan sangat berarti dalam proses ini.

Semoga penulisan ini dapat memberikan manfaat bagi pembaca dan berkontribusi dalam pengembangan ilmu pengetahuan.

## DAFTAR ISI

ABSTRAK .....	ii
ABSTRACT .....	iii
HALAMAN PENGESAHAN.....	iv
KATA PENGANTAR.....	vi
DAFTAR ISI .....	vii
DAFTAR GAMBAR .....	1
DAFTAR TABEL.....	x
BAB I PENDAHULUAN .....	2
I.1 Latar Belakang .....	2
I.2 Rumusan Masalah .....	4
I.3 Tujuan Tugas Akhir.....	5
I.4 Manfaat Tugas Akhir.....	5
I.5 Batasan dan Asumsi Tugas Akhir.....	6
I.6 Sistematika Laporan.....	6
BAB II LANDASAN TEORI .....	8
II.1 Literatur.....	8
II.1.1 <i>Mangrove</i> .....	8
II.1.2 <i>Land Cover</i> .....	9
II.1.3 <i>Feature Engineering</i> .....	9
II.1.4 <i>Resampling Data</i> .....	10
II.1.5 <i>Random Forest</i> .....	12
II.1.6 Evaluasi Model.....	13
II.1.7 Penelitian Terdahulu.....	15

II.2	Pemilihan Metode atau Kerangka Kerja .....	20
BAB III	METODE PENYELESAIAN MASALAH .....	21
III.1	Sistematika Penyelesaian Masalah.....	21
III.1.1	Pengembangan Model Konseptual.....	21
III.1.2	Metode Penyelesaian Masalah .....	22
III.1.3	Metode Pengumpulan Data .....	24
III.1.4	Metode Pengolahan Data .....	25
III.1.5	Metode Evaluasi.....	25
BAB IV	PENYELESAIAN MASALAH .....	27
IV.1	<i>Data Selection</i> .....	27
IV.1.1	Eksplorasi Data .....	27
IV.1.2	Pemilihan <i>Dataset</i> .....	29
IV.2	<i>Data Pre-processing</i> .....	30
IV.2.1	Pemotongan Data .....	30
IV.2.2	Reduksi <i>Noise</i> .....	31
IV.2.3	Standarisasi Warna .....	33
IV.2.4	<i>Downscaling Data</i> .....	34
IV.3	<i>Data Transformation</i> .....	34
IV.3.1	Kategorisasi Piksel.....	35
IV.3.2	<i>Feature Engineering</i> .....	36
IV.3.3	Penggabungan Data.....	37
IV.4	<i>Data Mining</i> .....	40
IV.4.1	<i>Data Splitting</i> .....	41
IV.4.2	Penanganan Data <i>Imbalance</i> .....	44



IV.4.3	<i>Modeling</i> .....	46
IV.5	<i>Evaluation</i> .....	53
BAB V	VALIDASI, ANALISIS, HASIL, DAN IMPLIKASI.....	55
V.1	Pengujian Model .....	55
V.2	Evaluasi Hasil Pengujian.....	56
V.2.1	<i>Confusion matrix</i> .....	56
V.2.2	Perhitungan Metriks Evaluasi .....	59
V.2.3	Analisis Hasil Pengujian Model Terbaik.....	62
BAB VI	KESIMPULAN DAN SARAN.....	72
VI.1	Kesimpulan .....	72
VI.2	Saran.....	72
DAFTAR PUSTAKA	.....	74

## DAFTAR TABEL

Tabel II-1 <i>State of Art</i> .....	15
Tabel II-2 Perbandingan Kerangka Kerja .....	20
Tabel IV-1 Data Peta Persebaran <i>Mangrove</i> dan Garis Pantai .....	27
Tabel IV-2 Data 18 Titik Pantai.....	28
Tabel IV-3 Label Piksel Berdasarkan Nilai RGB.....	35
Tabel IV-4 Seluruh Fitur pada <i>Dataframe</i> .....	37
Tabel IV-5 <i>Sample Dataset</i> untuk Model Prediksi.....	40
Tabel IV-6 Skenario Pembagian Data .....	41
Tabel IV-7 Pembagian Data .....	43
Tabel IV-8 Pembagian Data setelah Teknik SMOTE.....	43
Tabel IV-9 <i>Detail</i> Jumlah Distribusi Data pada Setiap Skenario Pembagian Data.....	45
Tabel IV-10 Parameter yang Diuji .....	47
Tabel IV-11 Hasil Kombinasi Parameter dari <i>Framework Optuna</i> .....	47
Tabel V-1 Hasil Pengujian Model .....	55
Tabel V-2 Hasil Perhitungan Metriks Evaluasi .....	60
Tabel V-3 Hasil Perhitungan Metriks Evaluasi setiap Kelas.....	61
Tabel V-4 Metriks Evaluasi pada Data Uji Pantai Bajul Mati .....	63
Tabel V-5 Metriks Evaluasi pada Data Uji Pantai Sempu.....	66
Tabel V-6 Metriks Evaluasi pada Data Uji Pantai Sipelot .....	68
Tabel V-7 Metriks Evaluasi pada Data Uji Pantai Tamban.....	70

## DAFTAR GAMBAR

Gambar II.1 Ilustrasi Teknik SMOTE.....	11
Gambar II.2 Ilustrasi Algoritma <i>Random Forest</i> .....	12
Gambar II.3 <i>Confusion matrix</i> .....	14
Gambar III.1 Model Konseptual .....	21
Gambar III.2 Ilustrasi Kerangka Kerja KDD.....	22
Gambar III.3 Sistematisa Penelitian menggunakan KDD .....	23
Gambar IV.1 Peta Persebaran <i>Mangrove</i> dan Garis Pantai Sendang Biru Tahun 2018 .....	29
Gambar IV.2 Perbandingan Sebelum dan Sesudah Pemotongan Data .....	31
Gambar IV.3 <i>Noise</i> pada Gambar.....	32
Gambar IV.4 Perbandingan Sebelum dan Sesudah Reduksi <i>Noise</i> .....	32
Gambar IV.5 Perbandingan Sebelum dan Sesudah Standarisasi Warna .....	33
Gambar IV.6 Perbandingan Sebelum dan Sesudah <i>Downscaling</i> .....	34
Gambar IV.7 Ukuran Jendela 3x3 dengan Delapan Piksel Tetangga.....	36
Gambar IV.8 Distribusi Jumlah Data pada Setiap Skenario Pembagian Data .....	44
Gambar IV.9 Hasil <i>Resampling</i> dengan Teknik SMOTE.....	46
Gambar IV.10 <i>Confusion matrix</i> 6x6.....	53
Gambar V.1 <i>Confusion matrix</i> untuk 12 Area Latih dan 4 Area Uji .....	57
Gambar V.2 <i>Confusion matrix</i> untuk 13 Area Latih dan 3 Area Uji .....	58
Gambar V.3 <i>Confusion matrix</i> untuk 14 Area Latih dan 2 Area Uji .....	59
Gambar V.4 <i>Confusion Matrix</i> pada Data Uji Pantai Bajul Mati.....	63
Gambar V.5 Visualisasi Hasil Prediksi Data Uji Pantai Bajul Mati.....	64
Gambar V.6 <i>Confusion Matrix</i> pada Data Uji Pantai Sempu.....	65
Gambar V.7 Visualisasi Hasil Prediksi Data Uji Pantai Sempu .....	66
Gambar V.8 <i>Confusion Matrix</i> pada Data Uji Pantai Sipelot.....	67
Gambar V.9 Visualisasi Hasil Prediksi Data Uji Pantai Sipelot.....	68
Gambar V.10 <i>Confusion Matrix</i> pada Data Uji Pantai Tamban .....	69
Gambar V.11 Visualisasi Hasil Prediksi Data Uji Pantai Tamban .....	70

# PENDAHULUAN

## I.1 Latar Belakang

Hutan *mangrove* merupakan salah satu ekosistem wilayah pesisir yang memiliki peran penting dalam kelangsungan lingkungan sekitarnya. Namun, kawasan hutan *mangrove* di Indonesia mengalami penurunan luasan seiring berjalannya waktu. Melansir dari BBC News Indonesia, hampir 13.000 hektar *mangrove* hilang setiap tahunnya dalam 20 tahun terakhir (Lubis, 2023). Selain itu, data Kementerian Lingkungan Hidup dan Kehutanan (KLHK) periode 2010 sampai 2020 menyatakan bahwa Indonesia telah kehilangan area *mangrove* hampir sekitar 200.000 hektar dengan kehilangan area *mangrove* tertinggi di tahun 2016 sebesar 60.000 hektar. Berdasarkan wawancara REPUBLIKA bersama dengan Direktur Jenderal Pengendalian Daerah Aliran Sungai dan Hutan Lindung KLHK di tahun 2018, Terjadi kerusakan hutan *mangrove* seluas 1,81 juta hektar dan hanya 1,67 juta hektar hutan *mangrove* di Indonesia yang berada pada kondisi baik. Penyebab utama dari kerusakan tersebut adalah faktor manusia yang melakukan alih fungsi lahan menjadi kawasan non-hutan, seperti Pembangunan pemukiman, infrastruktur, perkebunan, dan area tambak.

Daerah Pantai Selatan, Kabupaten Malang menjadi salah satu wilayah yang mengalami kerusakan hutan *mangrove* di Indonesia. Kerusakan ini sudah terjadi sejak tahun 1998. Krisis moneter yang melanda Indonesia dan mempengaruhi perekonomian masyarakat mengakibatkan deforestasi hutan *mangrove* di Pantai Dusun Sendangbiru, Kecamatan Sumberwijing Wetan akibat dijadikan lahan pertanian dan tambak liar dalam skala besar. Hal ini mempengaruhi stabilitas ekosistem wilayah pesisir yang dampaknya masih berlanjut hingga di tahun-tahun berikutnya. Dampak paling parah terjadi di rentang tahun 2000 sampai 2004. Warga Sendangbiru mengalami paceklik ikan hingga Pemerintah Kabupaten Malang harus mengirim beras sebagai bantuan pangan (Imaduddin & Krisnadi, 2020). Melansir dari media Tempo (2013), 57% dari hutan *mangrove* di pesisir selatan mengalami kerusakan, yaitu seluas 195 hektar dari 344 hektar. Di tahun yang sama, profil Desa Tambakrejo Kecamatan Sumbermanjing Wetan Kabupaten Malang mengungkapkan bahwa daerah tersebut mengalami kerusakan

hutan *mangrove* seluas 190 hektar dari total 200 hektar. Hal ini menunjukkan bahwa kerusakan hutan *mangrove* di Pantai Selatan Kabupaten Malang sangat signifikan dan memerlukan perhatian serius dari berbagai pihak. Kerusakan yang terjadi tidak hanya berdampak pada ekosistem pesisir, tetapi juga pada kesejahteraan masyarakat yang bergantung pada sumber daya alam tersebut.

Dalam mengatasi kerusakan ini, sudah terdapat upaya pemeliharaan dan rehabilitasi hutan *mangrove* melalui inisiatif masyarakat membentuk Lembaga Yayasan Bhakti Alam pada tahun 2014 sebagai upaya konservasi *mangrove* di Desa Tambakrejo. Lembaga tersebut dibentuk tidak hanya untuk konservasi hutan *mangrove* saja, tetapi juga untuk mengelola kawasan tersebut sebagai kawasan ekowisata, sehingga masyarakat dapat mandiri secara ekonomi melalui upaya konservasi ini (Imaduddien & Krisnadi, 2020). Meski sudah ada upaya pemeliharaan yang dilakukan, konservasi hutan *mangrove* masih diadakan di sebagian daerah dari Kabupaten Malang saja. Masih terdapat area hutan *mangrove* yang kondisinya belum diketahui di area Pantai Selatan.

Pemantauan yang efektif menjadi sangat penting untuk mendukung konservasi dan pemanfaatan berkelanjutan hutan *mangrove* (Win & Sasaki, 2024). Metode tradisional berbasis lapangan memerlukan biaya dan waktu yang besar, sehingga pendekatan melalui penginderaan jauh muncul sebagai solusi pemantauan skala besar yang lebih efisien (Maurya dkk., 2021). Dengan memanfaatkan citra satelit dan teknologi *machine learning*, pemetaan distribusi *mangrove*, penilaian kerapatan vegetasi, serta deteksi perubahan kondisi fisiknya dapat dilakukan dengan lebih akurat (Dzulfigar dkk., 2024). Citra satelit mampu menyediakan data spektral yang kaya di berbagai *bands*, yang memungkinkan pengembangan dan penerapan berbagai indeks spektral (Tran dkk., 2022). Indeks-indeks tersebut digunakan untuk analisis dengan menyediakan informasi terperinci tentang kondisi dan dinamika ekosistem. Namun, tidak semua wilayah memiliki akses terhadap citra satelit yang lengkap akibat kurangnya data historis, cakupan yang terbatas, atau kualitas data yang kurang optimal. Keterbatasan ini dapat menghambat analisis yang lebih mendetail, sehingga diperlukan pendekatan alternatif untuk tetap melakukan pemantauan tutupan lahan secara efektif.

Dalam penelitian ini, analisis prediktif terhadap perubahan ekosistem hutan *mangrove* dilakukan menggunakan peta tutupan lahan yang telah tersegmentasi, mengingat keterbatasan akses terhadap citra satelit yang lebih rinci. Tidak seperti metode berbasis penginderaan jauh yang menggunakan spektrum elektromagnetik luas, data yang digunakan dalam penelitian ini hanya mencakup nilai RGB, sehingga jumlah informasi yang dapat diekstraksi secara langsung menjadi lebih terbatas. Hal ini menjadi tantangan utama dalam pemodelan, karena informasi spektral tambahan seperti indeks vegetasi atau kedalaman spektral lainnya tidak tersedia dalam dataset yang digunakan.

Untuk mengatasi keterbatasan ini, penelitian ini menerapkan teknik *feature engineering* sebagai solusi utama dalam meningkatkan kualitas representasi data. *Feature engineering* terbukti mampu memperkaya representasi data dan meningkatkan akurasi klasifikasi dengan melakukan ekstraksi terhadap fitur spasial maupun temporal (Zhai dkk., 2018). Fitur yang dilakukan rekayasa pada penelitian ini mencakup fitur piksel tetangga yang merepresentasikan fitur spasial dan fitur perubahan piksel yang merepresentasikan fitur temporal.

Untuk membangun model prediksi perubahan hutan *mangrove*, digunakan *Random Forest* sebagai algoritma klasifikasi utama. Beberapa penelitian telah melakukan klasifikasi perubahan tutupan lahan menggunakan metode *machine learning*, termasuk *Random Forest* yang telah terbukti lebih unggul dalam tugas klasifikasi tutupan lahan dibandingkan metode lain seperti *Support Vector Machine* (SVM), *Artificial Neural Network* (ANN), dan *Maximum Likelihood* (ML) (Chowdhury, 2024). *Random Forest* memiliki keunggulan dalam klasifikasi berbasis spektral yang mampu menangani data dengan fitur yang bervariasi serta menghasilkan prediksi yang lebih stabil (Ramírez dkk., 2023).

## **I.2 Rumusan Masalah**

Adapun rumusan masalah yang mendasari penelitian ini antara lain:

1. Bagaimana menentukan konfigurasi terbaik pada model dalam melakukan prediksi perubahan hutan *mangrove* berdasarkan data tutupan lahan yang telah tersegmentasi di Kabupaten Malang?
2. Bagaimana performa model secara keseluruhan dalam memprediksi perubahan hutan *mangrove*?

### **I.3 Tujuan Tugas Akhir**

Berdasarkan uraian pada rumusan masalah, tujuan dilakukannya penelitian ini adalah:

1. Menentukan konfigurasi terbaik dari model yang digunakan dalam prediksi perubahan hutan *mangrove*, dengan mengeksplorasi berbagai parameter dan teknik optimasi untuk meningkatkan akurasi model.
2. Mengevaluasi performa model secara keseluruhan dalam memprediksi perubahan hutan *mangrove* serta menganalisis hasil model dengan performa terbaik.

### **I.4 Manfaat Tugas Akhir**

Adapun manfaat dari penelitian ini berdasarkan empat sudut pandang yaitu bagi mahasiswa, bagi Telkom University, bagi organisasi terkait, dan bagi keilmuan Sistem Informasi.

1. Bagi mahasiswa, penelitian ini bermanfaat untuk meningkatkan keterampilan dalam analisis data *time-series* dan pemodelan menggunakan algoritma *Random Forest* , serta memberikan pengalaman praktis dalam penerapan *machine learning* pada studi kasus nyata terkait analisis perubahan hutan *mangrove* di Pantai Selatan, Kabupaten Malang.
2. Bagi Telkom University, penelitian ini bermanfaat sebagai bahan ajar dan referensi bagi mahasiswa yang tertarik pada analisis data lingkungan serta berkontribusi dalam meningkatkan reputasi akademik melalui publikasi di tingkat nasional maupun internasional.
3. Bagi organisasi terkait, penelitian ini bermanfaat dalam mendukung perencanaan konservasi dan restorasi hutan *mangrove* yang lebih efektif serta sebagai dasar dalam perumusan kebijakan perlindungan dan pelestarian lingkungan berbasis data.

4. Bagi keilmuan Sistem Informasi, penelitian ini memberikan referensi tentang penerapan algoritma *Random Forest* dalam pengelolaan data lingkungan serta memperkaya studi kasus terkait pemanfaatan teknologi informasi dalam pemantauan ekosistem hutan *mangrove*.

### **I.5 Batasan dan Asumsi Tugas Akhir**

Penelitian ini berfokus pada pengolahan dan analisis data tutupan lahan untuk memprediksi perubahan hutan *mangrove* di area pesisir Kabupaten Malang, dengan menggunakan data yang telah tersegmentasi. Data yang digunakan terbatas pada 16 titik pantai di wilayah tersebut dan hanya mencakup periode 2018 hingga 2023. Penelitian ini mengasumsikan bahwa data yang digunakan telah melalui proses segmentasi yang valid serta memiliki tingkat akurasi yang memadai untuk mendukung pemodelan prediksi perubahan tutupan lahan.

### **I.6 Sistematika Laporan**

1. Bab I Pendahuluan, menjelaskan latar belakang, rumusan masalah, tujuan tugas akhir, manfaat tugas akhir, serta batasan dan asumsi tugas akhir. Selain itu, sistematika laporan disusun untuk memberikan gambaran mengenai alur penelitian
2. Bab II Landasan Teori, membahas berbagai literatur yang relevan untuk mendukung penelitian, penelitian terdahulu yang terkait dengan prediksi perubahan hutan *mangrove*, serta penjelasan metode atau kerangka kerja yang digunakan berdasarkan relevansinya dengan tujuan penelitian.
3. Bab III Metode Penyelesaian Masalah, menguraikan sistematika dalam menyelesaikan masalah yang mencakup pengembangan model konseptual, metode penyelesaian masalah, metode pengumpulan data, metode pengolahan data, dan metode evaluasi.
4. Bab IV Penyelesaian Masalah, menjelaskan penerapan langkah-langkah penyelesaian masalah sesuai dengan metode yang telah dirancang pada Bab III Metode Penyelesaian Masalah.



5. Bab V Validasi, Analisis, Hasil, dan Implikasi, membahas hasil pengujian model melalui beberapa eksperimen untuk menemukan kombinasi konfigurasi terbaik, evaluasi dan analisis hasil, serta dampak hasil tugas akhir.
6. Bab VI Kesimpulan dan Saran, menjelaskan hasil temuan dari penelitian dan menjawab tujuan penelitian yang telah didefinisikan, serta saran untuk pengembangan penelitian di masa mendatang.

## BAB I LANDASAN TEORI

### II.1 Literatur

#### II.1.1 *Mangrove*

*Mangrove* merupakan istilah untuk ekosistem hutan yang toleran terhadap garam, umumnya ditemukan di wilayah pesisir tropis dan subtropis. *Mangrove* memiliki adaptasi unik, seperti akar di atas permukaan tanah, daun yang mampu mengeluarkan garam, dan biji yang dapat mengapung. Adaptasi ini memungkinkan mereka untuk bertahan di lingkungan pasang surut dengan salinitas yang bervariasi, tanah anaerobik, dan kondisi angin yang kencang (Vilma dkk., 2022). Selain itu, *mangrove* umumnya hanya ditemukan di daerah dengan rentang suhu rata-rata 20 hingga 35 derajat celsius, curah hujan tahunan antara 1500 hingga 2500 mm, dan pasokan air tawar yang cukup dari sungai (Tran dkk., 2022).

Sebagai ekosistem pesisir yang penting, *mangrove* berfungsi untuk mengurangi erosi, melindungi dari cuaca ekstrem dan gelombang tinggi, serta menyediakan habitat bagi berbagai spesies. Selain itu, *mangrove* meningkatkan keanekaragaman hayati, memperbaiki kualitas air, menyimpan karbon dalam jumlah besar, dan memainkan peran penting dalam mendukung komunitas lokal melalui penyediaan sumber daya dan penghidupan (Farzanmanesh dkk., 2024). Meskipun memiliki manfaat yang sangat beragam, *mangrove* merupakan vegetasi yang rentan terhadap ancaman baik secara alami maupun aktivitas manusia (Ginancar dkk., 2023).

*Global Mangrove Alliance* melaporkan bahwa pada tahun 2020 luas hutan *mangrove* di dunia mencapai 14,7 juta hektar dengan Asia Tenggara sebagai area hutan *mangrove* paling luas. Indonesia sebagai negara maritim kepulauan memiliki garis pantai yang sangat panjang yang sebagiannya ditumbuhi hutan *mangrove* dengan luas yang beragam (Leal dkk., 2022). Berdasarkan data Kementerian Lingkungan Hidup dan Kehutanan tahun 2023, hutan *mangrove* Indonesia merupakan kawasan hutan *mangrove* terluas di dunia yang mencakup lebih dari 24 persen dari total luas *mangrove* dunia, yaitu 3,36 juta hektar.

Berdasarkan Standar Nasional Indonesia (SNI) 7717:2020 dengan judul Spesifikasi informasi geospasial – *Mangrove* skala 1:25.000 dan 1:50.000, kondisi *mangrove* di Indonesia diklasifikasikan menjadi tiga jenis berdasarkan persentase tutupan tajuknya, yaitu *mangrove* lebat, *mangrove* sedang, dan *mangrove* jarang. Kondisi *mangrove* lebat merupakan *mangrove* dengan tutupan tajuk >70%, *mangrove* sedang dengan tutupan tajuk dalam rentang 30% sampai dengan 70%, dan *mangrove* jarang dengan tutupan tajuk < 30%.

### **II.1.2 *Land Cover***

*Land cover* atau penutup lahan mengacu pada karakteristik fisik permukaan lahan, seperti hutan, perairan, area pertambangan, dan kategori penutup lainnya (Letsoin dkk., 2020). Pemahaman terhadap perubahan penutup lahan sangat penting untuk memberikan wawasan mendalam mengenai perubahan lingkungan. Dengan mengetahui bagaimana penutup lahan berubah, kita dapat mengidentifikasi faktor-faktor yang memengaruhi ekosistem dan membuat kebijakan yang mendukung pelestariannya (Cheng & Wang, 2019).

Pemantauan dan pemetaan penutup lahan memiliki peran penting dalam perencanaan tata ruang dan pengelolaan lingkungan (Aliani dkk., 2019). Analisis terhadap penutup juga memungkinkan kita untuk memprediksi kondisi lingkungan di masa depan secara lebih akurat. Sebagai contoh, perubahan penutup hutan di masa depan dapat diperkirakan dengan memanfaatkan data historis dan pengamatan dari teknologi penginderaan jauh, sehingga memberikan gambaran yang lebih andal untuk mendukung pengelolaan sumber daya alam yang berkelanjutan.

### **II.1.3 *Feature Engineering***

*Feature Engineering* atau rekayasa fitur, merupakan salah satu langkah krusial dalam proses *machine learning* yang berperan dalam mengubah data mentah menjadi format yang lebih terstruktur dan informatif. Langkah ini memiliki dampak signifikan terhadap performa prediksi model, karena fitur yang dihasilkan secara langsung memengaruhi kemampuan model dalam memahami pola dan hubungan yang ada dalam data (Katya, 2023). Dengan melakukan rekayasa fitur, data dapat diproses

sedemikian rupa sehingga lebih sesuai dengan kebutuhan analisis, memungkinkan model untuk mengolah informasi yang lebih relevan dengan permasalahan yang sedang ditangani (Derhab dkk., 2020). Proses ini tidak hanya meningkatkan kualitas data, tetapi juga dapat membantu mengurangi ketergantungan pada algoritma yang kompleks dengan menyediakan representasi data yang lebih optimal.

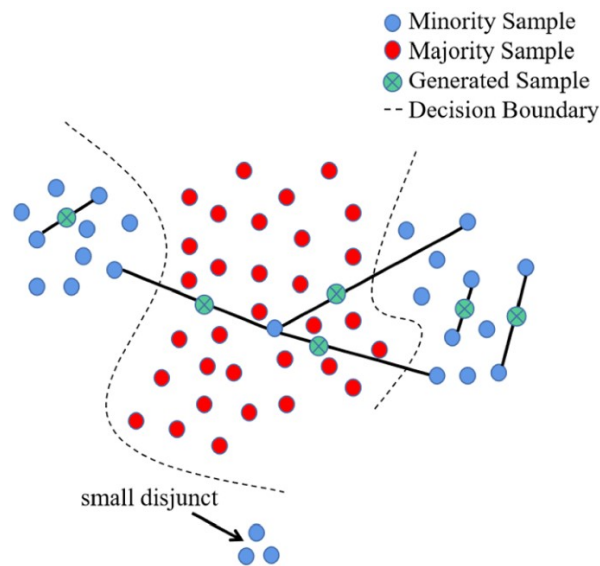
Selain berperan dalam mempersiapkan data, *feature engineering* juga bertujuan untuk meningkatkan kekuatan prediktif dan akurasi model dengan menambahkan fitur yang lebih informatif (F. Khan dkk., 2024). Fitur yang dirancang secara efektif dapat memperjelas pola yang tersembunyi dalam data, mempermudah model dalam membedakan kategori yang serupa, serta mengurangi risiko kesalahan prediksi. Dengan mengidentifikasi dan mengekstraksi elemen data yang paling relevan, proses ini memungkinkan model untuk belajar dengan lebih baik, bahkan ketika bekerja dengan data yang kompleks atau jumlah data yang terbatas (Alamri & Ykhlef, 2024).

#### **II.1.4 Resampling Data**

*Resampling* merupakan teknik yang digunakan untuk menangani ketidakseimbangan data dengan menciptakan distribusi yang lebih merata di antara kelas-kelas dalam dataset (Indrawati dkk., 2020). Teknik ini bekerja dengan membuat sampel baru yang diambil dari data yang sudah ada (Nurhopipah & Magnolia, 2023), baik melalui *oversampling* (menambahkan sampel pada kelas minoritas) maupun *undersampling* (mengurangi jumlah sampel pada kelas mayoritas). Dengan melakukan *resampling*, dataset yang dihasilkan menjadi lebih seimbang, yang dapat membantu proses klasifikasi menjadi lebih sederhana dan mencegah model menjadi bias terhadap kelas mayoritas.

Dalam penelitian ini, teknik *resampling* digunakan dengan metode *Synthetic Minority Over-sampling Technique* (SMOTE) untuk menangani ketidakseimbangan data pada kelas penutup lahan. SMOTE merupakan metode *resampling* berbasis interpolasi yang bertujuan menambah sampel pada kelas dengan jumlah data yang lebih rendah secara sintesis. Teknik ini bekerja dengan memilih sampel secara acak dari kelas minoritas dan menghasilkan data baru dengan melakukan interpolasi linier di antara dua titik data

yang berdekatan (Nurhopipah & Magnolia, 2023). Proses ini melibatkan tiga tahap utama, yaitu memilih sampel dari kelas minoritas secara acak, memilih tetangga terdekat menggunakan metode *K-Nearest Neighbors*, dan menghasilkan sampel sintetis dengan menghitung rata-rata berbobot antara dua titik data yang dipilih (Nurhopipah dkk., 2021). Adapun ilustrasi dari implementasi teknik SMOTE ditunjukkan pada Gambar I.1.



Gambar I.1 Ilustrasi Teknik SMOTE

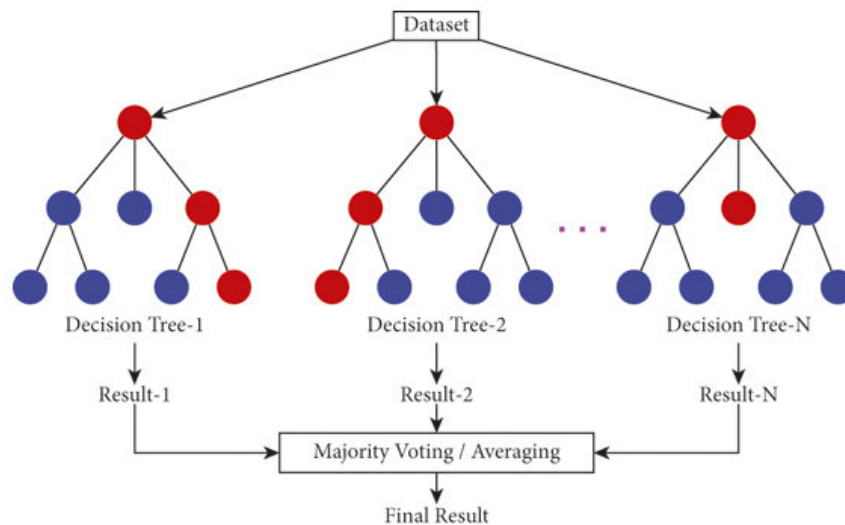
Sumber: Yi dkk., 2022

Dengan menerapkan metode SMOTE, distribusi data pada kelas minoritas menjadi lebih seimbang dengan kelas mayoritas, sehingga model dapat mempelajari pola yang lebih representatif dari setiap kelas penutup lahan. Penelitian sebelumnya menunjukkan bahwa penggunaan SMOTE secara signifikan mampu meningkatkan kinerja model klasifikasi, terutama ketika dikombinasikan dengan algoritma *Random Forest*, menghasilkan peningkatan akurasi, sensitivitas, dan spesifisitas (Hasanah dkk., 2024). Selain itu, SMOTE digunakan untuk mengatasi tantangan *overfitting* yang kerap muncul dalam teknik *oversampling* pada dataset yang tidak seimbang. Metode ini menciptakan data yang lebih representatif dibandingkan hanya menduplikasi data yang ada (Hartono & Ongko, 2022). Hal ini memungkinkan model *Random Forest* untuk

mempelajari pola yang lebih bervariasi dari kelas yang kurang terwakili, sehingga meningkatkan performa secara keseluruhan dalam mengklasifikasikan perubahan penutup lahan *mangrove* dengan lebih akurat.

### II.1.5 *Random Forest*

*Random Forest* merupakan sebuah algoritma yang menggunakan sekumpulan pohon keputusan (*decision tree*) untuk meningkatkan akurasi dalam pengelompokan data ke dalam kelas tertentu. Setiap pohon dibangun secara acak dengan atribut yang dihasilkan pada setiap *node*, sehingga memastikan keragaman dan keandalan model dalam proses klasifikasi (Amaliah dkk., 2022). Algoritma ini merupakan sebuah metode pembelajaran *ensemble* yang kuat yang mampu meningkatkan kinerja klasifikasi serta mengurangi korelasi data (Salman dkk., 2024). Proses ini dicapai melalui dua tingkat randomisasi selama konstruksi model, yaitu pemilihan acak subset data menggunakan *bootstrap* sampling dan pemilihan acak subset fitur pada setiap percabangan pohon keputusan. Dual randomisasi ini tidak hanya mengurangi korelasi antar pohon, tetapi juga meminimalkan risiko *overfitting*, sehingga menghasilkan model yang lebih akurat dan andal. Gambar I.2 berikut merupakan ilustrasi dari algoritma *Random Forest*.



Gambar I.2 Ilustrasi Algoritma *Random Forest*

Sumber: Khan dkk., 2021

Proses umum dalam membangun *Random Forest* melibatkan penyusunan sejumlah pohon keputusan, di mana setiap pohon dilatih menggunakan *bootstrap sample* dari data yang tersedia. Prediksi dari setiap pohon kemudian digabungkan untuk menghasilkan output akhir (Alita & Isnain, 2020). Adapun rumus dari *Random Forest* ditunjukkan pada (II-1).

$$y = \frac{1}{B} \sum_{b=1}^B t_b(x) \quad (\text{II-1})$$

Di mana  $B$  adalah jumlah pohon keputusan, dan  $t_b(x)$  merupakan prediksi dari pohon ke- $b$  untuk input  $x$ .

Salah satu keunggulan utama *Random Forest* adalah kemampuannya menangani dataset dengan jumlah observasi dan fitur yang terbatas, yang menjadi aspek penting dalam analisis tutupan lahan *mangrove* pada penelitian ini. Berbeda dengan banyak model kompleks yang membutuhkan dataset besar dan beragam untuk memberikan kinerja optimal, *Random Forest* sangat adaptif terhadap dataset kecil. Pendekatan *ensemble* dan mekanisme agregasinya memastikan prediksi yang andal, bahkan dalam kondisi keterbatasan sumber daya. Penelitian sebelumnya menunjukkan bahwa *Random Forest* mampu mencapai akurasi klasifikasi yang tinggi dan mempertahankan kinerja prediktif yang baik meskipun data yang tersedia terbatas (Han dkk., 2021).

### **II.1.6 Evaluasi Model**

Dalam penelitian ini, kinerja model *Random Forest* yang diusulkan dievaluasi menggunakan berbagai metrik evaluasi klasifikasi standar untuk menilai efektivitasnya dalam mengklasifikasikan data penutup lahan *mangrove*. Evaluasi ini bertujuan memberikan pemahaman komprehensif mengenai performa model dari berbagai aspek, termasuk kemampuannya dalam mengklasifikasikan piksel dengan benar, menangani data yang tidak seimbang, serta menjaga keseimbangan antara presisi dan sensitivitas. Untuk menganalisis kinerja model secara mendalam, digunakan

*confusion matrix* sebagai alat utama dalam menilai akurasi prediksi model (Riyanto, 2019).

		Predicted	
		Positive	Negative
Actual	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

Gambar I.3 *Confusion matrix*

*Confusion matrix* menyajikan hasil prediksi dalam bentuk tabel yang membandingkan label sebenarnya dengan hasil prediksi model, memungkinkan identifikasi kesalahan klasifikasi secara terperinci. Matriks ini terdiri dari empat elemen utama, seperti yang ditampilkan pada Gambar I.3, yaitu:

1. *True Positive* (TP): Jumlah data yang diklasifikasikan dengan benar sebagai positif.
2. *True Negative* (TN): Jumlah data yang diklasifikasikan dengan benar sebagai negatif.
3. *False Positive* (FP): Jumlah data yang salah diklasifikasikan sebagai positif.
4. *False Negative* (FN): Jumlah data yang salah diklasifikasikan sebagai negatif.

Berdasarkan nilai-nilai dalam *confusion matrix*, beberapa evaluasi dapat dihitung untuk mengukur performa model secara lebih mendalam (Ainurrohma, 2021), yang meliputi:



1. *Accuracy*

Mengukur proporsi piksel yang diklasifikasikan dengan benar dari total sampel.

Rumus untuk menghitung ditunjukkan pada Persamaan (II-2)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (II-2)$$

2. *Precision*

Mengevaluasi kemampuan model dalam meminimalkan *false positives*, yang mencerminkan akurasi prediksi positif, seperti ditunjukkan pada persamaan (II-3):

$$Precision = \frac{TP}{TP + FP} \quad (II-3)$$

3. *Recall*

Mengukur kemampuan model dalam menangkap *true positives*, yang mencerminkan sensitivitas model. Rumus untuk menghitung *Recall* ditunjukkan pada persamaan (II-4):

$$Recall = \frac{TP}{TP + FN} \quad (II-4)$$

4. *F1-score*

Menggabungkan *Precision* dan *Recall* ke dalam satu metrik, memberikan evaluasi yang seimbang, terutama untuk dataset yang tidak seimbang, seperti dinyatakan pada persamaan (II-5):

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (II-5)$$

Metriks-metriks tersebut digunakan dalam penelitian ini untuk mengevaluasi kinerja model *Random Forest* secara menyeluruh. Evaluasi ini diharapkan dapat memberikan gambaran akurat mengenai keandalan model dalam memprediksi berbagai kategori penutup lahan, baik yang memiliki frekuensi tinggi maupun rendah.

### II.1.7 Penelitian Terdahulu

Dalam penelitian ini, beberapa referensi terdahulu dikaji dan dipelajari sebagai dasar pemilihan algoritma dalam analisis prediktif perubahan hutan *mangrove*. Referensi yang mendasari penelitian ini ditampilkan pada Tabel I-1.

Tabel I-1 *State of Art*

No	Judul	Data Input	Output	Metode	Hasil	Relasi dengan Penelitian Ini
1	<i>Urban Land Use and Land Cover Change Analysis Using Random Forest Classification of Landsat Time Series</i> (Amini dkk., 2022).	Landsat 5, 7, dan 8 dengan indeks NDVI, EVI, MNDWI, serta data tambahan <i>Land Surface Temperature</i> (LST), <i>Digital Elevation Model</i> (DEM).	Peta <i>Land Use</i> dan <i>Land Cover</i> (LULC) untuk wilayah Isfahan, Iran, dari 1985 hingga 2019.	<i>Random Forest Classifier</i> .	Akurasi model keseluruhan mencapai 94,438% dengan koefisien <i>kappa</i> 0,93 serta data tambahan LST dan DEM terbukti meningkatkan akurasi.	Menjadi dasar penggunaan <i>Random Forest</i> untuk klasifikasi tutupan lahan dan integrasi data tambahan dapat meningkatkan akurasi.
2	<i>Spatiotemporal Hybrid Random Forest Model for Tea Yield Prediction Using Satellite-Derived Variables</i> (Jui dkk., 2022).	Satelit MERRA-2 dan CERES dengan 22 variabel hidro-meteorologi.	Prediksi hasil panen teh di Bangladesh.	<i>Random Forest</i> dengan <i>Dragonfly Optimization Algorithm</i> dan <i>Support Vector Machine</i> (SVM).	Model DRS–RF mencapai koefisien korelasi 0,993 dan MAPE 11,95%, menunjukkan peningkatan akurasi dibandingkan model standalone.	Menjadi dasar penggunaan model <i>Random Forest</i> dengan optimasi parameter untuk meningkatkan akurasi prediksi.
3	<i>Spatiotemporal Analysis of Mangrove in Subang Regency</i>	Data <i>Sentinel-2</i> (13 band spektral),	Peta distribusi spasial dan temporal	<i>Random Forest Classifier</i> .	Model mencapai <i>Overall Accuracy</i> 90.38%, <i>kappa</i>	Menjadi dasar penerapan analisis perubahan spasial

No	Judul	Data Input	Output	Metode	Hasil	Relasi dengan Penelitian Ini
	<i>using Sentinel-2 Timeseries Data (Dzulfigar dkk., 2024).</i>	beberapa indeks vegetasi (NDVI, SAVI, EVI) dan indeks air (MNDWI).	<i>mangrove</i> di Kabupaten Subang.		86.67%, dengan akurasi deteksi <i>mangrove</i> mencapai 98.44%.	dan temporal pada data tutupan lahan.
4	<i>Predictive Modelling of Land Cover Changes in the Greater Amanzule Peatlands Using Multi-Source Remote Sensing and Machine Learning Techniques (Amoakoh dkk., 2024).</i>	Data multisensor satelit <i>Landsat 7, Landsat 8, ALOS/PALSAR</i> , DEM.	Prediksi perubahan tutupan lahan di <i>Greater Amanzule Peatlands</i> hingga tahun 2040.	<i>Random Forest Classification.</i>	Model mendapatkan hasil <i>Overall Accuracy (OA)</i> 93%, <i>kappa</i> 0.70, dan <i>weighted f1-score</i> 80%.	Menjadi dasar penggunaan <i>Random Forest</i> untuk klasifikasi tutupan lahan.
5	<i>Random Forest - based analysis of land cover/land use (LCLU) dynamics associated with meteorological droughts in the desert ecosystem of Pakistan (Faheem dkk., 2024).</i>	Data citra satelit <i>Landsat</i> dari tahun 1990 hingga 2020 dengan indeks vegetasi dan tanah, yaitu NDVI, NDBaI, dan <i>Top Grain Soil Index (TGSI)</i> .	Pemetaan perubahan tutupan lahan dan penggunaan lahan selama tiga dekade.	<i>Random Forest Classifier</i> dengan uji <i>mann-kendall</i> dan estimasi kemiringan.	<i>Random Forest</i> mendapatkan akurasi sebesar 99% dengan model yang dibangun.	Menjadi dasar penggunaan algoritma <i>Random Forest</i> untuk menganalisis dinamika tutupan lahan.
6	<i>Modelling of land use and land cover changes and prediction using CA-</i>	Citra <i>Landsat</i> dari tahun 1990 hingga 2022 dan data elevasi,	Peta perubahan LULC dari tahun 1990 hingga 2022 dan prediksi	<i>Random Forest</i> dan <i>CA-Markov.</i>	Akurasi keseluruhan model <i>Random Forest</i> mencapai 90%.	Menjadi dasar penggunaan <i>Random Forest</i> untuk

No	Judul	Data Input	Output	Metode	Hasil	Relasi dengan Penelitian Ini
	<i>Markov and Random Forest</i> (Asif dkk., 2023).	jarak ke jalan utama, dan jarak ke pusat kota.	LULC untuk tahun 2038.			klasifikasi tutupan lahan.
7	<i>Random Forest Algorithm for the Classification of Spectral Data of Astronomical Objects</i> (Ramírez dkk., 2023).	Kumpulan data spektral yang mencakup tiga jenis objek langit: bintang, galaksi, dan kuasar.	Klasifikasi objek astronomi ke dalam kategori bintang, galaksi, atau kuasar berdasarkan data spektral mereka.	<i>Random Forest</i> dibandingkan dengan <i>Naive Bayes</i> , <i>K-Nearest Neighbors</i> , <i>Support Vector Machines</i> , dan ANN.	Algoritma <i>Random Forest</i> menunjukkan kinerja yang unggul dalam klasifikasi objek astronomi dibandingkan dengan algoritma lain yang diuji.	Menjadi dasar penggunaan <i>Random Forest</i> untuk klasifikasi data.
8	<i>The Random Forest Approach for Land Cover Mapping</i> (Aonpong dkk., 2016).	Data citra satelit dan fitur spektral serta tekstural.	Peta klasifikasi tutupan lahan yang akurat berdasarkan analisis citra satelit.	<i>Random Forest</i> berbasis piksel dan <i>neighbor-looking</i> .	Kombinasi pendekatan berbasis piksel dan <i>neighbor-looking</i> memberikan performa terbaik dalam pemetaan tutupan lahan.	Menjadi dasar penggunaan piksel tetangga dalam <i>feature engineering</i> .
9	<i>Random Forest Classification for Mangrove Land Cover Mapping Using Landsat 5 TM and Alos Palsar Imageries</i> (Jhonnerie dkk., 2015).	Citra <i>Landsat 5 TM</i> dan data ALOS PALSAR FBD dengan NDVI, NDBI, NDWI.	Peta klasifikasi tutupan lahan mangrove di Sungai Kambung, Pulau Bengkalis, Indonesia.	<i>Random Forest</i> dan <i>Maximum Likelihood (ML)</i> untuk perbandingan.	<i>Random Forest</i> menghasilkan OA sebesar 81,1% dan nilai <i>kappa</i> 0,76 lebih unggul dari <i>Maximum Likelihood</i> yang menghasilkan OA sebesar 77,7% dan nilai <i>Kappa</i> 0,71.	Menjadi dasar penggunaan <i>Random Forest</i> untuk klasifikasi tutupan lahan.
10	<i>Imbalanced Learning in Land Cover Classification:</i>	Data LULC , <i>area frame</i>	Peta klasifikasi tutupan lahan dengan fokus pada	G-SMOTE untuk <i>oversampling</i> kelas minoritas; evaluasi	G-SMOTE secara signifikan meningkatkan akurasi	Menjadi dasar penggunaan teknik SMOTE untuk

No	Judul	Data Input	Output	Metode	Hasil	Relasi dengan Penelitian Ini
	<i>Improving Minority Classes' Prediction Accuracy Using the Geometric SMOTE Algorithm</i> (Douzas dkk., 2019).	survey dan citra Landsat 8.	peningkatan akurasi kelas minoritas.	dengan berbagai classifier seperti <i>Logistic Regression</i> , <i>K-NN</i> , <i>Decision Tree</i> , <i>Gradient Boosting Classifier</i> , dan <i>Random Forest</i> .	prediksi kelas minoritas dibandingkan metode <i>oversampling</i> lainnya, meningkatkan robustitas <i>classifier</i> .	menangani ketidakseimbangan kelas dalam klasifikasi tutupan lahan.
11	<i>An Interpretable Model for Landslide Susceptibility Assessment Based on Optuna Hyperparameter Optimization and Random Forest</i> (Xiao dkk., 2024).	Data inventarisasi 4.447 longsor dan 23 faktor terkait di wilayah barat laut Provinsi Yunnan, Tiongkok.	Peta kerentanan longsor dengan interpretasi faktor-faktor yang mempengaruhi.	Kombinasi <i>Random Forest</i> dengan optimasi <i>hyperparameter</i> menggunakan <i>optuna</i> .	Model yang dioptimalkan dengan <i>optuna</i> menunjukkan peningkatan akurasi dalam evaluasi kerentanan longsor.	Menjadi dasar pemilihan <i>framework optuna</i> untuk melakukan pengujian kombinasi parameter model <i>Random Forest</i> .
12	<i>From Land Cover Map to Land Use Map: A Combined Pixel-Based and Object-Based Approach Using Multi-Temporal Landsat Data, a Random Forest Classifier, and Decision Rules</i> (Bui & Mucsi, 2021).	Citra multi-temporal Landsat-8 untuk tahun 2013–2015.	Peta tutupan lahan dan peta penggunaan lahan terpisah untuk Provinsi Binh Duong, Vietnam.	Kombinasi klasifikasi berbasis piksel dan objek menggunakan algoritma <i>Random Forest</i> .	Peta tutupan lahan mencapai OA 93,86% dan peta penggunaan lahan mencapai akurasi keseluruhan 93,45%.	Menjadi dasar penggunaan kombinasi pendekatan berbasis piksel dan objek dalam penggunaan algoritma <i>Random Forest</i> .
13	<i>Per-pixel Land Cover Accuracy Prediction:</i>	Citra satelit multi-sumber	Peta prediksi akurasi per-piksel	<i>Random Forest Classifier</i> yang	Metode ini efektif dalam memprediksi	Menjadi dasar penggunaan <i>Random</i>

No	Judul	Data Input	Output	Metode	Hasil	Relasi dengan Penelitian Ini
	<i>A Random Forest based Method with Limited Reference Sample Data</i> (Ebrahimiy dkk., 2021).	dan data referensi terbatas untuk pelatihan.	untuk klasifikasi tutupan lahan.	diterapkan untuk memprediksi akurasi klasifikasi pada tingkat piksel.	akurasi per-piksel, memungkinkan identifikasi area dengan ketidakpastian tinggi dalam peta tutupan lahan.	<i>Forest</i> untuk memprediksi akurasi klasifikasi pada tingkat piksel dengan data referensi terbatas.
14	Land Cover Classification Using Integrated Spectral, Temporal, and Spatial Features Derived from Remotely Sensed Images (Zhai dkk., 2018).	Citra satelit multi-temporal dari <i>Landsat</i> untuk tahun 2013–2015.	Peta klasifikasi tutupan lahan dengan peningkatan akurasi.	Metode klasifikasi yang dengan integrasi informasi spektral, temporal, dan spasial, reduksi dimensi non-linear, dan segmentasi grafis untuk ekstraksi fitur spasial.	Metode yang diusulkan meningkatkan akurasi klasifikasi tutupan lahan dengan memanfaatkan informasi spektral, temporal, dan spasial secara terpadu.	Menjadi dasar penggunaan integrasi fitur temporal dan spasial untuk meningkatkan akurasi klasifikasi tutupan lahan.
15	<i>Comparison of accuracy and reliability of Random Forest , support vector machine, artificial neural network and maximum likelihood method in land use/cover classification of urban setting</i> (Chowdhury, 2024).	Citra satelit multi-sumber dari area perkotaan.	Peta klasifikasi penggunaan dan tutupan lahan di lingkungan perkotaan.	Perbandingan algoritma <i>Random Forest</i> (RF), <i>Support Vector Machine</i> (SVM), <i>Artificial Neural Network</i> (ANN), dan <i>Maximum Likelihood</i> (ML).	RF menunjukkan akurasi keseluruhan tertinggi dibandingkan dengan algoritma pembandingan lainnya, berkisar antara 81,72% hingga 90,71%.	Menjadi dasar dalam memilih algoritma klasifikasi untuk analisis tutupan lahan yaitu <i>Random Forest</i> .

## II.2 Pemilihan Kerangka Kerja

Tabel I-2 Perbandingan Kerangka Kerja

Karakteristik	CRISP-DM	SEMMA	KDD	Penelitian Ini
Fokus dan tujuan	Digunakan untuk menyelesaikan permasalahan bisnis secara terstruktur.	Digunakan untuk pengembangan model dengan evaluasi mendalam.	Digunakan untuk menemukan pola atau pengetahuan baru yang belum dikenali sebelumnya.	Membutuhkan kerangka kerja yang berfokus untuk menemukan pengetahuan baru.
Pendekatan	Iteratif dengan mencoba model berdasarkan kebutuhan bisnis.	Penekanan bersifat statistik pada pemodelan data	Eksploratif dengan melakukan banyak eksperimen untuk menemukan pola terbaik.	Membutuhkan kerangka kerja dengan pendekatan yang lebih eksploratif.
Kebutuhan Data	Cocok untuk data yang cukup terstruktur.	Cocok untuk data yang terstruktur dan siap pakai.	Cocok untuk data yang kompleks dan tidak terstruktur	Membutuhkan kerangka kerja yang mampu menangani data kompleks.

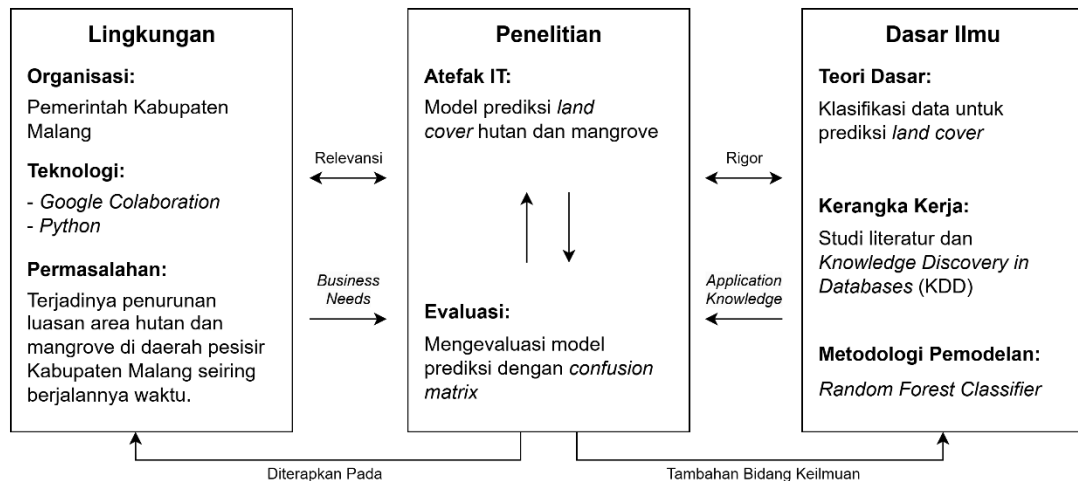
Berdasarkan Tabel I-2, kerangka kerja yang dipilih dalam penelitian ini adalah *Knowledge Discovery in Databases* (KDD) karena pendekatannya yang eksploratif dalam menggali pola serta menemukan pengetahuan baru dari data yang kompleks dan tidak terstruktur. KDD sesuai dengan kebutuhan penelitian ini karena fokusnya yang lebih spesifik pada proses pengolahan data mentah menjadi informasi yang dapat diinterpretasikan, dibandingkan dengan kerangka kerja lainnya yang lebih luas dan mencakup aspek bisnis maupun implementasi model. Dengan pendekatan berbasis eksplorasi dan eksperimen, KDD memungkinkan penelitian ini untuk mengeksplorasi berbagai kemungkinan pola dalam memprediksi perubahan hutan *mangrove* di Kabupaten Malang, sehingga dapat memperoleh wawasan yang lebih luas dan komprehensif.

## BAB III METODE PENYELESAIN MASALAH

### III.1 Sistematika Penyelesaian Masalah

#### III.1.1 Pengembangan Model Konseptual

Model konseptual adalah representasi abstrak dari sistem, proses, atau fenomena yang dipelajari. Model ini berperan sebagai gambaran yang jelas dan sistematis dari berbagai konsep yang terkait serta hubungan di antaranya, sehingga memudahkan dalam melakukan analisis, evaluasi, dan pengembangan sistem atau proses tersebut.



Gambar III.1 Model Konseptual

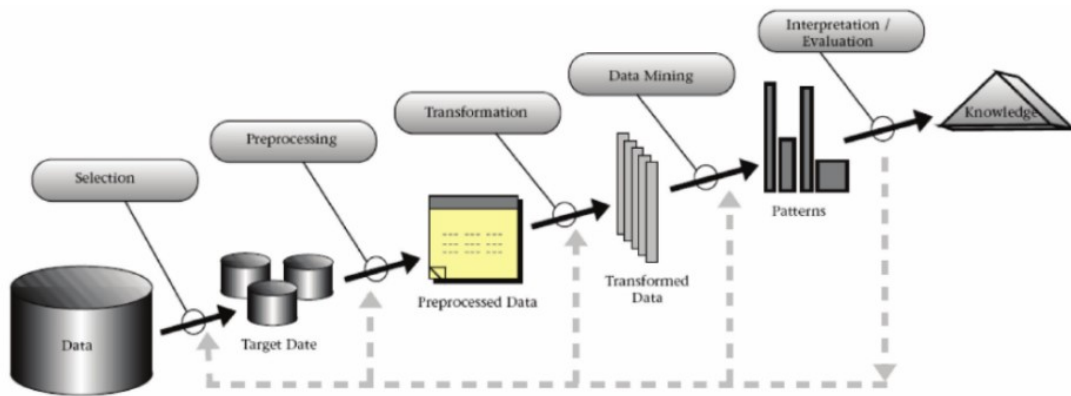
Berdasarkan Gambar III.1, terdapat tiga komponen pembahasan dalam model konseptual yang digunakan, yaitu lingkungan, penelitian, dan dasar ilmu. Tujuan dirancangnya model konseptual adalah untuk membangun solusi dari permasalahan di Pantai Selatan Kabupaten Malang, yaitu terjadinya penurunan luasan area hutan *mangrove* seiring berjalannya waktu. Oleh karena itu, solusi yang dibangun berfokus pada pembangunan model prediksi *land cover* berdasarkan data historis tahunan menggunakan algoritma *Random Forest*. Hasil dari model prediksi yang dibangun selanjutnya dievaluasi untuk mengukur tingkat keberhasilan dan akurasi prediksi menggunakan *confusion matrix* yang mencakup empat metrik utama yaitu *accuracy*, *precision*, *recall*, dan *f-1 score*. Adapun kerangka kerja atau metodologi yang digunakan dalam penelitian ini adalah kerangka kerja *Knowledge Discovery in*



*Databases* (KDD). Pemilihan metodologi ini didasarkan pada kebutuhan kerangka kerja yang berfokus untuk menemukan pengetahuan baru, memiliki pendekatan yang lebih eksploratif, serta mampu menangani data yang kompleks secara sistematis.

### III.1.2 Metode Penyelesaian Masalah

Metode penyelesaian masalah dalam penelitian ini dirancang berdasarkan langkah-langkah terstruktur untuk mengidentifikasi, menganalisis, dan menyelesaikan permasalahan penurunan luasan area hutan *mangrove* di Pantai Selatan Kabupaten Malang. Metodologi yang digunakan adalah *Knowledge Discovery in Databases* (KDD), yang merupakan pendekatan sistematis dalam *data mining* dan eksplorasi data yang menekankan pada penemuan pola yang dapat diinterpretasikan sebagai pengetahuan yang berguna atau menarik (Fayyad dkk., 1996). Model yang dihasilkan tidak hanya bertujuan untuk menjelaskan fenomena yang diamati dalam data tetapi juga digunakan untuk keperluan analisis dan prediksi (Maimon & Rokach, 2006).

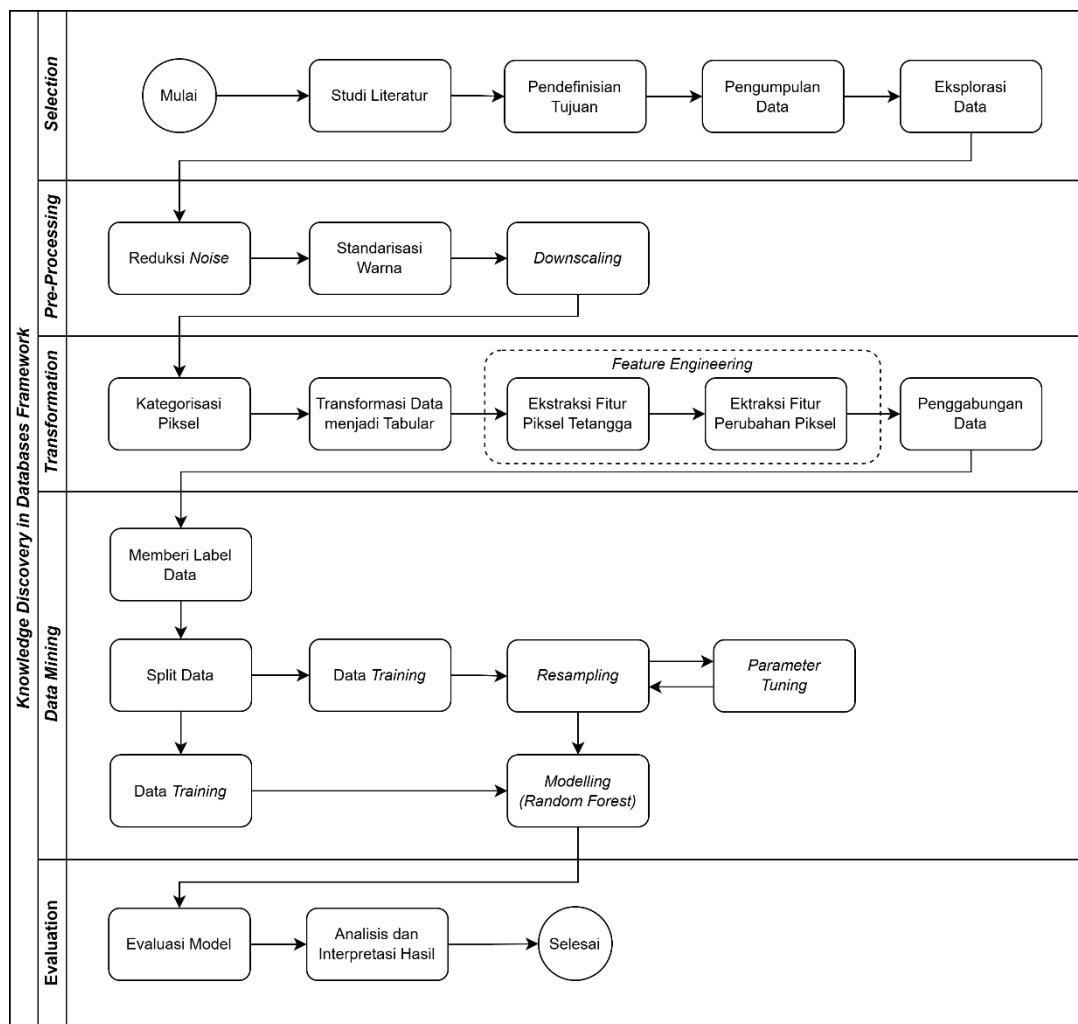


Gambar III.2 Ilustrasi Kerangka Kerja KDD

Sumber: Fayyad dkk., 1996

Berdasarkan Gambar III.2, proses KDD diawali dengan *data selection*, yaitu pemilihan data relevan dari sumber yang tersedia, baik data primer maupun sekunder, yang sesuai dengan tujuan penelitian. Setelah data terpilih, langkah berikutnya adalah *data preprocessing*, yang melibatkan pembersihan data untuk mengatasi masalah seperti data yang hilang, inkonsisten, atau *noise*, sehingga data menjadi lebih terstruktur dan siap untuk diolah. Selanjutnya adalah tahap *data transformation*, yaitu data yang telah

diproses diubah menjadi format atau fitur yang lebih cocok untuk analisis, seperti normalisasi, agregasi, atau pengelompokan variabel penting. Tahap inti dari KDD adalah *data mining*, yang menggunakan algoritma atau metode tertentu untuk mengeksplorasi pola, membuat model, atau menemukan hubungan tersembunyi di dalam data. Terakhir, hasil dari proses data mining dievaluasi pada tahap *evaluation*, yaitu model atau pola yang ditemukan diuji menggunakan metrik tertentu untuk memastikan bahwa hasilnya valid, relevan, dan dapat digunakan untuk mendukung pengambilan keputusan.



Gambar III.3 Sistematika Penelitian menggunakan KDD

Gambar III.3 menunjukkan penerapan kerangka kerja *Knowledge Discovery in Databases* (KDD) dalam penelitian ini. Tahap *data selection* mencakup studi literatur, pendefinisian tujuan, pengumpulan data, dan eksplorasi data untuk memahami konteks serta kebutuhan penelitian terkait prediksi perubahan hutan *mangrove* di Kabupaten Malang. Setelah itu, pada tahap *data pre-processing*, dilakukan proses reduksi noise, standarisasi warna, dan downscaling. Tahap berikutnya adalah *data transformation*, yang bertujuan untuk mengubah data mentah menjadi format yang lebih terstruktur dengan melakukan kategorisasi piksel, transformasi data menjadi tabular, proses *feature engineering*, lalu digabungkan untuk mendapatkan dataset yang siap digunakan untuk pemodelan. Pada tahap *data mining*, dilakukan pemberian label pada data lalu pembagian data menjadi *data training* dan *data testing* serta proses *resampling* untuk menangani ketidakseimbangan data. Model prediksi dibangun menggunakan algoritma *Random Forest*, dengan dilakukan pengujian kombinasi parameter untuk memperoleh performa terbaik. Tahap akhir adalah *evaluation*, di mana model yang telah dibangun dievaluasi menggunakan *confusion matrix* dan perhitungan *accuracy*, *precision*, *recall*, dan *f1-score*. Terakhir, dilakukan analisis dan interpretasi hasil untuk mendapatkan wawasan yang lebih komprehensif terkait perubahan hutan *mangrove*.

### **III.1.3 Metode Pengumpulan Data**

Tahap pengumpulan data pada penelitian ini dilakukan melalui proses wawancara oleh pihak sumber, yaitu Politeknik Negeri Malang. Data yang didapatkan pada wawancara merupakan hasil dokumentasi data berupa Peta Persebaran *Mangrove* dan Garis Pantai dari 18 titik pantai di kawasan Pantai Selatan Kabupaten Malang yang secara geografis dibatasi oleh koordinat antara bujur  $112.0371^{\circ}$  hingga  $112.9468^{\circ}$  dan lintang  $-7.2299^{\circ}$  hingga  $-8.4021^{\circ}$  dan mencakup rentang waktu dari tahun 2018 hingga tahun 2023. Data peta tersebut mencakup gambar tutupan lahan yang sudah tersegmentasi yang diambil dari digitalisasi citra satelit *Google Earth Imagery* untuk periode tahun 2018 hingga 2023 serta data batas administratif yang diambil dari *Global Administrative Areas Database* (GADM) Indonesia tahun 2019.

### III.1.4 Metode Pengolahan Data

Pengolahan data merupakan langkah penting dalam penelitian ini untuk memastikan data memiliki kualitas yang baik, konsisten, dan siap digunakan dalam proses analisis. Tahap pertama dalam pengolahan data adalah *pre-processing*, yang bertujuan untuk meningkatkan kualitas data dan mengurangi gangguan yang dapat memengaruhi hasil analisis. Pada tahap ini, dilakukan langkah reduksi *noise* untuk menghilangkan gangguan berupa tutupan awan yang ada pada data. Selanjutnya, dilakukan standarisasi warna yaitu data dikoreksi agar setiap klasifikasi penutup lahan dapat dibedakan dengan jelas. Selanjutnya *downscaling* data juga dilakukan untuk menyesuaikan resolusi data agar lebih efisien dalam pengolahan tanpa kehilangan informasi penting. Tahap selanjutnya adalah transformasi data, yang melibatkan sejumlah langkah untuk mengubah data mentah menjadi format yang sesuai untuk proses analisis dan pemodelan. Proses ini mencakup kategorisasi piksel, yaitu mengelompokkan piksel-piksel berdasarkan nilai warna RGB. Data kemudian diubah menjadi bentuk tabular sehingga memudahkan untuk diproses menggunakan algoritma pemodelan. Untuk meningkatkan kualitas model, dilakukan *feature engineering* melalui ekstraksi fitur piksel tetangga, yang memperhitungkan informasi dari piksel di sekitarnya, serta perubahan piksel, yang mempertimbangkan perubahan temporal untuk setiap piksel. Terakhir, semua data yang telah melalui *pre-processing* dan transformasi digabungkan menjadi satu dataset komprehensif yang siap digunakan dalam tahap analisis lebih lanjut.

### III.1.5 Metode Evaluasi

Evaluasi merupakan tahap penting dalam penelitian ini sebelum dilakukan analisis hasil, dengan tujuan untuk menilai seberapa baik dan akurat performa model dalam memprediksi perubahan penutup lahan, khususnya dalam mengidentifikasi pertumbuhan dan kerusakan hutan *mangrove*. Evaluasi dilakukan menggunakan *confusion matrix*, yang merupakan alat analisis berbasis matriks yang membandingkan prediksi model dengan label aktual untuk mengidentifikasi keberhasilan dan kesalahan klasifikasi pada setiap kelas. *confusion matrix* tidak hanya memberikan gambaran

sederhana tentang jumlah prediksi yang benar dan salah, tetapi juga menjadi dasar untuk perhitungan metrik evaluasi yang lebih mendalam, yaitu *accuracy*, *precision*, *recall*, dan *f1-score*. *Accuracy* digunakan untuk mengukur tingkat kesesuaian prediksi model terhadap data aktual, *precision* mengukur tingkat ketepatan model dalam mengklasifikasikan kelas tertentu, *recall* berfokus pada sensitivitas model dalam mendeteksi seluruh instance yang benar dari suatu kelas, dan *f1-score* menjadi metrik yang menggabungkan *precision* dan *recall* dan memberikan gambaran kinerja model secara keseluruhan dengan mempertimbangkan keseimbangan antara kedua metrik tersebut, terutama pada dataset yang tidak seimbang.

## BAB IV PENYELESAIAN MASALAH

### IV.1 Data Selection

*Data selection* merupakan tahap pemilihan data yang digunakan dalam penelitian dan memiliki peran penting dalam menentukan kualitas analisis yang dilakukan. Pada tahap ini, proses yang dilakukan meliputi eksplorasi data untuk memahami karakteristik awal dataset yang didapatkan berdasarkan hasil pengumpulan data serta pemilihan dataset yang sesuai dengan kebutuhan penelitian berdasarkan eksplorasi yang telah dilakukan.

#### IV.1.1 Eksplorasi Data

Dataset yang bersumber dari dokumentasi Politeknik Negeri Malang berupa Peta Persebaran *Mangrove* dan Garis Pantai dari 18 titik pantai yang dapat dilihat pada Tabel IV-1. Tiap titik pantai memiliki data peta tahunan dengan rentang tahun 2018 sampai 2023 seperti yang ditampilkan secara rinci pada Tabel IV-2.

Tabel IV-1 Data Peta Persebaran *Mangrove* dan Garis Pantai

No.	Nama Pantai	Lokasi	Koordinat
1	Pantai Tamban	Desa Tambakrejo, Kecamatan Sumbermanjing Wetan	<i>Longitude:</i> 112.7018 <i>Latitude:</i> -7.46659
2	Pulau Sempu	Desa Tambakrejo, Kecamatan Sumbermanjing Wetan	<i>Longitude:</i> 112.51150596534 <i>Latitude:</i> -7.2298789341889
3	Pantai Goa Cina	Desa Sitarjo, Kecamatan Sumbermanjing Wetan	<i>Longitude:</i> 112.6034239139 <i>Latitude:</i> -7.9639513169792
4	Pantai Balekambang	Desa Srigonco, Kecamatan. Bantur	<i>Longitude:</i> 112.6094945 <i>Latitude:</i> -7.954794
5	Pantai Ngliyep	Desa Kedungsalam, Kecamatan Donomulyo	<i>Longitude:</i> 112.037134165 <i>Latitude:</i> -8.027891665
6	Pantai Sendang Biru	Desa Tambakrejo, Kecamatan Sumbermanjing Wetan	<i>Longitude:</i> 112.65934 <i>Latitude:</i> -8.26647
7	Pantai Jembatan Panjang	Desa Srigoco, Kecamatan Bantur	<i>Longitude:</i> 112.53021240234 <i>Latitude:</i> -8.4020736287669
8	Pantai Kondang Merak	Desa Sumberbening, Kecamatan Bantur	<i>Longitude:</i> 112.7156779 <i>Latitude:</i> -7.9588961
9	Pantai Bajul Mati	Desa Gajahrejo, Kecamatan Gedangan	<i>Longitude:</i> 112.60025840336 <i>Latitude:</i> -7.9428244537815

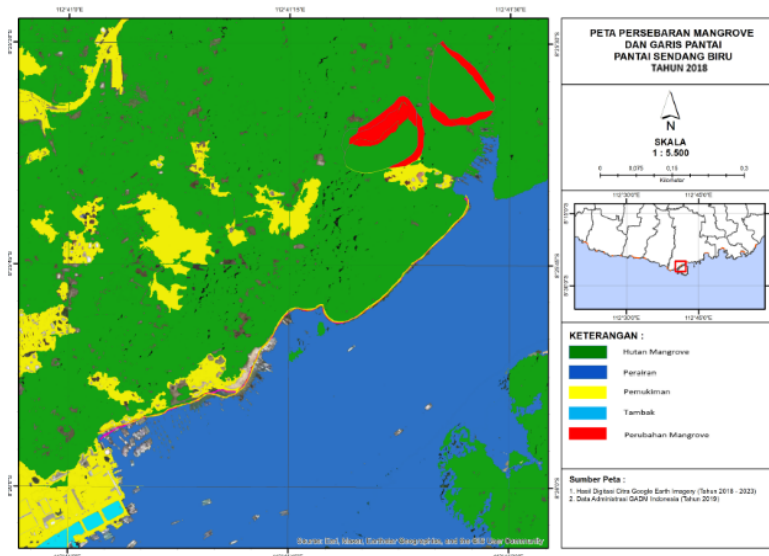
No.	Nama Pantai	Lokasi	Koordinat
10	Pantai Modangan	Desa Sumberoto, Kecamatan Donomulyo	<i>Longitude: 112.359496</i> <i>Latitude: -8.348821</i>
11	Pantai Jonggring Saloko	Desa Mentaraman, Kecamatan Donomulyo	<i>Longitude: 112.3961</i> <i>Latitude: -8.379265</i>
12	Pantai Bantol	Desa Banjarejo, Kecamatan Donomulyo	<i>Longitude: 112.452129</i> <i>Latitude: -8.395116</i>
13	Pantai Kondang Iwak	Desa Tulungrejo, Kecamatan Donomulyo	<i>Longitude: 112.483216</i> <i>Latitude: -8.395517</i>
14	Pantai Wonogoro	Desa Tumpakrejo, Kecamatan Gedangan	<i>Longitude: 112.565455</i> <i>Latitude: -8.399322</i>
15	Pantai Tambaksari	Desa Sidoasri, Kecamatan Sumbermanjing	<i>Longitude: 112.773109</i> <i>Latitude: -8.385217</i>
16	Pantai Lenggoksono	Desa Purwodadi, Kecamatan Tirto Yudo	<i>Longitude: 112.838931</i> <i>Latitude: -8.372234</i>
17	Pantai Sipelot	Desa Tirtoyudo, Kecamatan Gunung	<i>Longitude: 112.896795</i> <i>Latitude: -8.380837</i>
18	Pantai Licin	Desa Lebakharjo, Kecamatan Ampelgading	<i>Longitude: 112.946832</i> <i>Latitude: -8.363322</i>

Tabel IV-2 Data 18 Titik Pantai

No.	Nama Pantai	Jumlah Data	Tahun
1	Pantai Tamban	6	2018, 2019, 2020, 2021, 2022, 2023
2	Pulau Sempu	6	2018, 2019, 2020, 2021, 2022, 2023
3	Pantai Goa Cina	6	2018, 2019, 2020, 2021, 2022, 2023
4	Pantai Balekambang	6	2018, 2019, 2020, 2021, 2022, 2023
5	Pantai Ngliyep	6	2018, 2019, 2020, 2021, 2022, 2023
6	Pantai Sendang Biru	6	2018, 2019, 2020, 2021, 2022, 2023
7	Pantai Jembatan Panjang	6	2018, 2019, 2020, 2021, 2022, 2023
8	Pantai Kondang Merak	6	2018, 2019, 2020, 2021, 2022, 2023
9	Pantai Bajul Mati	6	2018, 2019, 2020, 2021, 2022, 2023
10	Pantai Modangan	6	2018, 2019, 2020, 2021, 2022, 2023
11	Pantai Jonggring Saloko	6	2018, 2019, 2020, 2021, 2022, 2023
12	Pantai Bantol	6	2018, 2019, 2020, 2021, 2022, 2023
13	Pantai Kondang Iwak	6	2018, 2019, 2020, 2021, 2022, 2023

No.	Nama Pantai	Jumlah Data	Tahun
14	Pantai Wonogoro	6	2018, 2019, 2020, 2021, 2022, 2023
15	Pantai Tambaksari	1	2023
16	Pantai Lenggoksono	1	2023
17	Pantai Sipelot	6	2018, 2019, 2020, 2021, 2022, 2023
18	Pantai Licin	6	2018, 2019, 2020, 2021, 2022, 2023

Peta Persebaran *Mangrove* dan Garis Pantai pada dataset mencakup beberapa informasi di dalamnya. Peta tersebut memiliki gambar tutupan lahan yang telah tersegmentasi menjadi lima warna, yaitu warna hijau yang mewakili hutan *mangrove*, warna biru tua yang mewakili perairan, warna kuning yang mewakili pemukiman, warna biru muda yang mewakili tambak, dan warna merah yang mewakili penurunan hutan *mangrove*. Selanjutnya, terdapat informasi skala dari tutupan lahan pada peta yang berbeda-beda di tiap titik pantai. Terakhir, terdapat informasi gambar yang menunjukkan letak area tutupan lahan secara garis besar pada daerah pesisir Kabupaten Malang. Adapun salah satu sampel data peta dapat dilihat pada Gambar IV.1.



Gambar IV.1 Peta Persebaran *Mangrove* dan Garis Pantai Sendang Biru Tahun 2018

#### IV.1.2 Pemilihan *Dataset*

Pada tahap pemilihan dataset dalam penelitian ini, dilakukan seleksi berdasarkan kelengkapan data historis yang tersedia di setiap titik pantai. Dari hasil eksplorasi data



awal, ditemukan total 18 titik pantai yang terdata dalam wilayah pesisir Kabupaten Malang. Namun, dua di antaranya, yaitu Pantai Lenggoksono dan Pantai Tambaksari, hanya memiliki data untuk satu tahun, yaitu tahun 2023. Mengingat bahwa penelitian ini menggunakan analisis berbasis *time series* yang memerlukan data historis yang berkelanjutan, kedua pantai tersebut tidak diikutsertakan dalam analisis lebih lanjut.

Oleh karena itu, dipilih 16 pantai yang memiliki data lengkap dan kontinu dari tahun 2018 hingga 2023 untuk memastikan ketersediaan informasi yang cukup dalam menganalisis perubahan penutup lahan *mangrove* secara temporal. Pantai-pantai yang terpilih meliputi Pantai Tamban, Pantai Sempu, Pantai Goa Cina, Pantai Balekambang, Pantai Ngliyep, Pantai Sendang Biru, Pantai Jembatan Panjang, Pantai Kondang Merak, Pantai Bajul Mati, Pantai Modangan, Pantai Jonggring Saloko, Pantai Bantol, Pantai Kondang Iwak Sumberpucung, Pantai Wonogoro, Pantai Sipelot, dan Pantai Licin.

## **IV.2 Data Pre-processing**

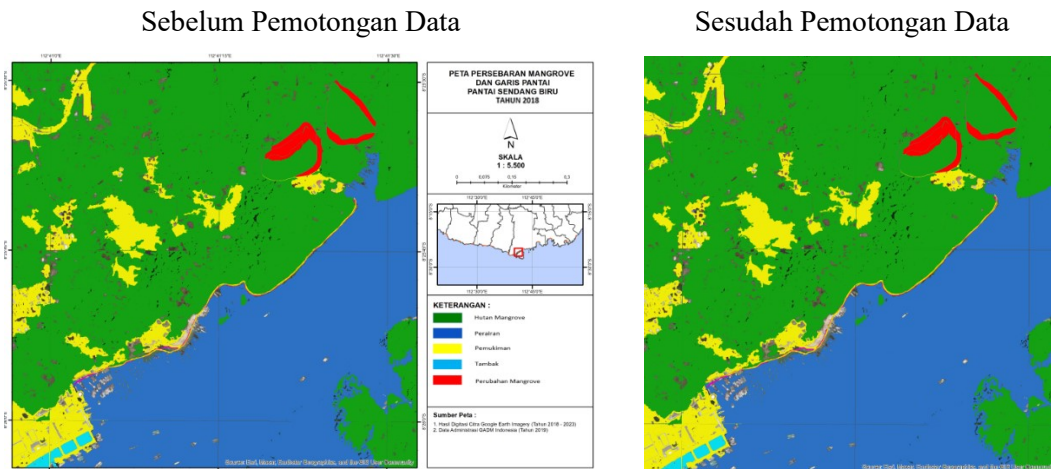
Proses *data pre-processing* merupakan langkah penting dalam memastikan kualitas data yang digunakan sesuai dengan kebutuhan analisis. Adapun beberapa tahapan yang dilakukan dalam pre-processing yaitu pemotongan gambar, reduksi *noise*, standarisasi warna, dan *downscaling data*.

### **IV.2.1 Pemotongan Data**

Tahap pemotongan data dilakukan karena dataset yang sudah terpilih untuk dianalisis masih berupa Peta Persebaran *Mangrove* dan Garis Pantai yang memiliki elemen tambahan seperti legenda, informasi skala, serta keterangan lainnya. Maka dari itu, diperlukan serangkaian langkah untuk menghilangkan informasi yang tidak relevan dan memastikan area yang dipilih sesuai dengan kebutuhan penelitian dan memiliki ukuran yang seragam.

Proses pemotongan data dilakukan menggunakan *tools Figma*. Setiap data peta dipotong mengikuti ukuran area tutupan lahan yang ada pada Peta Persebaran *Mangrove* dan Garis Pantai dalam bentuk persegi dan menghasilkan dimensi 590 x 590

piksel pada 16 titik pantai dengan data historis dari tahun 2018 hingga 2023. Pemilihan ukuran ini bertujuan untuk menjaga konsistensi dalam analisis spasial dan memastikan bahwa data yang diambil mewakili area penutup lahan secara proporsional. Hasil dari proses pemotongan data ditampilkan pada Gambar IV.2, yang menunjukkan perbandingan antara peta asli dan hasil pemotongan yang difokuskan pada area penutup lahan.

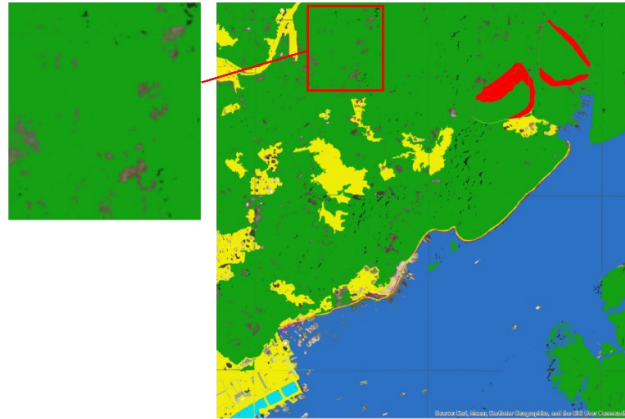


Gambar IV.2 Perbandingan Sebelum dan Sesudah Pemotongan Data

#### IV.2.2 Reduksi *Noise*

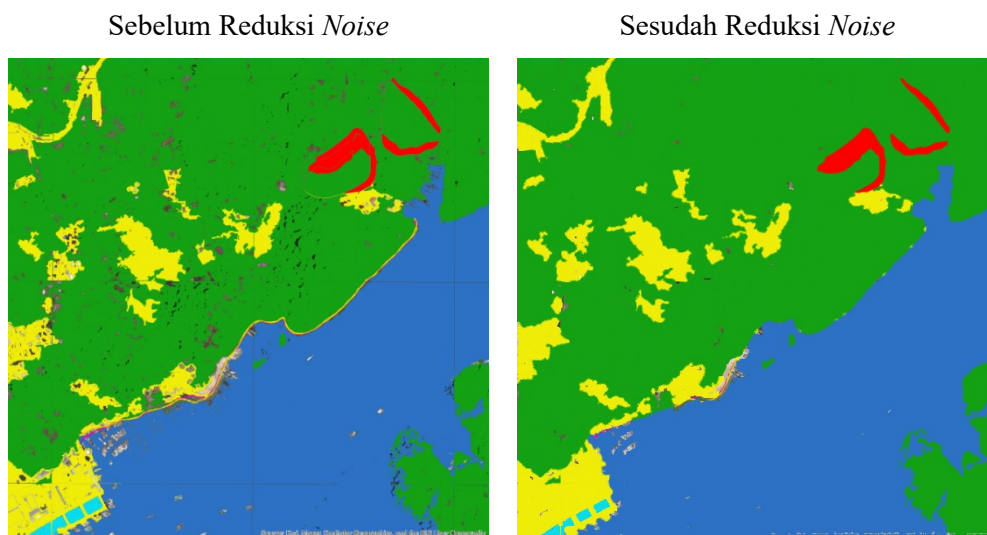
Tahap reduksi *noise* dilakukan untuk meningkatkan kualitas data dengan menghilangkan elemen gangguan yang dapat mengurangi keakuratan hasil analisis. *Noise* merupakan gangguan visual berupa piksel acak atau elemen tidak diinginkan yang tidak mencerminkan informasi sebenarnya dalam data. Dalam konteks penelitian ini, *noise* merujuk pada piksel dengan warna yang tidak sesuai dengan legenda peta, yang seharusnya hanya mencakup warna spesifik untuk klasifikasi hutan, penurunan hutan, tambak, perairan, dan area pemukiman. Namun, ditemukan adanya piksel dengan warna di luar klasifikasi tersebut, seperti abu-abu, hitam, atau putih. Ketidaksiuaian ini dapat disebabkan oleh elemen eksternal seperti tutupan awan yang mengaburkan area peta yang sebenarnya. Pada data penutup lahan berbasis citra, keberadaan *noise* dapat menyebabkan distorsi pada pola yang sebenarnya, sehingga

mempengaruhi proses analisis yang membutuhkan informasi visual yang akurat. Contoh keberadaan *noise* pada data disajikan pada Gambar IV.3.



Gambar IV.3 *Noise* pada Gambar

Untuk mengatasi masalah ini, penelitian ini menggunakan metode *spatial filtering* dengan teknik *mean filtering*. Teknik ini merupakan metode pemfilteran yang bekerja dengan menganalisis nilai piksel di sekitar area yang terdeteksi mengandung *noise*. Piksel yang teridentifikasi sebagai gangguan digantikan dengan nilai rata-rata dari piksel di sekitarnya untuk menghaluskan perbedaan yang tidak relevan. Pendekatan ini terbukti mampu memperbaiki kejernihan citra dan menjaga representasi visual yang lebih akurat. Hasil dari proses reduksi *noise* ditampilkan pada Gambar IV.4.

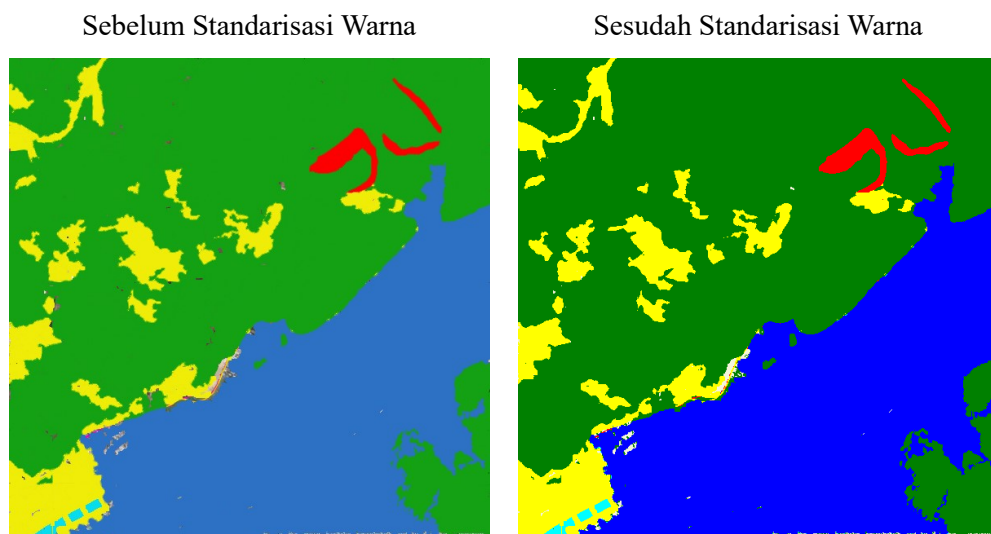


Gambar IV.4 Perbandingan Sebelum dan Sesudah Reduksi *Noise*

### IV.2.3 Standarisasi Warna

Setelah melalui tahap reduksi *noise*, data gambar penutup lahan selanjutnya diproses dengan tahap standarisasi warna. Proses ini dilakukan karena pada saat eksplorasi data ditemukan beberapa gambar yang menunjukkan perbedaan tingkatan warna antara satu titik pantai dengan titik lainnya serta inkonsistensi pada masing-masing segmentasi warna dalam satu gambar. Oleh karena itu, standarisasi warna diterapkan untuk menyelaraskan nilai RGB setiap piksel agar dapat diinterpretasikan secara konsisten dalam proses klasifikasi.

Proses standarisasi warna bertujuan untuk menyamakan representasi warna di seluruh dataset, sehingga setiap kelas penutup lahan, dapat diidentifikasi dengan lebih konsisten. Dengan penyelarasan warna ini, variasi antar lokasi dapat diminimalkan dan inkonsistensi dalam satu gambar dapat dikurangi, sehingga model memiliki referensi warna yang lebih seragam dalam proses klasifikasi. Dengan diterapkannya standarisasi warna, data gambar penutup lahan menjadi lebih homogen, memungkinkan model untuk bekerja lebih efektif dalam mengenali pola dan membedakan setiap kelas dengan lebih akurat. Perbandingan visual antara gambar sebelum dan sesudah proses standarisasi warna dapat dilihat pada Gambar IV.5.

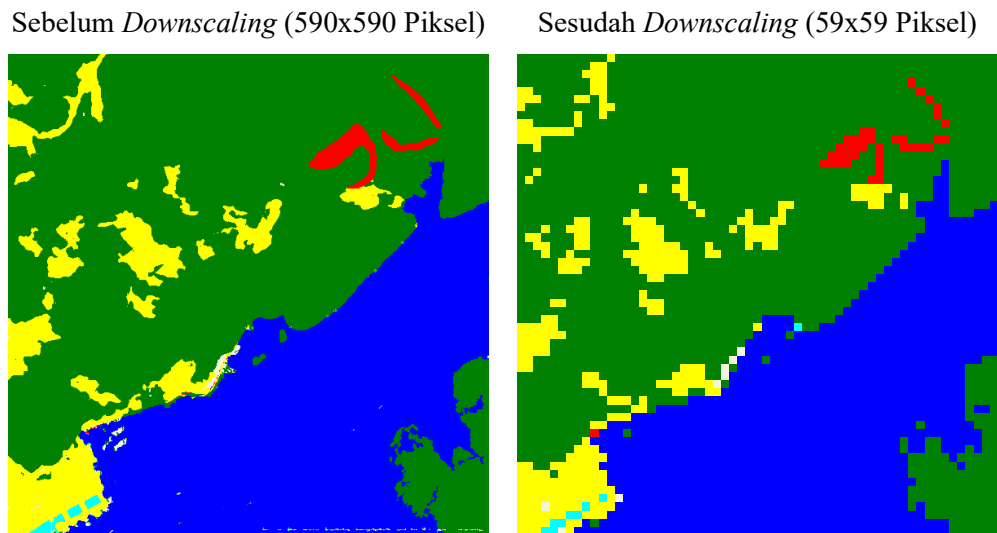


Gambar IV.5 Perbandingan Sebelum dan Sesudah Standarisasi Warna

#### IV.2.4 *Downscaling Data*

Tahap selanjutnya dalam proses *data pre-processing* adalah *downscaling* data, yaitu tahap mengurangi resolusi citra penutup lahan dari ukuran awal 590x590 piksel menjadi 59x59 piksel. Proses ini dilakukan menggunakan metode *nearest-neighbor interpolation*, yang dipilih karena kesederhanaannya dan kemampuannya dalam mempertahankan integritas nilai piksel asli tanpa memperkenalkan noise atau distorsi tambahan pada data.

Proses *downscaling* bertujuan untuk mengurangi ukuran data secara signifikan, sehingga lebih efisien secara komputasi saat diproses dalam model klasifikasi. Dengan ukuran yang lebih kecil, waktu pemrosesan menjadi lebih cepat, namun tetap mempertahankan informasi spasial yang penting untuk analisis penutup lahan. Metode ini memastikan bahwa pola spasial yang ada dalam citra tetap dapat terdeteksi dengan baik. Adapun hasil dari *downscaling* data dapat dilihat pada Gambar IV.6.



Gambar IV.6 Perbandingan Sebelum dan Sesudah *Downscaling*

#### IV.3 *Data Transformation*

Setelah melalui proses *data pre-processing*, tahapan selanjutnya dalam penelitian ini adalah transformasi data. Tahap ini bertujuan untuk mengubah data citra penutup lahan menjadi format yang dapat diolah secara optimal oleh model. Transformasi data dilakukan dengan mengkategorikan setiap piksel berdasarkan nilai warnanya,

mengubah data gambar menjadi format yang sesuai untuk analisis, serta menerapkan *feature engineering* guna memperkaya informasi yang tersedia dalam dataset. Setelah tahap transformasi selesai, data digabungkan menjadi satu.

### IV.3.1 Kategorisasi Piksel

Tahap pertama dalam proses transformasi data adalah kategorisasi piksel, yang bertujuan untuk mengubah nilai numerik piksel menjadi label klasifikasi yang sesuai dengan kelas penutup lahan yang diamati. Pada penelitian ini, fitur utama yang diekstraksi dari citra penutup lahan adalah nilai RGB (*Red, Green, Blue*) dari setiap piksel, yang merepresentasikan intensitas warna merah, hijau, dan biru dalam gambar. Setiap nilai RGB kemudian dikategorikan ke dalam label tertentu yang merepresentasikan kelas tutupan lahan berdasarkan ambang batas yang telah ditentukan sebelumnya seperti yang ditampilkan pada Tabel IV-3.

Tabel IV-3 Label Piksel Berdasarkan Nilai RGB

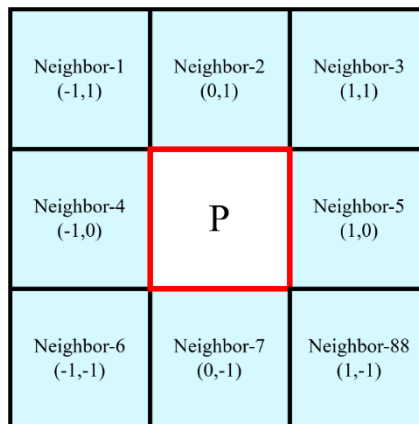
Nilai Merah (R)	Nilai Hijau (G)	Nilai Biru (B)	Keterangan	Warna	Label
255	0	0	Penurunan Hutan	Merah	0
0	128	0	Hutan <i>Mangrove</i>	Hijau	1
0	0	255	Perairan	Biru Tua	2
255	255	0	Pemukiman	Kuning	3
0	255	255	Tambak	Biru Muda	4

Namun, dalam proses pengolahan data, ditemukan bahwa beberapa piksel dengan warna di luar klasifikasi yang telah ditentukan tetap muncul pada citra, seperti piksel berwarna abu-abu, hitam, atau putih yang tidak merepresentasikan kelas penutup lahan yang valid. Piksel-piksel tersebut merupakan residu dari proses reduksi *noise* yang belum sepenuhnya menghilangkan gangguan visual pada data citra. Oleh karena itu, label tambahan yaitu kelas 5 diperkenalkan untuk merepresentasikan piksel-piksel yang tidak sesuai dengan kelas penutup lahan yang ditentukan. Pemberian label ini memungkinkan piksel yang terdeteksi sebagai *noise* dapat diidentifikasi secara terpisah, sehingga model dapat mengabaikannya dalam proses klasifikasi utama.

### IV.3.2 Feature Engineering

Pada tahap *feature engineering*, dilakukan proses rekayasa fitur untuk memperkaya informasi dalam dataset dengan menambahkan fitur yang dapat meningkatkan kemampuan model dalam mengenali pola perubahan penutup lahan secara lebih mendalam. Proses ini cukup penting dalam penelitian ini karena fitur yang dapat diekstraksi langsung dari citra penutup lahan terbatas hanya pada nilai RGB saja. Nilai RGB yang digunakan sudah melalui proses pelabelan menjadi kelas penutup lahan tertentu yang membuat informasi yang tersedia menjadi sederhana dan kurang mendalam dalam merepresentasikan variasi spasial dan temporal dalam data. Dengan adanya keterbatasan tersebut, *feature engineering* dilakukan untuk memperkaya dataset dengan menambahkan fitur tambahan yang dapat membantu model memahami hubungan spasial dan temporal antar piksel. Dua metode utama yang diterapkan dalam tahap ini adalah pengambilan piksel tetangga dan perubahan piksel.

Proses pertama, yaitu pengambilan piksel tetangga, dilakukan dengan menganalisis hubungan antara suatu piksel dengan piksel di sekitarnya. Dalam penelitian ini, digunakan area jendela berukuran 3x3 piksel. Setiap piksel dianalisis bersama dengan nilai piksel dari delapan piksel tetangganya untuk menambahkan informasi spasial yang lebih kaya seperti yang ditampilkan pada Gambar IV.7. Fitur ini memungkinkan model untuk mengenali pola spasial yang lebih kompleks, seperti perbedaan warna yang signifikan antar area dalam citra.



Gambar IV.7 Ukuran Jendela 3x3 dengan Delapan Piksel Tetangga

Selanjutnya, diterapkan fitur perubahan piksel untuk menangkap pola perubahan temporal pada dataset. Proses ini dilakukan dengan membandingkan nilai kelas piksel antara satu tahun dengan tahun berikutnya, di mana jika terjadi perubahan kelas piksel, maka akan diberikan nilai 1, sedangkan jika tidak ada perubahan, nilainya tetap 0. Dengan pendekatan ini, model dapat mengidentifikasi dinamika perubahan penutupan lahan dari waktu ke waktu. Fitur ini memungkinkan model untuk lebih sensitif terhadap perubahan yang terjadi secara bertahap, sehingga perubahan hutan *mangrove* dapat terdeteksi dengan lebih baik.

### IV.3.3 Penggabungan Data

Tahap terakhir dalam proses transformasi data adalah penggabungan data. Seluruh data yang telah diklasifikasikan dan ditambahkan dengan fitur hasil *feature engineering* digabungkan menjadi satu kesatuan dalam bentuk *dataframe*. Proses ini dilakukan dengan menggabungkan data dari seluruh titik pantai yang digunakan dalam penelitian. Setiap data dari satu pantai dikumpulkan, diberi label area, dan disatukan dalam satu struktur data yang terorganisir. Tujuan dari penggabungan ini adalah untuk menyatukan seluruh informasi yang diperlukan dalam format tabular yang seragam. Penggabungan data menghasilkan total 53 fitur yang secara lengkap dapat dilihat pada Tabel IV-4.

Tabel IV-4 Seluruh Fitur pada *Dataframe*

No	Nama Kolom	Type Data	Deskripsi
1	x	<i>Integer</i>	Menunjukkan poisisi piksel pada koordinat x.
2	y	<i>Integer</i>	Menunjukkan poisisi piksel pada koordinat y.
3	<i>Year_2018</i>	<i>Integer</i>	Menunjukkan kelas piksel di tahun 2018.
4	<i>Year_2019</i>	<i>Integer</i>	Menunjukkan kelas piksel di tahun 2019.
5	<i>Year_2020</i>	<i>Integer</i>	Menunjukkan kelas piksel di tahun 2020.
6	<i>Year_2021</i>	<i>Integer</i>	Menunjukkan kelas piksel di tahun 2021.
7	<i>Year_2022</i>	<i>Integer</i>	Menunjukkan kelas piksel di tahun 2022.
8	<i>Year_2023</i>	<i>Integer</i>	Menunjukkan kelas piksel di tahun 2023.
9	<i>Change_2022_2021</i>	<i>Integer</i>	Menunjukkan apakah terdapat perubahan piksel pada tahun 2022 dari tahun 2021.



<b>No</b>	<b>Nama Kolom</b>	<b>Type Data</b>	<b>Deskripsi</b>
10	<i>Change_2021_2020</i>	<i>Integer</i>	Menunjukkan apakah terdapat perubahan piksel pada tahun 2021 dari tahun 2020.
11	<i>Change_2020_2019</i>	<i>Integer</i>	Menunjukkan apakah terdapat perubahan piksel pada tahun 2020 dari tahun 2019.
12	<i>Change_2019_2018</i>	<i>Integer</i>	Menunjukkan apakah terdapat perubahan piksel pada tahun 2019 dari tahun 2018.
13	<i>Neighbor_1_2018</i>	<i>Integer</i>	Menunjukkan piksel tetangga 1 pada tahun 2018.
14	<i>Neighbor_2_2018</i>	<i>Integer</i>	Menunjukkan piksel tetangga 2 pada tahun 2018.
15	<i>Neighbor_3_2018</i>	<i>Integer</i>	Menunjukkan piksel tetangga 3 pada tahun 2018.
16	<i>Neighbor_4_2018</i>	<i>Integer</i>	Menunjukkan piksel tetangga 4 pada tahun 2018.
17	<i>Neighbor_5_2018</i>	<i>Integer</i>	Menunjukkan piksel tetangga 5 pada tahun 2018.
18	<i>Neighbor_6_2018</i>	<i>Integer</i>	Menunjukkan piksel tetangga 6 pada tahun 2018.
19	<i>Neighbor_7_2018</i>	<i>Integer</i>	Menunjukkan piksel tetangga 7 pada tahun 2018.
20	<i>Neighbor_8_2018</i>	<i>Integer</i>	Menunjukkan piksel tetangga 8 pada tahun 2018.
21	<i>Neighbor_1_2019</i>	<i>Integer</i>	Menunjukkan piksel tetangga 1 pada tahun 2019.
22	<i>Neighbor_2_2019</i>	<i>Integer</i>	Menunjukkan piksel tetangga 2 pada tahun 2019.
23	<i>Neighbor_3_2019</i>	<i>Integer</i>	Menunjukkan piksel tetangga 3 pada tahun 2019.
24	<i>Neighbor_4_2019</i>	<i>Integer</i>	Menunjukkan piksel tetangga 4 pada tahun 2019.
25	<i>Neighbor_5_2019</i>	<i>Integer</i>	Menunjukkan piksel tetangga 5 pada tahun 2019.
26	<i>Neighbor_6_2019</i>	<i>Integer</i>	Menunjukkan piksel tetangga 6 pada tahun 2019.
27	<i>Neighbor_7_2019</i>	<i>Integer</i>	Menunjukkan piksel tetangga 7 pada tahun 2019.
28	<i>Neighbor_8_2019</i>	<i>Integer</i>	Menunjukkan piksel tetangga 8 pada tahun 2019.
29	<i>Neighbor_1_2020</i>	<i>Integer</i>	Menunjukkan piksel tetangga 1 pada tahun 2020.
30	<i>Neighbor_2_2020</i>	<i>Integer</i>	Menunjukkan piksel tetangga 2 pada tahun 2020.
31	<i>Neighbor_3_2020</i>	<i>Integer</i>	Menunjukkan piksel tetangga 3 pada tahun 2020.
32	<i>Neighbor_4_2020</i>	<i>Integer</i>	Menunjukkan piksel tetangga 4 pada tahun 2020.

No	Nama Kolom	Tipe Data	Deskripsi
33	<i>Neighbor_5_2020</i>	<i>Integer</i>	Menunjukkan piksel tetangga 5 pada tahun 2020.
34	<i>Neighbor_6_2020</i>	<i>Integer</i>	Menunjukkan piksel tetangga 6 pada tahun 2020.
35	<i>Neighbor_7_2020</i>	<i>Integer</i>	Menunjukkan piksel tetangga 7 pada tahun 2020.
36	<i>Neighbor_8_2020</i>	<i>Integer</i>	Menunjukkan piksel tetangga 8 pada tahun 2020.
37	<i>Neighbor_1_2021</i>	<i>Integer</i>	Menunjukkan piksel tetangga 1 pada tahun 2021.
38	<i>Neighbor_2_2021</i>	<i>Integer</i>	Menunjukkan piksel tetangga 2 pada tahun 2021.
39	<i>Neighbor_3_2021</i>	<i>Integer</i>	Menunjukkan piksel tetangga 3 pada tahun 2021.
40	<i>Neighbor_4_2021</i>	<i>Integer</i>	Menunjukkan piksel tetangga 4 pada tahun 2021.
41	<i>Neighbor_5_2021</i>	<i>Integer</i>	Menunjukkan piksel tetangga 5 pada tahun 2021.
42	<i>Neighbor_6_2021</i>	<i>Integer</i>	Menunjukkan piksel tetangga 6 pada tahun 2021.
43	<i>Neighbor_7_2021</i>	<i>Integer</i>	Menunjukkan piksel tetangga 7 pada tahun 2021.
44	<i>Neighbor_8_2021</i>	<i>Integer</i>	Menunjukkan piksel tetangga 8 pada tahun 2021.
45	<i>Neighbor_1_2022</i>	<i>Integer</i>	Menunjukkan piksel tetangga 1 pada tahun 2022.
46	<i>Neighbor_2_2022</i>	<i>Integer</i>	Menunjukkan piksel tetangga 2 pada tahun 2022.
47	<i>Neighbor_3_2022</i>	<i>Integer</i>	Menunjukkan piksel tetangga 3 pada tahun 2022.
48	<i>Neighbor_4_2022</i>	<i>Integer</i>	Menunjukkan piksel tetangga 4 pada tahun 2022.
49	<i>Neighbor_5_2022</i>	<i>Integer</i>	Menunjukkan piksel tetangga 5 pada tahun 2022.
50	<i>Neighbor_6_2022</i>	<i>Integer</i>	Menunjukkan piksel tetangga 6 pada tahun 2022.
51	<i>Neighbor_7_2022</i>	<i>Integer</i>	Menunjukkan piksel tetangga 7 pada tahun 2022.
52	<i>Neighbor_8_2022</i>	<i>Integer</i>	Menunjukkan piksel tetangga 8 pada tahun 2022.
53	<i>Area</i>	<i>Object</i>	Menunjukkan lokasi titik pantai.

Berdasarkan informasi yang telah disajikan, pada dataframe terdapat 53 fitur yang masing-masing memiliki 55.696 baris mewakili masing-masing piksel di setiap titik pantai. Fitur x dan y merupakan titik koordinat dimana piksel utama berada. Fitur yang

berupa *Year* seperti *Year\_2018*, *Year\_2019*, dan seterusnya, merupakan fitur yang menunjukkan kelas piksel utama pada tahun tersebut di titik (x,y). Hasil dari *feature engineering* menunjukkan terdapat empat fitur perubahan piksel yang ditunjukkan pada fitur *Change* serta delapan fitur piksel tetangga dari tahun 2018 sampai tahun 2022 sehingga total terdapat 40 fitur piksel tetangga yang ditunjukkan dengan fitur *Neighbor*. Fitur perubahan piksel dan piksel tetangga untuk tahun 2023 tidak diekstraksi karena 2023 menjadi label baik dalam data latih maupun data uji. Adapun bentuk *dataset* hasil penggabungan data yang selanjutnya akan digunakan untuk melakukan prediksi perubahan hutan *mangrove* disajikan pada Tabel IV-5

Tabel IV-5 *Sample Dataset* untuk Model Prediksi

x	y	Year_2018	Year_2019	...	Neighbor_1_2018	Neighbor_2_2018	...	Neighbor_1_2019	Neighbor_2_2019	...	Change_2018_2019	Change_2019_2020	...	Area
0	0	1	1	...	1	1	...	4	4	...	0	0	...	Bajul_Mati
1	0	1	1	...	1	1	...	4	4	...	0	0	...	Bajul_Mati
2	0	4	4	...	4	4	...	4	4	...	0	0	...	Bajul_Mati

#### IV.4 *Data Mining*

Tahap data mining merupakan inti dalam kerangka kerja *Knowledge Discovery in Databases* (KDD) yang bertujuan untuk mengekstraksi pola dan pengetahuan yang bermakna dari data yang telah diproses. Setelah data melewati tahap transformasi, data mining menjadi tahap utama dalam membangun model yang mampu menganalisis dan memprediksi perubahan penutup lahan *mangrove*. Pada penelitian ini, tahapan *data mining* meliputi data *splitting* untuk membagi data menjadi set pelatihan dan pengujian, *resampling* untuk mengatasi ketidakseimbangan data dan meningkatkan performa model, serta *modeling* yang mencakup pembangunan model *Random Forest*.

#### IV.4.1 Data Splitting

Tahap data splitting atau pembagian data dilakukan untuk memisahkan data menjadi dua set, yaitu data latih dan data uji. Proses ini bertujuan untuk memastikan bahwa model dapat dilatih menggunakan sebagian besar data yang representatif, sementara sisanya digunakan untuk menguji kinerja model pada data yang belum pernah dilihat sebelumnya. Dalam penelitian ini, pembagian data dilakukan dengan menggunakan tiga skenario pembagian berdasarkan jumlah titik pantai. Skenario pertama terbagi atas 12 titik pantai sebagai data latih dengan 4 titik pantai sebagai data uji, skenario kedua terbagi atas 13 titik pantai sebagai data latih dengan 3 titik pantai sebagai data uji, dan skenario ketiga terbagi atas 14 titik pantai sebagai data latih dengan 3 titik pantai sebagai data uji. Setiap skenario akan digunakan untuk mengevaluasi performa model guna menentukan konfigurasi pembagian data terbaik.

Tabel IV-6 Skenario Pembagian Data

Pembagian Data	Titik Pantai Data Latih	Titik Pantai Data Uji
12 area latih dan 4 area uji	<ol style="list-style-type: none"><li>1. Pantai Balekambang</li><li>2. Pantai Bantol</li><li>3. Pantai Goa Cina</li><li>4. Pantai Jembatan Panjang</li><li>5. Pantai Jonggring Saloko</li><li>6. Pantai Kondang Iwak</li><li>7. Pantai Kondang Merak</li><li>8. Pantai Licin</li><li>9. Pantai Modangan</li><li>10. Pantai Ngliyep</li><li>11. Pantai Wonogoro</li><li>12. Pantai Sendang Biru</li></ol>	<ol style="list-style-type: none"><li>1. Pantai Sipelot</li><li>2. Pantai Tamban</li><li>3. Pantai Sempu</li><li>4. Pantai Bajul Mati</li></ol>
13 area latih dan 3 area uji	<ol style="list-style-type: none"><li>1. Pantai Bajul Mati</li><li>2. Pantai Balekambang</li><li>3. Pantai Bantol</li><li>4. Pantai Goa Cina</li><li>5. Pantai Jembatan Panjang</li><li>6. Pantai Jonggring Saloko</li><li>7. Pantai Kondang Iwak</li><li>8. Pantai Kondang Merak</li><li>9. Pantai Licin</li><li>10. Pantai Modangan</li><li>11. Pantai Ngliyep</li><li>12. Pantai Wonogoro</li><li>13. Pantai Sendang Biru</li></ol>	<ol style="list-style-type: none"><li>1. Pantai Sipelot</li><li>2. Pantai Tamban</li><li>3. Pantai Sempu</li></ol>

Pembagian Data	Titik Pantai Data Latih	Titik Pantai Data Uji
14 area latih dan 2 area uji	<ol style="list-style-type: none"> <li>1. Pantai Bajul Mati</li> <li>2. Pantai Balekambang</li> <li>3. Pantai Bantol</li> <li>4. Pantai Goa Cina</li> <li>5. Pantai Jembatan Panjang</li> <li>6. Pantai Jonggring Saloko</li> <li>7. Pantai Kondang Iwak</li> <li>8. Pantai Kondang Merak</li> <li>9. Pantai Licin</li> <li>10. Pantai Modangan</li> <li>11. Pantai Ngliyep</li> <li>12. Pantai Wonogoro</li> <li>13. Pantai Sendang Biru</li> <li>14. Pantai Sempu</li> </ol>	<ol style="list-style-type: none"> <li>1. Pantai Sipelot</li> <li>2. Pantai Tamban</li> </ol>

Tabel IV-6 di atas menunjukkan pembagian titik pantai untuk setiap skenario data splitting. Pemilihan titik pantai sebagai data latih dan data uji dilakukan dengan mempertimbangkan lokasi geografis untuk memastikan bahwa model diuji pada wilayah yang berbeda secara spasial. Pendekatan ini bertujuan untuk mengevaluasi kemampuan model dalam memprediksi perubahan tutupan lahan di area yang belum pernah digunakan dalam proses pelatihan, sehingga menghasilkan penilaian yang lebih objektif terhadap performa model dalam menghadapi data baru.

Selain mempertimbangkan aspek geografis, pembagian data juga memperhatikan keberagaman karakteristik tutupan lahan. Beberapa wilayah memiliki cakupan *mangrove* yang dominan, sementara yang lain lebih terbatas atau bahkan tidak memilikinya. Data uji juga dipilih dengan memperhatikan ketersediaan seluruh kelas tutupan lahan, seperti hutan, tambak, perairan, dan pemukiman, meskipun beberapa titik pantai mungkin tidak memiliki kelas tertentu. Variasi ini memastikan bahwa model diuji pada pola perubahan yang beragam, sehingga evaluasi dapat dilakukan secara menyeluruh untuk mengukur kemampuannya dalam mengenali dan memprediksi perubahan lahan di berbagai kondisi lingkungan.

Setelah menentukan skenario pembagian data berdasarkan aspek geografis dan karakteristik tutupan lahan, langkah selanjutnya adalah menyesuaikan data untuk keperluan pemodelan. Dalam penelitian ini, label yang diprediksi adalah data tutupan

lahan pada tahun 2023, sementara fitur independen mencakup berbagai variabel yang menggambarkan kondisi tutupan lahan pada tahun-tahun sebelumnya. Maka dari itu, dari 53 fitur yang ada pada dataset, empat fitur yang tidak relevan tidak diikuti sertakan dalam dataset pelatihan, yaitu fitur *Year\_2023* sebagai label, dan fitur *x*, fitur *y*, dan fitur *Area*, sehingga total fitur yang ada pada dataset pegujian adalah 49.

Tabel IV-7 Pembagian Data

Pembagian Data	Data Latih	Data Uji	Total
12 area latih dan 4 area uji	41.772	13.924	55.696
13 area latih dan 3 area uji	45.253	10.443	55.696
14 area latih dan 2 area uji	48.734	6.962	55.696

Berdasarkan Tabel IV-7, data dibagi menjadi data latih dan data uji dalam tiga skenario pembagian. Setiap skenario memiliki jumlah total data yang sama, tetapi proporsi data latih dan data uji berbeda tergantung pada jumlah pembagian yang digunakan. Selanjutnya, diterapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) untuk menyeimbangkan distribusi kelas dalam data latih sebelum model dilatih.

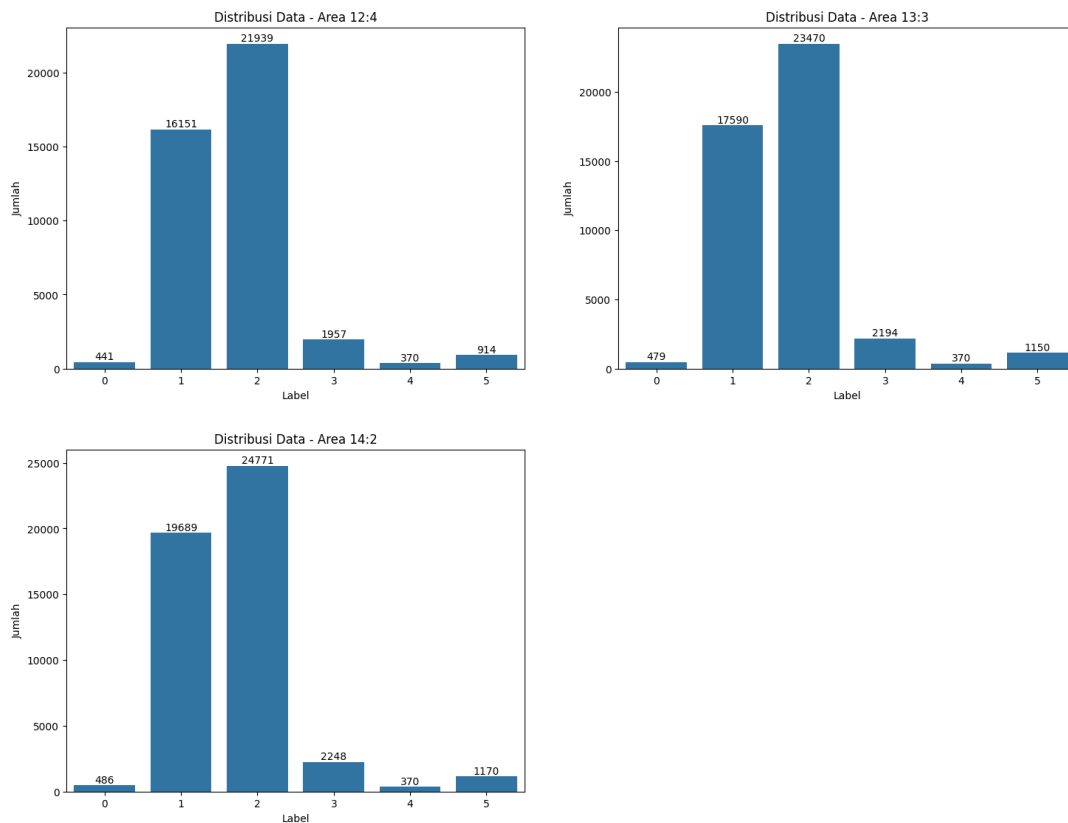
Tabel IV-8 Pembagian Data setelah Teknik SMOTE

Pembagian Data	Data Latih	Data Uji	Total
12 area latih dan 4 area uji	109.695	13.924	123.619
13 area latih dan 3 area uji	117.350	10.443	127.793
14 area latih dan 2 area uji	123.855	6.962	130.817

Tabel IV-8 menunjukkan hasil pembagian data setelah diterapkannya SMOTE pada data latih. Terlihat bahwa jumlah data latih meningkat secara signifikan dibandingkan dengan sebelum penyeimbangan, sedangkan jumlah data uji tetap tidak berubah. Peningkatan ini disebabkan oleh penambahan sampel sintetis dalam kelas minoritas, yang bertujuan untuk menyeimbangkan distribusi kelas dalam data latih. Dengan distribusi yang lebih merata, model diharapkan dapat belajar lebih baik dari seluruh kelas, sehingga meningkatkan akurasi prediksi dalam mengenali pola perubahan tutupan lahan.

#### IV.4.2 Penanganan Data *Imbalance*

Analisis distribusi kelas pada label di data latih untuk setiap skenario pembagian data dilakukan sebagai upaya untuk memastikan bahwa setiap kelas penutup lahan memiliki representasi yang cukup dalam dataset pelatihan. Pemeriksaan ini bertujuan untuk mengidentifikasi apakah distribusi antar kelas seimbang atau terdapat ketimpangan yang dapat memengaruhi kinerja model. Ketidakseimbangan dalam distribusi kelas dapat menyebabkan model lebih sulit mengenali pola pada kelas yang kurang terwakili.



Gambar IV.8 Distribusi Jumlah Data pada Setiap Skenario Pembagian Data

Distribusi kelas label di data latih pada setiap pembagian data telah divisualisasikan pada Gambar IV.8. Hasil distribusi menunjukkan adanya ketidakseimbangan yang signifikan antar kelas, sebagaimana dirinci pada Tabel IV-9.

Tabel IV-9 *Detail* Jumlah Distribusi Data pada Setiap Skenario Pembagian Data

Kelas	Keterangan	Area 12:4	Area 13:3	Area 14:2
0	Penurunan Hutan	441	479	486
1	Hutan <i>Mangrove</i>	16.151	17.590	19.689
2	Perairan	21.939	23.470	24.771
3	Pemukiman	1.957	2.194	2.248
4	Tambak	370	370	370
5	Lainnya (Noise)	914	1.150	1.170

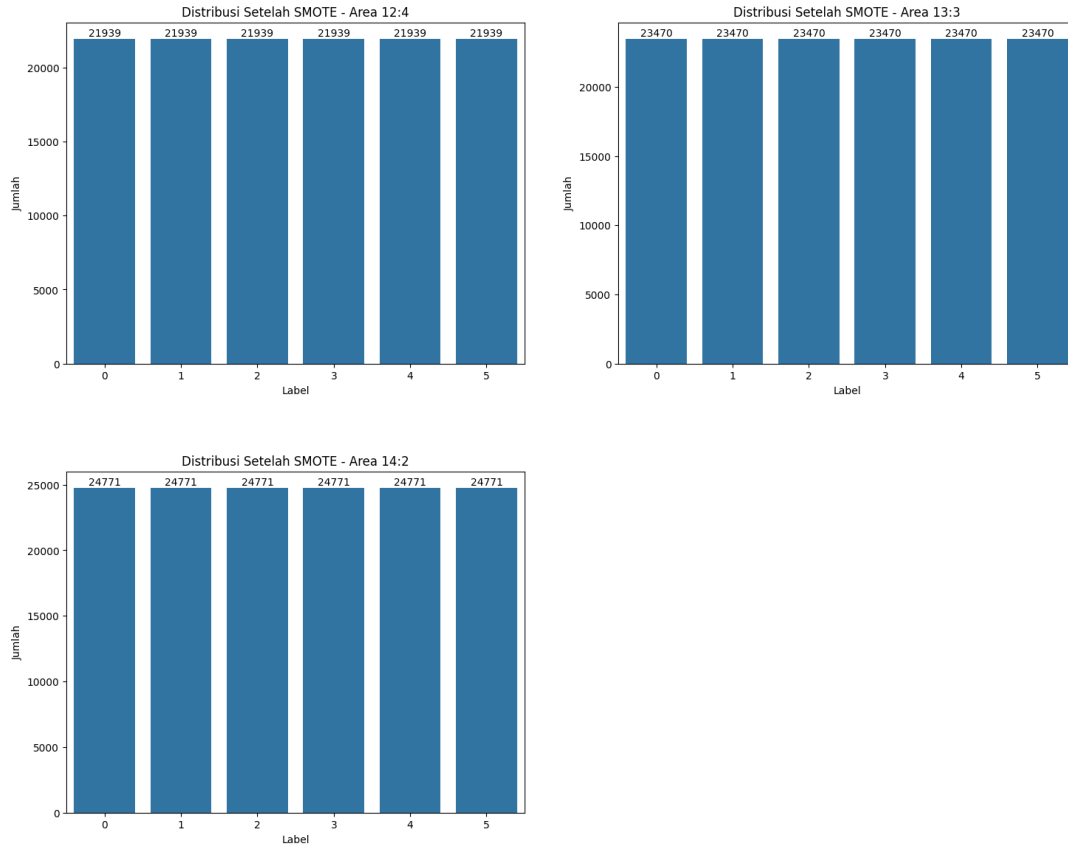
Dari ketiga skenario pembagian data, distribusi kelas menunjukkan pola yang cenderung tidak merata, dengan dominasi kelas tertentu, seperti perairan (kelas 2) dan hutan *mangrove* (kelas 1), sementara kelas lainnya, seperti tambak (kelas 4) dan penurunan hutan (kelas 0), memiliki jumlah sampel yang jauh lebih sedikit. Maka dari itu, dilakukan penerapan metode *resampling* untuk mengeksplorasi bagaimana penyeimbangan data dapat memengaruhi performa model.

Penyeimbangan distribusi dalam data dilakukan dengan *resampling* menggunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE). Teknik ini diterapkan hanya pada data latih, karena tujuan utamanya adalah memastikan bahwa model memiliki representasi yang lebih seimbang untuk setiap kelas selama proses pelatihan, tanpa mengubah distribusi data uji. Dengan cara ini, model dapat belajar lebih baik dalam mengenali pola pada kelas dengan jumlah sampel yang lebih sedikit.

SMOTE bekerja dengan menambahkan sampel sintetis pada kelas minoritas melalui interpolasi antara sampel yang ada dan tetangga terdekatnya. Dalam penelitian ini, SMOTE diterapkan pada kelas dengan jumlah sampel yang jauh lebih kecil dibandingkan kelas dominan, yaitu penurunan hutan (kelas 0), pemukiman (kelas 3), dan tambak (kelas 4). Sebelum diterapkan SMOTE, ketiga kelas ini memiliki jumlah sampel yang sangat terbatas dibandingkan kelas lainnya, sebagaimana terlihat dalam distribusi awal pada Tabel IV-9. Dengan SMOTE, sampel tambahan dihasilkan berdasarkan pola yang ada dan bukan sekadar duplikasi. Kelas minoritas memperoleh variasi data yang tetap merepresentasikan karakteristik aslinya sehingga dapat



mengurangi bias terhadap kelas mayoritas dan meningkatkan kemampuan model dalam mengenali pola dari seluruh kelas.



Gambar IV.9 Hasil *Resampling* dengan Teknik SMOTE

Berdasarkan Gambar IV.9, penerapan SMOTE pada setiap skenario pembagian data berhasil menyeimbangkan distribusi kelas dalam data latih. Pada area 12:4, jumlah sampel di setiap kelas meningkat menjadi 21.939 sampel, sedangkan pada area 13:3 meningkat menjadi 23.470 sampel, dan pada area 14:2 mencapai 24.771 sampel. Peningkatan ini menunjukkan bahwa kelas dengan jumlah sampel yang sebelumnya lebih sedikit kini memiliki representasi yang lebih seimbang dengan kelas lainnya.

### IV.4.3 *Modeling*

#### IV.4.3.1 Pengujian Kombinasi Parameter

Pada penelitian ini, pengujian kombinasi parameter pada model *Random Forest* dilakukan untuk mengoptimalkan kinerja model dalam memprediksi perubahan

tutupan lahan. Pengujian ini dilakukan dengan menggunakan *framework optuna*, yang memungkinkan eksplorasi nilai optimal dari setiap parameter secara otomatis melalui pendekatan *bayesian optimization*. Dalam penelitian ini, rentang nilai parameter utama telah ditentukan berdasarkan pengaturan yang umum digunakan dalam literatur dan praktik pembelajaran mesin. Tabel IV-10 menunjukkan daftar parameter beserta rentang nilai yang digunakan selama eksperimen.

Tabel IV-10 Parameter yang Diuji

Parameter	Nilai
<i>n_estimators</i>	100, 200, 300
<i>max_depth</i>	5, 10
<i>min_samples_split</i>	10,20
<i>min_samples_leaf</i>	6,8,10
<i>criterion</i>	gini, entropy

Parameter yang dipilih memiliki fungsi spesifik yang berperan penting dalam proses pelatihan model. Parameter *n\_estimators* menentukan jumlah pohon dalam *ensemble*, dengan lebih banyak pohon memungkinkan model menangkap pola data yang kompleks, tetapi meningkatkan waktu komputasi. Parameter *max\_depth* mengontrol kedalaman maksimum pohon, membatasi kompleksitas untuk menghindari *overfitting*. Parameter *min\_samples\_split* dan *min\_samples\_leaf* masing-masing memastikan bahwa pemisahan simpul dan daun pohon memiliki cukup data untuk menghasilkan keputusan yang kuat, sehingga meningkatkan generalisasi model. Parameter *criterion* digunakan untuk mengukur kualitas pemisahan pada simpul, dengan opsi gini dan entropy, yang dipilih karena mampu memberikan evaluasi pemisahan yang efektif pada data klasifikasi. Rentang nilai untuk setiap parameter dirancang berdasarkan eksperimen awal untuk mencapai keseimbangan antara akurasi prediksi dan efisiensi komputasi.

Tabel IV-11 Hasil Kombinasi Parameter dari *Framework Optuna*

Parameter	Nilai
<i>n_estimators</i>	100

Parameter	Nilai
<i>max_depth</i>	10
<i>min_samples_split</i>	20
<i>min_samples_leaf</i>	10
<i>criterion</i>	gini

Tabel IV-11 menunjukkan hasil pencarian kombinasi parameter terbaik menggunakan *framework optuna*. Parameter pertama, *n\_estimators* bernilai 100, menunjukkan 100 jumlah pohon dalam model. Parameter *min\_samples\_split*, bernilai 20, yang berarti setiap *node* dalam *decision tree* harus memiliki minimal 20 sampel sebelum dipisah. Parameter *min\_samples\_leaf* bernilai 10, menentukan bahwa setiap *leaf node* minimal memiliki 10 sampel untuk menghindari *overfitting*. Selanjutnya, *max\_depth* optimal pada 10, yang membatasi kedalaman pohon agar model tetap generalisasi. Terakhir, *criterion* terbaik adalah gini, yang digunakan untuk mengukur impuritas dalam node guna menentukan pemisahan terbaik.

#### IV.4.3.1 Desain Model *Random Forest*

##### 1. Pengambilan Sampel dengan Pengembalian (*Bootstrapping*)

Bootstrapping adalah metode pengambilan sampel secara acak dengan pengembalian yang digunakan dalam *Random Forest* untuk membangun *dataset* pelatihan bagi setiap *decision tree* (Salman dkk., 2024). Artinya, setiap pohon dalam *Random Forest* tidak menggunakan seluruh dataset, melainkan menggunakan *subset* yang diambil secara acak dengan pengembalian, sehingga satu sampel bisa muncul lebih dari satu kali dalam subset. Sekitar 36,8% dari total data asli tidak akan masuk ke dataset bootstrap dan dapat digunakan sebagai *Out-of-Bag* (OOB) samples untuk validasi.

Tabel IV-12 Sampel *Boostraping Dataset*

<i>Year_2018</i>	<i>Year_2019</i>	...	<i>Neighbor_1_2018</i>	<i>Neighbor_2_2018</i>	...	<i>Neighbor_1_2019</i>	<i>Neighbor_2_2019</i>	...	<i>Change_2018_2019</i>	<i>Change_2019_2020</i>	...	<i>Year_20213 (label)</i>
1	1	...	1	1	...	4	4	...	0	0	...	1
1	1	...	1	1	...	4	4	...	0	0	...	1
4	4	...	4	4	...	4	4	...	0	0	...	4
2	1	...	3	3	...	3	2	...	1	0	...	0
3	3	...	4	4	...	1	1	...	0	1	...	1

Tabel IV-12 di atas menunjukkan gambaran dari dilakukannya *bootstraping* pada dataset dengan pembagian data area 12:4 sehingga data latih memiliki jumlah data 109.695. Terlihat bahwa beberapa data muncul lebih dari sekali karena pengambilan sampel dilakukan dengan pengembalian. Dari seluruh data latih tersebut hanya 63,2% saja yang dipilih dalam *bootstrapping*, yaitu sekitar 69.328 data.

## 2. Pemilihan Subset Acak dari Atribut

Selanjutnya adalah dilakukan pemilihan *subset* acak dari atribut pada dataset pelatihan. *Random Forest* tidak mengevaluasi semua atribut dalam dataset, tetapi memilih *subset* acak dari atribut untuk setiap node dalam *decision tree*. Jika dataset memiliki  $M$  atribut, maka hanya  $\sqrt{M}$  atribut yang dipilih secara acak di setiap node. Dalam penelitian ini, terdapat 49 fitur atau atribut pada dataset pelatihan, maka  $\sqrt{49} \approx 7$  atribut akan dipilih secara acak untuk dihitung impurity-nya. Pada contoh ini, 7 atribut acak yang terpilih adalah *Neighbor\_1\_2022*, *Change\_2022\_2021*, *Year\_2021*, *Neighbor\_3\_2021*, *Change\_2019\_2018*, *Neighbor\_5\_2022*, dan *Year\_2020*.

## 3. Pemilihan Atribut *Split*

Berdasarkan hasil pengujian kombinasi parameter, di dapatkan kombinasi dengan *gini impurity* sebagai pemilihan atribut mendapat hasil yang paling optimal. *gini*

*impurity* adalah ukuran seberapa tidak murni suatu node dalam *decision tree*. Jika semua sampel dalam node berasal dari satu kelas, maka nilai *impurity*nya adalah 0 dan dianggap sebagai node murni. Jika semua kelas terdistribusi secara merata, *impurity* akan lebih tinggi. Setelah subset atribut dipilih, *gini Impurity* dihitung untuk masing-masing atribut tersebut. Atribut yang menghasilkan pengurangan *impurity* terbesar akan dipilih sebagai atribut terbaik untuk *split*. Rumus *gini impurity* ditunjukkan pada (IV-1) berikut:

$$Gini(S) = 1 - \sum_{i=1}^c p_i^2 \quad (IV-1)$$

Dimana:

- $C$  merupakan jumlah kelas dalam dataset
- $p_i$  merupakan proporsi sampel yang termasuk dalam kelas ke-  $i$  dalam suatu node.

Tabel IV-13 Jumlah Sampel Hasil *Bootstrapping*

Kelas	Label Kelas	Jumlah Sampel	Proporsi ( $p_i$ )
0	Penurunan Hutan	10.399	0.15
1	Hutan <i>Mangrove</i>	24.264	0.35
2	Perairan	13.866	0.20
3	Pemukiman	6.933	0.10
4	Tambak	10.399	0.15
5	<i>Noise</i>	3.467	0.05

Tabel IV-13 menunjukkan contoh 69.328 sampel hasil *bootstrapping* yang terbagi ke dalam 6 kelas. Selanjutnya, akan dihitung *gini impurity* sebelum *split* dengan rumus (IV-1).

$$Gini(S) = 1 - \sum_{i=1}^6 p_i^2$$

$$Gini(S) = 1 - (0.15^2 + 0.35^2 + 0.20^2 + 0.10^2 + 0.15^2 + 0.05^2)$$

$$Gini(S) = 1 - (0.0225 + 0.1225 + 0.04 + 0.01 + 0.0225 + 0.0025)$$

$$Gini(S) = 1 - 0.22 = 0.78$$

Hasil *gini impurity* sebelum split mendapatkan nilai 0.78 yang berarti node ini masih sangat tidak murni. Maka dari itu, perlu dilakukan perhitungan *split* berdasarkan atribut yang sebelumnya sudah dipilih secara acak. Setelah melakukan perhitungan *gini split*, langkah selanjutnya adalah menentukan atribut mana yang paling optimal untuk digunakan sebagai *node* pembagian dalam *decision tree*. Pemilihan akhir didasarkan pada *gini gain*, yang mengukur seberapa besar pengurangan impuritas setelah dilakukan pemisahan. Semakin besar nilai *gini gain*, semakin baik atribut tersebut dalam memisahkan kelas. Rumus *gini gain* dijelaskan pada (IV-2) berikut.

$$Gini\ Gain = Gini_{awal} - Gini_{split} \quad (IV-2)$$

Dimana:

- *Gini<sub>awal</sub>* merupakan nilai *gini* sebelum *split*.
- *Gini<sub>split</sub>* merupakan nilai *gini* setelah melakukan *split* dengan atribut tertentu.

Tabel IV-14 Hasil *Gini Split* dan *Gini Gain* pada Atribut Acak

Atribut	<i>Gini Split</i>	<i>Gini Gain</i>
<i>Neighbor_1_2022</i>	0.771	0.009
<i>Year_2021</i>	<b>0.693</b>	<b>0.087</b>
<i>Change_2022_2021</i>	0.720	0.06

Tabel IV-14 menunjukkan hasil dari *gini split* dan *gini gain* pada beberapa atribut acak yang terpilih. Hasil tersebut menunjukkan bahwa atribut *Year\_2021* memiliki *gini split* terkecil yaitu 0.693 serta memiliki *gini gain* terbesar yaitu 0.087 sehingga atribut ini dipilih sebagai atribut pertama untuk *split* dalam *decision tree*.

#### 4. *Majority Voting*

Setelah proses pembentukan *decision tree* dengan pemilihan atribut berdasarkan nilai *gini gain*, langkah selanjutnya dalam algoritma *Random Forest* adalah melakukan *majority voting*. *Majority voting* digunakan untuk menggabungkan

prediksi dari masing-masing *decision tree* yang terbentuk dalam *ensemble model*. Setiap pohon dalam *Random Forest* memberikan hasil prediksi independen untuk setiap sampel, dan keputusan akhir diambil berdasarkan kelas yang paling banyak dipilih oleh pohon-pohon tersebut. *Majority voting* dalam *Random Forest* dirumuskan pada (IV-3).

$$\hat{y} = \arg \max_k \sum_{i=1}^N 1(T_i = k) \quad (IV-3)$$

Dimana:

- $\hat{y}$  adalah label akhir yang diprediksi.
- $k$  adalah setiap kelas yang ada dalam dataset.
- $N$  adalah jumlah *decision tree* dalam model.
- $T_i$  adalah prediksi dari *decision tree* ke- $i$ .
- $1(T_i = k)$  adalah fungsi indikator yang bernilai 1 jika prediksi pohon ke- $i$  adalah  $k$ , dan 0 jika bukan.

Berdasarkan hasil pengujian kombinasi parameter, didapatkan bahwa parameter  $n\_estimator$  terbaik memiliki nilai 100 sehingga dalam model akan terdapat 100 *decision tree*. Tabel IV-15 berikut merupakan contoh hasil jumlah *tree* yang memprediksi pada masing-masing kelas.

Tabel IV-15 Jumlah *Tree* yang Memprediksi Kelas

Kelas	Label Kelas	Jumlah <i>Tree</i> yang Memprediksi
0	Penurunan Hutan	25 <i>tree</i>
1	Hutan <i>Mangrove</i>	30 <i>tree</i>
2	Perairan	10 <i>tree</i>
3	Pemukiman	20 <i>tree</i>
4	Tambak	10 <i>tree</i>
5	<i>Noise</i>	5 <i>tree</i>

Hasil pada Tabel IV-15 selanjutnya dimasukkan ke dalam rumus *majority voting* seperti berikut:

$$\hat{y} = \arg \max_{k \in \{0,1,2,3,4,5\}} \sum_{i=1}^N 1(T_i = k)$$

$$\hat{y} = \arg \max_{k \in \{0,1,2,3,4,5\}} (25, 30, 10, 20, 10, 5) = \text{kelas 1 (hutan mangrove)}$$

Berdasarkan hasil perhitungan *majority voting*, model menentukan bahwa kelas dengan jumlah prediksi terbanyak adalah hutan *mangrove* (kelas 1) dengan 30 *tree*, sehingga sampel tersebut diklasifikasikan sebagai hutan *mangrove*.

#### IV.5 Evaluation

Dalam penelitian ini, model dikembangkan untuk mengklasifikasikan data ke dalam lima kelas tutupan lahan, sehingga evaluasi menggunakan *confusion matrix* multikelas dengan dimensi 6×6 seperti yang disajikan pada Gambar IV.10. *Confusion matrix* ini merepresentasikan hubungan antara prediksi model dan label sebenarnya dalam enam kelas, dengan setiap baris menunjukkan jumlah sampel dari suatu kelas yang diprediksi sebagai kelas lain. Matriks ini membantu mengidentifikasi sejauh mana model mampu mengenali kelas dengan benar serta mengetahui jenis kesalahan yang terjadi dalam klasifikasi.

		Predicted					
		0	1	2	3	4	5
Actual	0	TN	FP	TN	TN	TN	TN
	1	FN	TP	FN	FN	FN	FN
	2	TN	FP	TN	TN	TN	TN
	3	TN	FP	TN	TN	TN	TN
	4	TN	FP	TN	TN	TN	TN
	5	TN	FP	TN	TN	TN	TN

Gambar IV.10 *Confusion matrix* 6×6



Pada *confusion matrix*, setiap nilai di diagonal utama menunjukkan jumlah prediksi yang benar atau *True Positive* (TP) untuk masing-masing kelas, sementara nilai di luar diagonal menunjukkan kesalahan prediksi. Kesalahan ini dapat dibagi menjadi dua jenis yaitu *False Positive* (FP), ketika model memprediksi suatu kelas tetapi sebenarnya sampel berasal dari kelas lain, dan *False Negative* (FN), ketika model gagal mengklasifikasikan sampel ke dalam kelas yang benar. Sebagai contoh, jika baris kedua memiliki angka yang tinggi di kolom kedua, maka model memiliki performa yang baik dalam mengklasifikasikan kelas tersebut. Sebaliknya, jika terdapat angka tinggi di luar diagonal utama, hal itu menunjukkan adanya kesalahan prediksi yang signifikan. *True Negative* (TN) dalam kasus multikelas dihitung dengan cara yang berbeda dibandingkan dengan klasifikasi biner, yaitu dengan menjumlahkan seluruh nilai dalam *confusion matrix* kecuali baris dan kolom yang sedang dievaluasi. Misalnya, jika mengevaluasi kelas tertentu, TN mencakup semua sampel yang tidak termasuk dalam kelas tersebut dan juga tidak diprediksi sebagai kelas tersebut. Dengan memahami pola dari *confusion matrix* ini, dapat dilakukan analisis lebih lanjut untuk menilai performa model secara lebih mendalam, termasuk melalui perhitungan *accuracy*, *precision*, *recall*, dan *f1-score*.

## BAB V VALIDASI, ANALISIS, HASIL, DAN IMPLIKASI

### V.1 Pengujian Model

Pengujian model dilakukan setelah proses penyeimbangan data menggunakan teknik SMOTE untuk mengevaluasi kemampuannya dalam memprediksi perubahan tutupan lahan. Data dibagi ke dalam tiga skenario pembagian yang sudah ditentukan sebelumnya. Pada skenario 12 area latih dan 4 area uji, jumlah data latih yang digunakan sebanyak 109.695, sedangkan data uji berjumlah 13.924. Skenario 13 area latih dan 3 area uji menggunakan 117.350 data latih dan 10.443 data uji. Sementara itu, skenario 14 area latih dan 2 area uji terdiri dari 123.855 data latih dengan 6.962 data uji.

Tabel V-1 Hasil Pengujian Model

Pembagian Data	Akurasi
Area 12:4	99,10%
Area 13:3	98,87%
Area 14:2	98,49%

Hasil pengujian model pada Tabel V-1 merupakan hasil pengujian data uji secara keseluruhan. Skenario 12 area latih dan 4 area uji menghasilkan akurasi tertinggi sebesar 99,10%, diikuti oleh skenario 13 area latih dan 3 area uji dengan akurasi 98,87%, sedangkan skenario 14 area latih dan 2 area uji memiliki akurasi paling rendah, yaitu 98,49%. Penurunan akurasi seiring dengan meningkatnya jumlah area latih menunjukkan bahwa meskipun lebih banyak data latih tersedia, peningkatan variasi dalam dataset tidak selalu meningkatkan performa model. Hal ini dapat terjadi karena semakin banyak area latih yang digunakan, semakin beragam pola yang dipelajari oleh model, yang berpotensi meningkatkan kompleksitas dan membuat model lebih sulit melakukan generalisasi terhadap data uji yang lebih sedikit. Selain itu, skenario dengan jumlah area uji yang lebih kecil, seperti 14:2, memiliki data uji yang lebih terbatas, sehingga kurang optimal dalam mengevaluasi performa model terhadap kondisi yang lebih luas. Sebaliknya, skenario 12:4 dengan proporsi data uji yang lebih besar

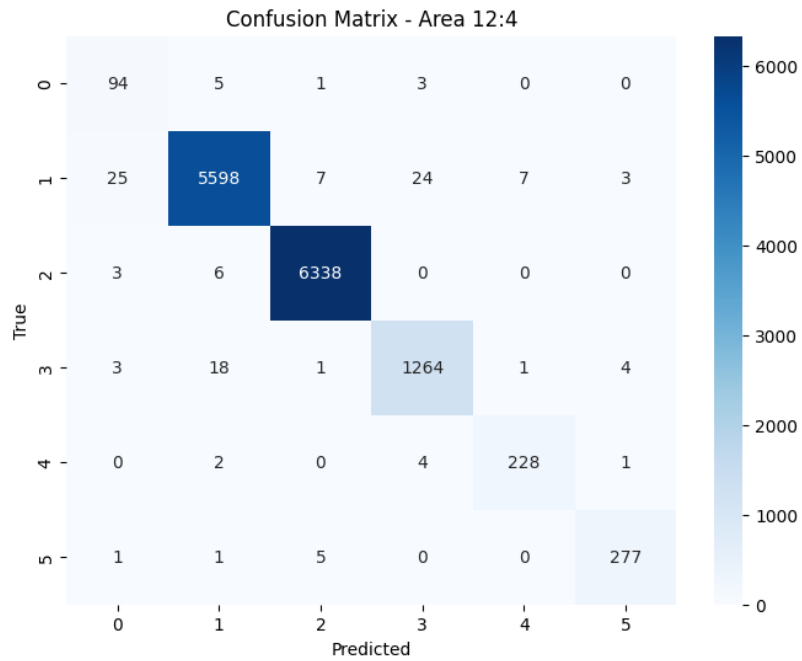
memberikan hasil yang lebih stabil karena memungkinkan evaluasi yang lebih menyeluruh terhadap generalisasi model.

Melihat hasil dari penelitian sebelumnya, algoritma *Random Forest* telah banyak digunakan dalam berbagai kasus prediksi dan klasifikasi data tutupan lahan. Penelitian oleh Bui & Mucsi (2021) melaporkan akurasi keseluruhan 93,86% untuk klasifikasi tutupan lahan berbasis kombinasi pendekatan piksel dan objek menggunakan *Random Forest*. Penelitian lain oleh Dzulfikar dkk. (2024) menunjukkan bahwa akurasi deteksi mangrove menggunakan *Random Forest* mencapai 98,44%, sementara penelitian oleh Faheem dkk. (2024) menunjukkan akurasi sebesar 99% dalam menganalisis dinamika tutupan lahan di ekosistem gurun Pakistan. Hasil penelitian ini, dengan akurasi tertinggi mencapai 99,10% pada skenario 12:4, menunjukkan bahwa model *Random Forest* yang digunakan mampu mempertahankan performa tinggi. Dibandingkan dengan hasil penelitian sebelumnya, akurasi keseluruhan di atas 98% dalam semua skenario menempatkan penelitian ini dalam kategori yang sangat baik dan menghasilkan performa yang kompetitif, bahkan melebihi beberapa hasil penelitian terdahulu.

## **V.2 Evaluasi Hasil Pengujian**

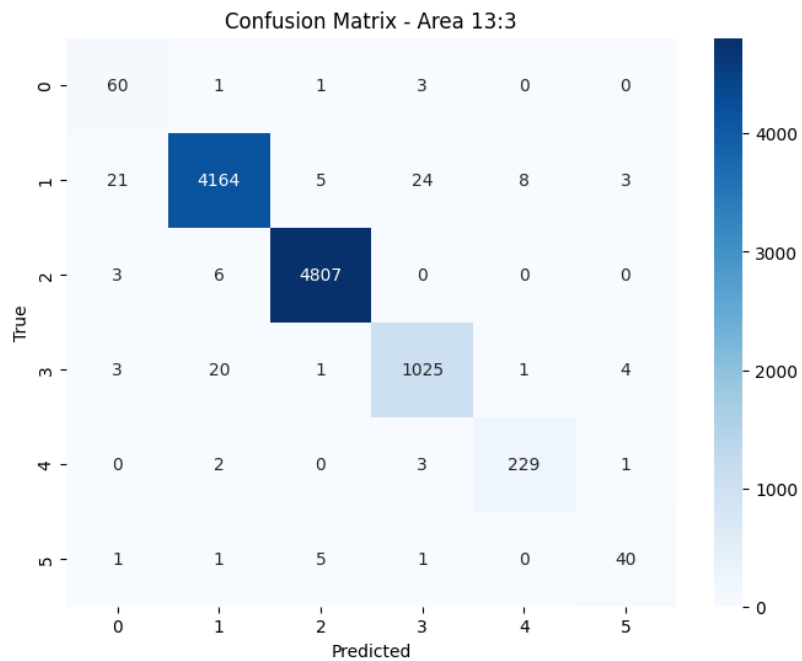
### **V.2.1 *Confusion matrix***

Pada tahap evaluasi hasil pengujian, dilakukan analisis menggunakan *confusion matrix* untuk setiap skenario pembagian data guna memahami performa prediksi model secara lebih mendetail. *Confusion matrix* berukuran 6x6 digunakan untuk membandingkan prediksi model dengan label sebenarnya dalam data uji, sehingga memungkinkan identifikasi kesalahan klasifikasi di setiap kelas tutupan lahan. Dengan adanya *confusion matrix*, dapat diketahui sejauh mana model mampu mengenali setiap kelas dengan baik serta apakah terdapat kelas yang lebih sering mengalami kesalahan prediksi dibandingkan kelas lainnya.



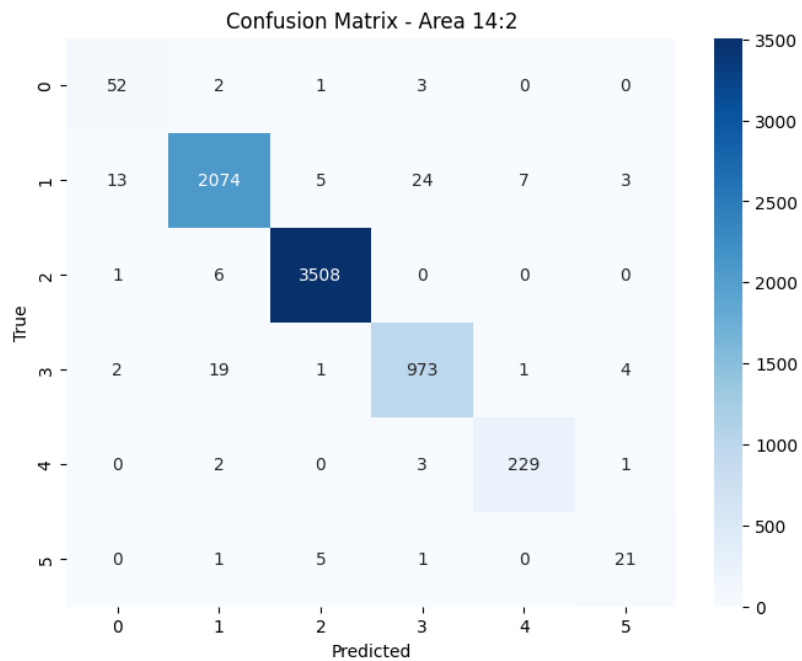
Gambar V.1 *Confusion matrix* untuk 12 Area Latih dan 4 Area Uji

Skenario pembagian data dengan 12 area latih dan 4 area uji pada Gambar V.1 menunjukkan performa klasifikasi yang cukup baik dengan jumlah prediksi yang benar cukup tinggi pada setiap kelas. Kelas 2 (perairan) memiliki prediksi paling akurat dengan 6.338 sampel yang diklasifikasikan dengan benar. Kelas 1 (hutan *mangrove*) juga memiliki akurasi yang tinggi dengan 5.598 prediksi yang benar, meskipun terdapat beberapa kesalahan klasifikasi ke kelas lain. Beberapa kesalahan terlihat pada kelas 3 (pemukiman), di mana 18 sampel salah diklasifikasikan sebagai kelas 1 dan 4 sampel salah diklasifikasikan sebagai kelas 5. Kelas 4 (tambak) memiliki 228 prediksi yang benar, tetapi juga mengalami beberapa kesalahan klasifikasi ke kelas lain.



Gambar V.2 *Confusion matrix* untuk 13 Area Latih dan 3 Area Uji

Selanjutnya pada skenario pembagian data dengan 13 area latih dan 3 area uji yang ditunjukkan pada Gambar V.2, hasil klasifikasi masih menunjukkan pola yang hampir serupa dengan skenario sebelumnya, meskipun terdapat sedikit penurunan jumlah prediksi yang benar pada beberapa kelas. Kelas 2 (perairan) masih menjadi kelas dengan prediksi paling akurat, dengan 4.807 sampel yang diklasifikasikan dengan benar. Kelas 1 (Hutan *Mangrove*) mengalami sedikit penurunan akurasi dibandingkan skenario area 12:4, dengan 4.164 prediksi benar, sementara beberapa sampel salah diklasifikasikan ke kelas. Kelas 3 (pemukiman) memiliki 1.025 prediksi yang benar, tetapi beberapa sampel masih salah diklasifikasikan. Kelas 4 (tambak) mempertahankan jumlah prediksi benar yang cukup stabil dengan 229 prediksi benar, namun beberapa sampel masih diklasifikasikan sebagai kelas 3.



Gambar V.3 *Confusion matrix* untuk 14 Area Latih dan 2 Area Uji

Gambar V.3 menunjukkan hasil skenario pembagian data dengan 14 area latih dan 2 area uji yang terjadi sedikit penurunan akurasi pada hampir semua kelas. Kelas 2 (perairan) tetap menjadi kelas dengan prediksi paling akurat. Namun, akurasi pada kelas 1 (hutan *mangrove*) mengalami penurunan yang cukup signifikan dibandingkan skenario sebelumnya, dengan hanya 2.074 prediksi yang benar. Kelas 3 (pemukiman) memiliki 973 prediksi benar, tetapi juga masih mengalami kesalahan klasifikasi ke kelas lain. Kelas 4 (tambak) tetap memiliki jumlah prediksi benar yang stabil, dengan 229 prediksi yang benar. Secara keseluruhan, hasil ini menunjukkan bahwa meskipun jumlah area latih lebih besar, model tidak selalu menghasilkan klasifikasi yang lebih akurat, kemungkinan karena peningkatan variasi dalam data latih yang menyebabkan model mengalami kesulitan dalam menggeneralisasi pola untuk data uji yang lebih terbatas.

### V.2.2 Perhitungan Metriks Evaluasi

Selanjutnya dilakukan perhitungan metriks kinerja seperti *precision*, *recall*, dan *f1-score* berdasarkan hasil *confusion matrix* yang telah diperoleh. Metriks ini digunakan untuk menilai seberapa baik model dalam mengklasifikasikan setiap kelas dalam

prediksi perubahan tutupan lahan. *Precision* mengukur ketepatan model dalam memberikan prediksi positif untuk setiap kelas, sedangkan *recall* menunjukkan sejauh mana model dapat mengenali semua sampel yang benar dalam suatu kelas. *F1-score* merupakan rata-rata antara *precision* dan *recall*, yang memberikan gambaran keseimbangan antara kedua metrik tersebut. Perhitungan dilakukan pada masing-masing skenario pembagian data untuk mengevaluasi konsistensi performa model dalam berbagai proporsi data latih dan uji.

Tabel V-2 Hasil Perhitungan Metriks Evaluasi

<b>Pembagian Data</b>	<b><i>Precision</i></b>	<b><i>Recall</i></b>	<b><i>F1-score</i></b>
Area 12:4	94%	97%	95%
Area 13:3	91%	95%	92%
Area 14:2	90%	93%	91%

Tabel V-2 menyajikan hasil perhitungan metrik evaluasi secara keseluruhan berdasarkan *macro-average* untuk *precision*, *recall*, dan *f1-score* pada masing-masing skenario pembagian data. Penggunaan *macro-average* dalam perhitungan metrik evaluasi dipilih karena pendekatan ini memberikan gambaran yang lebih seimbang terhadap performa model di seluruh kelas, tanpa dipengaruhi oleh jumlah sampel dalam setiap kelas. Setiap kelas diberi bobot yang sama dalam perhitungan *precision*, *recall*, dan *F1-score*, sehingga model dapat dievaluasi berdasarkan kemampuannya dalam mengenali semua kelas secara setara.

Model menunjukkan performa terbaik pada skenario area 12:4, dengan *precision* sebesar 94%, *recall* 97%, dan *F1-score* 95%, mengindikasikan bahwa model mampu menjaga keseimbangan yang baik antara ketepatan dan sensitivitas dalam mengklasifikasikan perubahan tutupan lahan. Pada skenario area 13:3, terjadi sedikit penurunan dengan *precision* 91%, *recall* 95%, dan *F1-score* 92%, yang menunjukkan bahwa meskipun model masih cukup presisi, kemampuannya dalam mengenali sampel dari setiap kelas sedikit berkurang dibandingkan skenario pertama. Skenario area 14:2 memiliki performa paling rendah dengan *precision* 90%, *recall* 93%, dan *F1-score* 91%, yang kemungkinan disebabkan oleh jumlah data uji yang lebih kecil sehingga

model kurang mendapatkan variasi dalam pengujian. Meskipun demikian, perbedaan nilai antar skenario tidak terlalu signifikan, yang menunjukkan bahwa model tetap memiliki generalisasi yang baik di berbagai proporsi data latih dan uji.

Tabel V-3 Hasil Perhitungan Metriks Evaluasi setiap Kelas

Pembagian Data	Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Area 12:4	0	75%	91%	82%
	1	99%	99%	99%
	2	100%	100%	100%
	3	98%	98%	98%
	4	97%	97%	97%
	5	97%	98%	97%
Area 13:3	0	68%	92%	78%
	1	99%	99%	99%
	2	100%	100%	100%
	3	97%	97%	97%
	4	96%	97%	97%
	5	83%	83%	83%
Area 14:2	0	76%	90%	83%
	1	99%	98%	98%
	2	100%	100%	100%
	3	97%	97%	97%
	4	97%	97%	97%
	5	72%	75%	74%

Tabel V-3 menunjukkan hasil metriks evaluasi yang diterapkan pada masing-masing kelas di tiap pembagian data. Model menunjukkan performa yang sangat baik secara keseluruhan, dengan mayoritas kelas memiliki *precision*, *recall*, dan *f1-score* di atas 95%. Kelas 2 (perairan) secara konsisten mendapatkan skor sempurna di semua skenario, menunjukkan bahwa model mampu mengenali dan mengklasifikasikan kelas ini dengan sangat baik. Namun, pada kelas 0 (penurunan hutan), skor yang dimiliki lebih rendah dibandingkan kelas lainnya, terutama pada skenario area 13:3, di mana kelas 0 memiliki *precision* terendah yaitu 68% meskipun *recall*-nya tetap tinggi sebesar



92%. Hal ini menunjukkan bahwa model lebih sering salah memprediksi kelas lain sebagai kelas 0, yang menyebabkan nilai *precision* menurun. Secara keseluruhan, meskipun terdapat beberapa perbedaan dalam performa antar kelas, model tetap menunjukkan tingkat akurasi yang tinggi dengan distribusi prediksi yang cukup seimbang di berbagai skenario pembagian data.

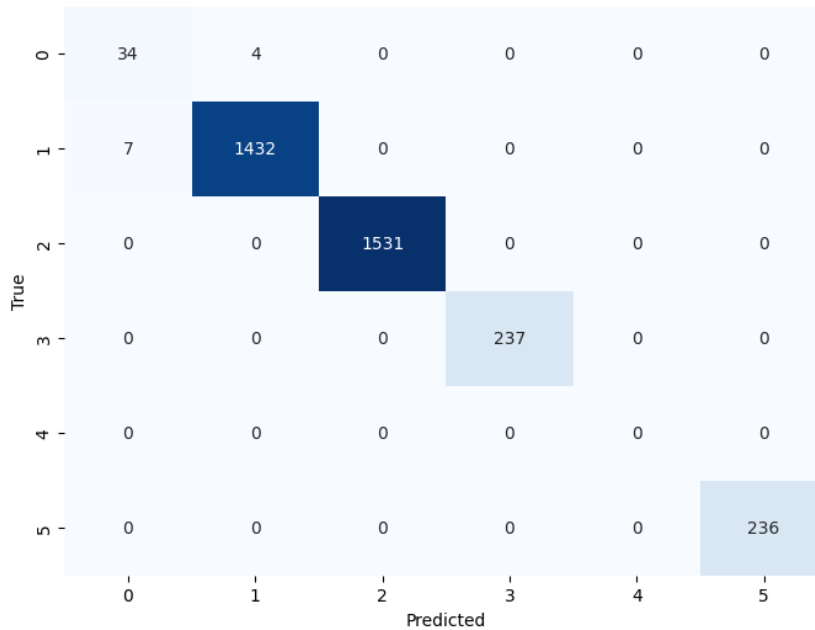
### **V.2.3 Analisis Hasil Pengujian Model Terbaik**

Berdasarkan pengujian model yang telah dilakukan, kombinasi model terbaik diperoleh pada skenario pembagian data area 12:4. Skenario ini menunjukkan performa tertinggi dalam hal akurasi serta keseimbangan antara *precision*, *recall*, dan *f1-score* di hampir semua kelas. Proporsi 12 area latih dan 4 area uji memberikan keseimbangan optimal antara jumlah data pelatihan yang cukup untuk model belajar serta data uji yang memadai untuk mengukur generalisasi model. Pada pembagian area ini, data uji mencakup empat titik pantai yang telah didefinisikan sebelumnya, yaitu Pantai Bajul Mati, Pantai Sipelot, Pantai Tamban, dan Pantai Sempu. Setiap titik pantai memiliki 3.481 data uji, sehingga total data uji yang digunakan dalam skenario ini adalah 13.924.

Analisis hasil prediksi model terbaik dilakukan pada setiap lokasi data uji, mencakup tiga aspek utama yaitu *confusion matrix*, metrik evaluasi yaitu *accuracy*, *precision*, *recall*, dan *f1-score*, serta visualisasi perbandingan antara hasil prediksi model dengan kondisi aktual pada tahun 2023. *Confusion matrix* digunakan untuk memahami distribusi kesalahan klasifikasi model pada setiap kelas, sedangkan metrik evaluasi memberikan gambaran kuantitatif mengenai performa model di setiap lokasi uji. Sementara itu, visualisasi hasil prediksi dibandingkan dengan data aktual memberikan pemahaman lebih lanjut mengenai sejauh mana model dapat merepresentasikan perubahan tutupan lahan secara spasial.

#### **V.2.3.1 Analisis Pantai Bajul Mati**

Pada skenario pengujian terbaik dengan pembagian data area 12:4, dilakukan evaluasi hasil prediksi model untuk Pantai Bajul Mati. Evaluasi ini mencakup analisis *confusion matrix*, perhitungan metrik evaluasi, serta perbandingan visual antara data aktual dan prediksi model.



Gambar V.4 *Confusion Matrix* pada Data Uji Pantai Bajul Mati

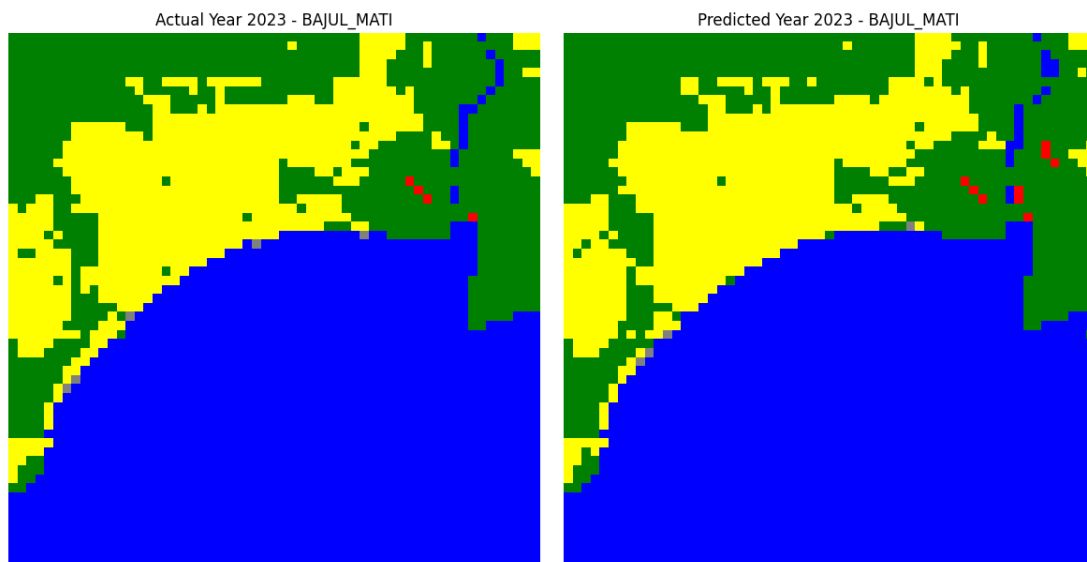
Berdasarkan Gambar V.4, *confusion matrix* untuk Pantai Bajul Mati menunjukkan bahwa model secara umum mampu mengklasifikasikan kelas tutupan lahan dengan baik. Nilai diagonal utama dalam *confusion matrix* mengindikasikan jumlah prediksi yang benar untuk setiap kelas. Kelas dengan jumlah prediksi yang benar paling banyak adalah kelas 1 (hutan *mangrove*) dengan 1.432 sampel yang seluruhnya diklasifikasikan dengan benar, diikuti oleh kelas 2 (perairan) dengan 1.531 sampel yang seluruhnya teridentifikasi secara tepat. Namun, masih terdapat sedikit kesalahan prediksi, terutama pada kelas 0 (perubahan hutan), yang memiliki 4 sampel yang salah diklasifikasikan sebagai kelas 1. Hal ini menunjukkan bahwa model memiliki kecenderungan untuk mengklasifikasikan beberapa perubahan tutupan lahan sebagai kelas yang cukup dominan dalam area tersebut.

Tabel V-4 Metriks Evaluasi pada Data Uji Pantai Bajul Mati

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
0	83%	89%	86%
1	100%	100%	100%
2	100%	100%	100%

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
3	100%	100%	100%
4	0%	0%	0%
5	100%	100%	100%

Hasil perhitungan metrik evaluasi pada Tabel V-4 menunjukkan bahwa kinerja model sangat baik pada sebagian besar kelas, khususnya kelas 1 (hutan *mangrove*) , 2 (perairan), dan 3 (pemukiman) dengan nilai *precision*, *recall*, dan *f1-score* sempurna sebesar 100%. Hal ini menunjukkan bahwa model mampu mengenali dan mengklasifikasikan seluruh data tanpa kesalahan. Pada kelas 0 (penurunan hutan), meskipun hasilnya masih tergolong baik dengan nilai *precision* sebesar 83% dan *recall* 89%, terdapat indikasi beberapa kesalahan prediksi yang menurunkan nilai *precision*. Kondisi ini menunjukkan bahwa model masih sering salah memprediksi kelas lain sebagai kelas 0. Sementara itu, pada kelas 4 (tambak), semua metrik bernilai nol karena tidak terdapat data dengan kelas 4 pada data di Pantai Bajul Mati.

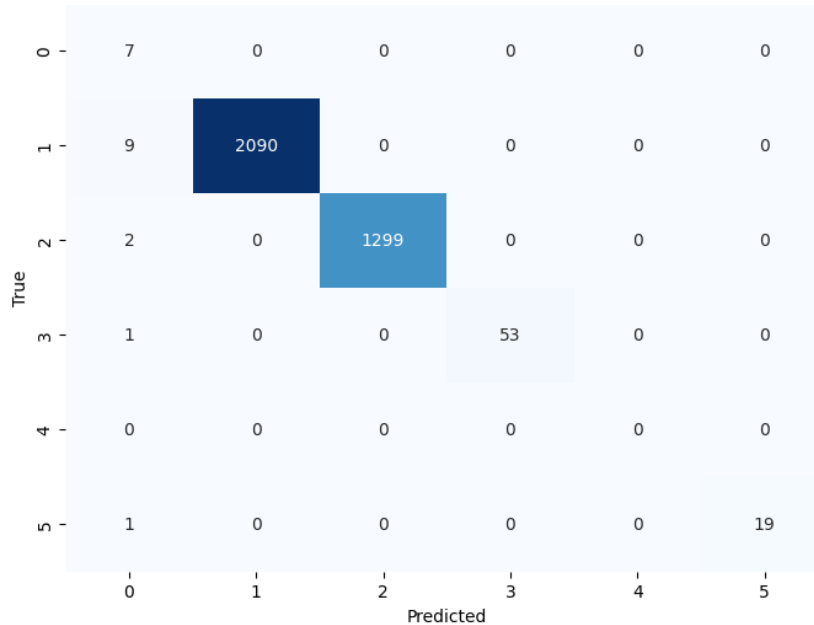


Gambar V.5 Visualisasi Hasil Prediksi Data Uji Pantai Bajul Mati

Selanjutnya, pada Gambar V.5, visualisasi hasil prediksi terhadap data aktual menunjukkan bahwa model berhasil mereplikasi pola tutupan lahan di Pantai Bajul Mati dengan cukup baik. Secara keseluruhan, visualisasi hasil prediksi memiliki

kesamaan tinggi dengan peta aktualnya dengan distribusi warna yang menunjukkan kesesuaian antara prediksi model dengan kondisi sebenarnya pada tahun 2023. Namun, masih terdapat perbedaan yang dengan mudah terdeteksi yaitu pada kelas penurunan hutan.

### V.2.3.1 Analisis Pantai Sempu



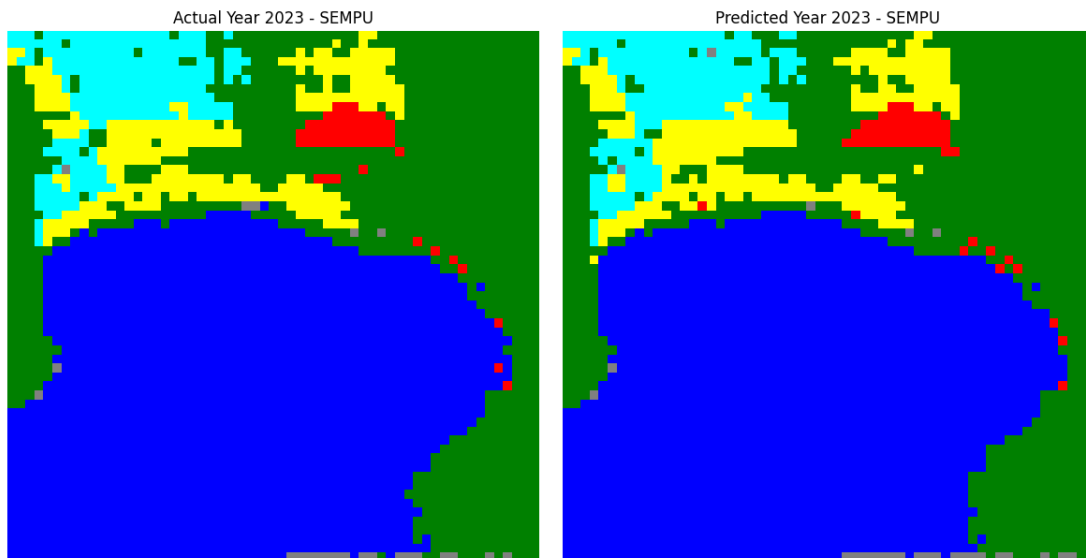
Gambar V.6 *Confusion Matrix* pada Data Uji Pantai Sempu

Berdasarkan Gambar V.6, *confusion matrix* untuk Pantai Sempu menunjukkan bahwa model secara umum masih mampu mengklasifikasikan sebagian besar kelas dengan baik, tetapi terdapat lebih banyak kesalahan klasifikasi dibandingkan Bajul Mati.. Hasil pada *confusion matrix* menunjukkan hasil yang sangat baik pada hampir seluruh kelas dalam melakukan prediksi. 2.090 sampel pada kelas 1 (hutan *mangrove*), 1.299 sampel pada kelas 2 (perairan), dan 53 sampel pada kelas 3 (pemukiman) berhasil diprediksi benar seluruhnya. Satu-satunya kelas yang melakukan kesalahan prediksi menjadi kelas lainnya hanya kelas 0 (penurunan hutan) dengan total 13 sampel.

Tabel V-5 Metriks Evaluasi pada Data Uji Pantai Sempu

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
0	35%	100%	52%
1	100%	100%	100%
2	100%	100%	100%
3	100%	98%	99%
4	0%	0%	0%
5	100%	95%	97%

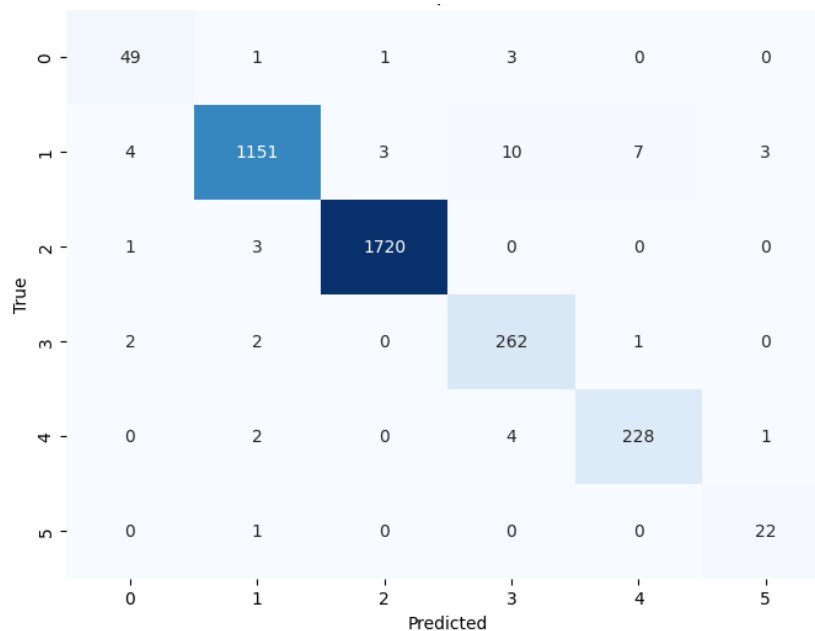
Pada Tabel V-5, metriks evaluasi menunjukkan bahwa model menunjukkan kinerja yang sangat baik pada sebagian besar kelas. Kelas 1 (hutan *mangrove*) dan kelas 2 (perairan) memiliki nilai *precision*, *recall*, dan *f1-score* sempurna sebesar 100%. Pada kelas 3 (pemukiman) menunjukkan performa yang hampir sempurna dengan *recall* 98%, *precision* 100%, dan *f-score* 99%. Namun, pada kelas 0 (penurunan hutan), model mengalami kesulitan dengan *precision* hanya 35%. Hal ini menandakan bahwa terdapat banyak *false positives* meskipun recall 100% menunjukkan bahwa semua kasus positif berhasil terdeteksi. Untuk kelas 4 (tambak), semua metrik bernilai 0% karena tidak terdapat data kelas 4 di Pantai ini.



Gambar V.7 Visualisasi Hasil Prediksi Data Uji Pantai Sempu

Pada Gambar V.7, hasil visualisasi prediksi menunjukkan pola yang cukup mendekati data aktual, tetapi masih terdapat beberapa perbedaan yang cukup signifikan, terutama di bagian yang menunjukkan penurunan hutan. Visualisasi ini memiliki kasus yang serupa pada Pantai Bajul Mati, dimana misklasifikasi pada kelas penurunan hutan terlihat cukup jelas saat dilakukan perbandingan hasil data prediksi dengan aktualnya.

### V.2.3.1 Analisis Pantai Sipelot



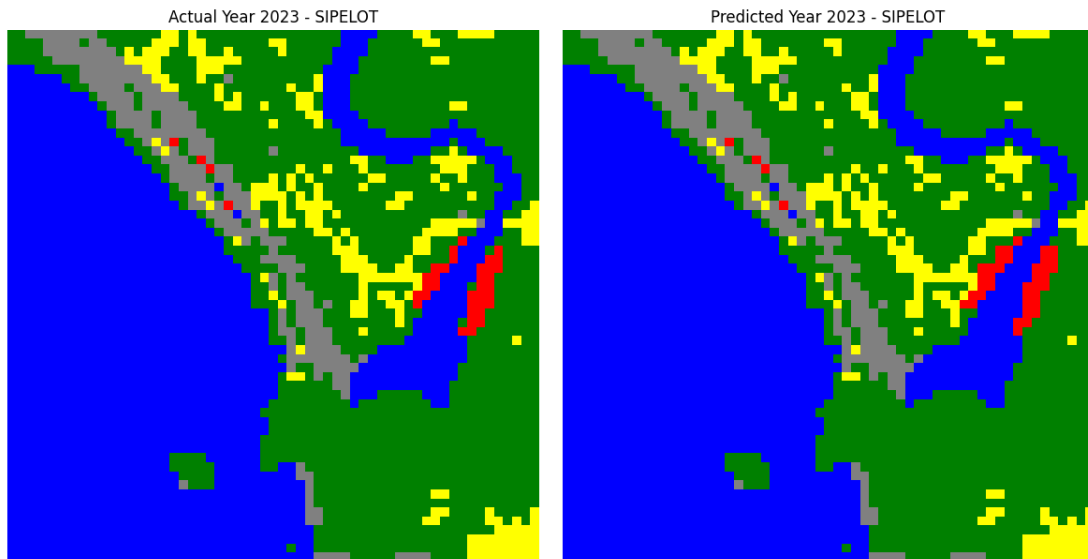
Gambar V.8 *Confusion Matrix* pada Data Uji Pantai Sipelot

Berdasarkan Gambar V.8, *confusion matrix* menunjukkan bahwa model berhasil mengklasifikasikan sebagian besar kelas dengan baik, tetapi terdapat beberapa kesalahan yang lebih signifikan pada kelas 1 (hutan *mangrove*) dan 3 (pemukiman) yang terlihat lebih banyak menghasilkan *false positive* pada masing-masing kelas dibandingkan dengan dua pantai sebelumnya. Kelas 2 (perairan) memiliki jumlah klasifikasi benar tertinggi, yaitu 1.720 sampel, diikuti oleh kelas 1 dengan 1.151 sampel yang benar diprediksi.

Tabel V-6 Metriks Evaluasi pada Data Uji Pantai Sipelot

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
0	88%	91%	89%
1	99%	98%	98%
2	100%	100%	100%
3	94%	98%	96%
4	97%	97%	97%
5	85%	96%	90%

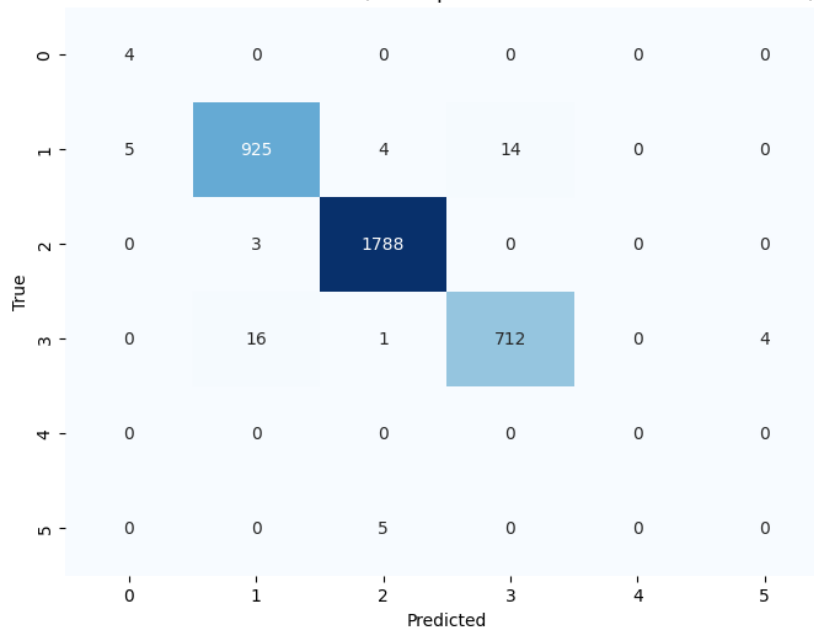
Pada Tabel V-6, hasil akurasi model menunjukkan performa yang sangat baik di hampir semua kelas untuk Pantai Sipelot. Kelas 2 (perairan) tidak mengalami kesalahan prediksi dengan *f-1 score* 100%. Kelas 1 (hutan *mangrove*), 3 (pemukiman), dan 4 (tambak) juga menunjukkan kinerja tinggi dengan *precision* dan *recall* di atas 94%, menghasilkan *f1-score* yang hampir mendekati nilai optimal. Pada kelas 0 (penurunan hutan), model memiliki nilai *precision* 88% dan *recall* 91%, yang berarti terdapat beberapa kesalahan prediksi berupa *false positives* dan *false negatives*. Namun, masih memiliki nilai *f-1 score* cukup baik yaitu 89%.



Gambar V.9 Visualisasi Hasil Prediksi Data Uji Pantai Sipelot

Pada Gambar V.9, visualisasi hasil prediksi secara garis besar sudah menunjukkan pola yang cukup akurat dibandingkan dengan data aktual. Tidak seperti dua pantai sebelumnya dimana kesalahan prediksi muncul pada posisi piksel yang cukup acak, pada Pantai Sipelot misklasifikasi terjadi karena sebagian besar piksel aktual diprediksi sebagai piksel tetangganya. Hal ini yang membuat hasil visualisasi pada area ini terlihat cukup serupa karena hampir tidak terdapat piksel salah prediksi yang muncul secara acak.

### V.2.3.1 Analisis Pantai Tamban



Gambar V.10 *Confusion Matrix* pada Data Uji Pantai Tamban

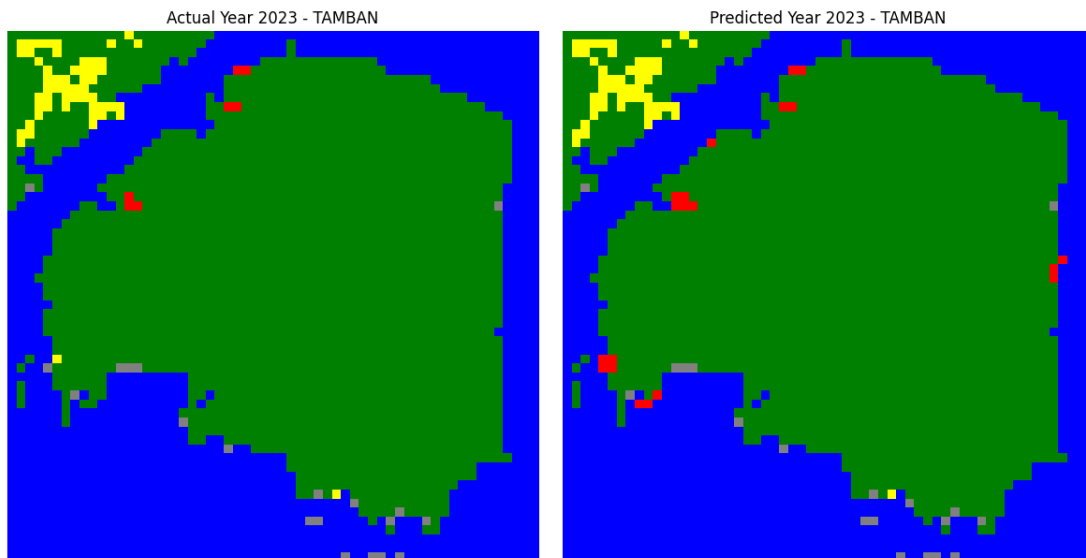
Berdasarkan Gambar V.10, *confusion matrix* menunjukkan bahwa Pantai Tamban menjadi lokasi paling menantang bagi model dalam melakukan klasifikasi perubahan tutupan lahan. Hasil *confusion matrix* menunjukkan model mengalami lebih banyak kesalahan klasifikasi dibandingkan dengan tiga lokasi sebelumnya. Kelas 2 (perairan) memiliki jumlah prediksi benar tertinggi, yaitu 1.788 sampel, sedangkan kelas 1 (hutan *mangrove*) memiliki 925 sampel yang diklasifikasikan dengan benar. Namun, kelas 3 (pemukiman) memiliki lebih banyak misprediksi, dengan 14 sampel yang salah diklasifikasi.



Tabel V-7 Metriks Evaluasi pada Data Uji Pantai Tamban

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
0	44%	100%	62%
1	98%	98%	98%
2	99%	100%	100%
3	98%	97%	98%
4	0%	0%	0%
5	0%	0%	0%

Pada Tabel V-7, model menunjukkan kinerja yang sudah baik. Kelas 2 menunjukkan hasil hampir sempurna dengan *precision* 99%, *recall* 100%, dan *f1-score* 100%. Kelas 1 (hutan *mangrove*) dan 3 (pemukiman) juga memiliki nilai dengan *f1-score* sebesar 98% untuk kedua kelas. Namun, pada kelas 0 (penurunan hutan), model menunjukkan hasil yang kurang optimal, dengan *precision* 44% dan *f1-score* 62%, meskipun *recall* mencapai 100%. Hal ini menunjukkan bahwa model mampu mendeteksi seluruh kasus positif pada kelas 0, tetapi masih terdapat cukup banyak *false positives*, sehingga *precision* menjadi rendah. Sementara itu, pada kelas 4 (tambak), semua metrik bernilai nol karena tidak terdapat data dengan kelas 4 pada data di Pantai Bajul Mati.



Gambar V.11 Visualisasi Hasil Prediksi Data Uji Pantai Tamban

Pada Gambar V.11, visualisasi hasil prediksi secara keseluruhan memberikan hasil yang cukup serupa dengan visualisasi aktualnya. Namun, dapat terlihat secara langsung bahwa terdapat misklasifikasi pada kelas hutan *mangrove* menjadi penurunan hutan. Hal ini juga terlihat dari hasil perhitungan metrik evaluasi yang menunjukkan rendahnya *precision* pada kelas tersebut. Sedangkan untuk prediksi kelas lain, secara visual sudah terlihat cukup baik dari visualisasi aktualnya.

## BAB VI KESIMPULAN DAN SARAN

### VI.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa:

1. Konfigurasi terbaik dalam prediksi perubahan hutan *mangrove* dilakukan melalui eksplorasi kombinasi parameter menggunakan *framework optuna*. Dari berbagai skenario pengujian, konfigurasi model terbaik diperoleh pada pembagian data dengan skenario 12 area latih dan 4 area uji. Model dengan parameter  $n\_estimators=100$ ,  $max\_depth=10$ ,  $min\_samples\_leaf=10$ ,  $min\_samples\_split=20$ , dan  $criterion='gini'$  menunjukkan performa paling optimal. Konfigurasi ini memberikan keseimbangan terbaik antara jumlah data latih dan uji.
2. Performa model *Random Forest* dalam memprediksi perubahan hutan *mangrove* menunjukkan hasil yang sangat baik, dengan akurasi tertinggi pada skenario pembagian data 12:4 sebesar 99,10%. Model mampu mengidentifikasi hutan *mangrove* dengan baik, sebagaimana ditunjukkan oleh *precision*, *recall*, dan *f1-score* pada kelas hutan *mangrove* yang masing-masing mencapai 99%. Secara keseluruhan, nilai *precision*, *recall*, dan *f1-score* model berturut-turut adalah 94%, 97%, dan 95%. Kelas perairan memiliki akurasi tertinggi, sementara tantangan utama terdapat pada prediksi perubahan hutan yang masih mengalami beberapa misprediksi.

### VI.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa langkah yang dapat diterapkan sebagai saran untuk meningkatkan akurasi dan generalisasi model dalam memprediksi perubahan hutan *mangrove* antara lain:

1. Penelitian selanjutnya dapat mengombinasikan model dengan sumber data lain, seperti penginderaan jauh dari satelit, untuk meningkatkan ketepatan dalam mendeteksi perubahan tutupan lahan. Data tambahan ini dapat memberikan informasi spektral yang lebih kaya, membantu model dalam mengenali pola

perubahan lahan yang lebih kompleks, serta meningkatkan kemampuan dalam membedakan kelas tutupan lahan yang memiliki karakteristik serupa.

2. penelitian selanjutnya dapat mengeksplorasi model lain seperti *Gradient Boosting* atau *Deep Learning* untuk membandingkan pendekatan yang lebih efektif dalam mengklasifikasikan perubahan tutupan lahan. Model berbasis *deep learning*, khususnya *Convolutional Neural Networks* (CNN), berpotensi lebih mampu menangkap pola spasial yang lebih kompleks, yang mungkin tidak sepenuhnya tertangkap oleh metode berbasis pohon keputusan seperti *Random Forest* .

## DAFTAR PUSTAKA

- Agus Supriyanto. (2013, Maret 13). *57 Persen Bakau di Pantai Selatan Malang Rusak*. Tempo. <https://nasional.tempo.co/read/466700/57-persen-bakau-di-pantai-selatan-malang-rusak>
- Ainurrohma. (2021). Akurasi Algoritma Klasifikasi pada Software Rapidminer dan Weka. *PRISMA, Prosiding Seminar Nasional Matematika*, 4.
- Alamri, M., & Ykhlef, M. (2024). Hybrid Feature Engineering Based on Customer Spending Behavior for Credit Card Anomaly and Fraud Detection. *Electronics (Switzerland)*, 13(20). <https://doi.org/10.3390/electronics13203978>
- Aliani, H., Malmir, M., Sourodi, M., & Kafaky, S. B. (2019). Change detection and prediction of urban land use changes by CA–Markov model (case study: Talesh County). *Environmental Earth Sciences*, 78(17). <https://doi.org/10.1007/s12665-019-8557-9>
- Alita, D., & Isnain, A. R. (2020). Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier. *jurnal komputasi*, 8(2). <https://doi.org/10.23960/komputasi.v8i2.2615>
- Amini, S., Saber, M., Rabiei-Dastjerdi, H., & Homayouni, S. (2022). Urban Land Use and Land Cover Change Analysis Using Random Forest Classification of Landsat Time Series. *Remote Sensing*, 14(11). <https://doi.org/10.3390/rs14112654>
- Amoakoh, A. O., Aplin, P., Rodríguez-Veiga, P., Moses, C., Alonso, C. P., Cortés, J. A., Delgado-Fernandez, I., Kankam, S., Mensah, J. C., & Nortey, D. D. N. (2024). Predictive Modelling of Land Cover Changes in the Greater Amanzule Peatlands Using Multi-Source Remote Sensing and Machine Learning Techniques. *Remote Sensing*, 16(21). <https://doi.org/10.3390/rs16214013>
- Aonpong, P., Kasetkasem, T., Rakwatin, P., Kumazawa, I., & Chanwimaluang, T. (2016). The Random forest approach for land cover mapping. *7th International Conference on Information Communication Technology for Embedded Systems 2016, IC-ICTES 2016*. <https://doi.org/10.1109/ICTEmSys.2016.7467112>
- Asif, M., Kazmi, J. H., Tariq, A., Zhao, N., Guluzade, R., Soufan, W., Almutairi, K. F., Sabagh, A. El, & Aslam, M. (2023). Modelling of land use and land cover changes and prediction using CA-Markov and Random Forest. *Geocarto International*, 38(1). <https://doi.org/10.1080/10106049.2023.2210532>
- Bui, D. H., & Mucsi, L. (2021). From land cover map to land use map: A combined pixel-based and object-based approach using multi-temporal landsat data, a

- random forest classifier, and decision rules. *Remote Sensing*, 13(9). <https://doi.org/10.3390/rs13091700>
- Cheng, K., & Wang, J. (2019). Forest type classification based on integrated spectral-spatial-temporal features and random forest algorithm-A case study in the Qinling Mountains. *Forests*, 10(7). <https://doi.org/10.3390/f10070559>
- Chowdhury, M. S. (2024). Comparison of accuracy and reliability of random forest, support vector machine, artificial neural network and maximum likelihood method in land use/cover classification of urban setting. *Environmental Challenges*, 14. <https://doi.org/10.1016/j.envc.2023.100800>
- Derhab, A., Aldweesh, A., Emam, A. Z., & Khan, F. A. (2020). Intrusion Detection System for Internet of Things Based on Temporal Convolution Neural Network and Efficient Feature Engineering. *Wireless Communications and Mobile Computing*, 2020. <https://doi.org/10.1155/2020/6689134>
- Douzas, G., Bacao, F., Fonseca, J., & Khudinyan, M. (2019). Imbalanced learning in land cover classification: Improving minority classes' prediction accuracy using the geometric SMOTE algorithm. *Remote Sensing*, 11(24). <https://doi.org/10.3390/rs11243040>
- Dzulfigar, A., Asy'ari, R., Rahmawati, A. D., Ulfa, A., Marfi, K. P., Puspitasari, R. F., Puspita, S., Adila, S., Firmansyah, J. C., Zamani, L. P., Pramulya, N. P., & Setiawan, R. (2024). Spatiotemporal Analysis of Mangrove in Subang Regency using Sentinel-2 Timeseries Data. Dalam *ojs SSRS Journal A: Agro-Environmental Research* (Vol. 2).
- Ebrahimi, H., Mirbagheri, B., Matkan, A. A., & Azadbakht, M. (2021). Per-pixel land cover accuracy prediction: A random forest-based method with limited reference sample data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 172. <https://doi.org/10.1016/j.isprsjprs.2020.11.024>
- Faheem, Z., Kazmi, J. H., Shaikh, S., Arshad, S., Noreena, & Mohammed, S. (2024). Random forest-based analysis of land cover/land use LCLU dynamics associated with meteorological droughts in the desert ecosystem of Pakistan. *Ecological Indicators*, 159. <https://doi.org/10.1016/j.ecolind.2024.111670>
- Farzanmanesh, R., Khoshelham, K., Volkova, L., Thomas, S., Ravelonjatovo, J., & Weston, C. J. (2024). Temporal Analysis of Mangrove Forest Extent in Restoration Initiatives: A Remote Sensing Approach Using Sentinel-2 Imagery. *Forests*, 15(3). <https://doi.org/10.3390/f15030399>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3).

- Ginanjar, C., Harfinda, E. M., & Saputra, R. (2023). Analisis Perubahan Garis Pantai dengan Pendekatan Penginderaan Jauh di Kecamatan Mempawah Hilir. *Jurnal Laut Khatulistiwa*, 6(3). <https://doi.org/10.26418/lkuntan.v6i3.68186>
- Han, S., Williamson, B. D., & Fong, Y. (2021). Improving random forest predictions in small datasets from two-phase sampling designs. *BMC Medical Informatics and Decision Making*, 21(1). <https://doi.org/10.1186/s12911-021-01688-3>
- Hartono, & Ongko, E. (2022). *Avoiding Overfitting dan Overlapping in Handling Class Imbalanced Using Hybrid Approach with Smoothed Bootstrap Resampling and Feature Selection*. [www.joiv.org/index.php/joiv](http://www.joiv.org/index.php/joiv)
- Hasanah, U., Soleh, A. M., & Sadik, K. (2024). Effect of Random Under sampling, Oversampling, and SMOTE on the Performance of Cardiovascular Disease Prediction Models. *Jurnal Matematika, Statistika dan Komputasi*, 21(1), 88–102. <https://doi.org/10.20956/j.v21i1.35552>
- Imaduddin, M. R., & Krisnadi, I. G. (2020). Konservasi Mangrove oleh Masyarakat Pesisir Malang Selatan 2012-2016. *Jurnal Penelitian Pendidikan Guru Sekolah Dasar*, 2(2).
- Indrawati, A., Subagyo, H., Sihombing, A., Wagiyah, W., & Afandi, S. (2020). Analyzing The Impact Of Resampling Method For Imbalanced Data Text In Indonesian Scientific Articles Categorization. *Baca: Jurnal Dokumentasi Dan Informasi*, 41(2). <https://doi.org/10.14203/j.baca.v41i2.702>
- Jhonnerie, R., Siregar, V. P., Nababan, B., Prasetyo, L. B., & Wouthuyzen, S. (2015). Random Forest Classification for Mangrove Land Cover Mapping Using Landsat 5 TM and Alos Palsar Imageries. *Procedia Environmental Sciences*, 24. <https://doi.org/10.1016/j.proenv.2015.03.028>
- Jui, S. J. J., Ahmed, A. A. M., Bose, A., Raj, N., Sharma, E., Soar, J., & Chowdhury, M. W. I. (2022). Spatiotemporal Hybrid Random Forest Model for Tea Yield Prediction Using Satellite-Derived Variables. *Remote Sensing*, 14(3). <https://doi.org/10.3390/rs14030805>
- Katya, E. (2023). Exploring Feature Engineering Strategies for Improving Predictive Models in Data Science. *Research Journal of Computer Systems and Engineering*, 4(2). <https://doi.org/10.52710/rjcse.88>
- Khan, F., Ibrahim Channa, M., Ali Soomro, M., Professor, A., Zaman Nizamani, S., & Aamir Bhutto, M. (2024). Advancing Machine Learning: Development, Evaluation, and Feature Engineering in Domain-Specific Applications. *International Journal on Recent and Innovation Trends in Computing and Communication*, 12. <http://www.ijritcc.org>

- Khan, M. Y., Qayoom, A., Nizami, M. S., Siddiqui, M. S., Wasi, S., & Raazi, S. M. K. U. R. (2021). Automated Prediction of Good Dictionary EXamples (GDEX): A Comprehensive Experiment with Distant Supervision, Machine Learning, and Word Embedding-Based Deep Learning Techniques. *Complexity*, 2021. <https://doi.org/10.1155/2021/2553199>
- Leal, M., Spalding Cartography by Kate Longley-Wood with, M. D., An, T., Andradi-Brown, D., Arquiza, Y., Barnes, E., Beeston, M., Benbow, S., Berzina-Rodrigo, A., Bood, N., Brown, C., Bunting, P., Cameron, C., Canty, S., Cummings-Krueger, E., Dahdouh-Guebas, F., de Hoog, A., Debney, A., Claudia Díazgranados, M., ... zu Ermgassen, P. (2022). *Ecosystems of Hope Map credits The Global Mangrove Alliance*. [www.mangrovealliance.org](http://www.mangrovealliance.org) and [www.globalmangrovewatch.org](http://www.globalmangrovewatch.org)
- Letsoin, S. M. A., Herak, D., Rahmawan, F., & Purwestri, R. C. (2020). Land cover changes from 1990 to 2019 in Papua, Indonesia: Results of the remote sensing imagery. *Sustainability (Switzerland)*, 12(16). <https://doi.org/10.3390/su12166623>
- Lubis, R. E. L. (2023, Agustus 15). *Hutan mangrove terancam hilang di Kalimantan - Dilema masyarakat Kubu Raya antara lingkungan dan minimnya lapangan kerja*. BBC News Indonesia. <https://www.bbc.com/indonesia/indonesia-66269380>
- Maimon, O., & Rokach, L. (2006). Introduction to Knowledge Discovery in Databases. Dalam *Data Mining and Knowledge Discovery Handbook*. [https://doi.org/10.1007/0-387-25465-x\\_1](https://doi.org/10.1007/0-387-25465-x_1)
- Maurya, K., Mahajan, S., & Chaube, N. (2021). Remote sensing techniques: mapping and monitoring of mangrove ecosystem—a review. Dalam *Complex and Intelligent Systems* (Vol. 7, Nomor 6). <https://doi.org/10.1007/s40747-021-00457-z>
- Nurhopipah, A., Ceasar, Y., & Priadana, A. (2021). Improving Machine Learning Accuracy using Data Augmentation in Recruitment Recommendation Process. *3rd 2021 East Indonesia Conference on Computer and Information Technology, EIConCIT 2021*. <https://doi.org/10.1109/EIConCIT50028.2021.9431908>
- Nurhopipah, A., & Magnolia, C. (2023). Perbandingan Metode Resampling Pada Imbalanced Dataset Untuk Klasifikasi Komentar Program MBKM. *Jurnal Publikasi Ilmu Komputer dan Multimedia*, 2(1). <https://doi.org/10.55606/jupikom.v2i1.862>
- Putri, M. R., & Aini, N. (2018, Januari 10). *Kementerian LHK Sebut 1,81 Juta Hektare Mangrove Rusak*. REPUBLIKA.



- Ramírez, J. L. S., Jiménez-Cruz, R., Villuendas-Rey, Y., & Yáñez-Márquez, C. (2023). Random forest Algorithm for the Classification of Spectral Data of Astronomical Objects. *Algorithms*, 16(6). <https://doi.org/10.3390/a16060293>
- Riyanto, U. (2019). Analisis Perbandingan Algoritma Naive Bayes Dan Support Vector Machine Dalam Mengklasifikasikan Jumlah Pembaca Artikel Online. *JIKA (Jurnal Informatika)*, 2(2). <https://doi.org/10.31000/v2i2.1521>
- Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random Forest Algorithm Overview. *Babylonian Journal of Machine Learning*, 2024, 69–79. <https://doi.org/10.58496/bjml/2024/007>
- Suci Amaliah, Nusrang, M., & Aswi, A. (2022). Penerapan Metode Random Forest Untuk Klasifikasi Variasi Minuman Kopi di Kedai Kopi Konijiwa Bantaeng. *VARIANSI: Journal of Statistics and Its application on Teaching and Research*, 4(3). <https://doi.org/10.35580/variansiunm31>
- Tran, T. V., Reef, R., & Zhu, X. (2022). A Review of Spectral Indices for Mangrove Remote Sensing. Dalam *Remote Sensing* (Vol. 14, Nomor 19). <https://doi.org/10.3390/rs14194868>
- Vilma, M. A., Fernando, A., Cravo, M., Massingue, M., Lima, H., Macamo, C., Bandeira, S., & Paula, J. (2022). A Comparison of Mangrove Forest Structure and Ecosystem Services in Maputo Bay (Eastern Africa) and Príncipe Island (Western Africa). *Forests*, 13(9). <https://doi.org/10.3390/fl3091466>
- Win, K. S., & Sasaki, J. (2024). The Change Detection of Mangrove Forests Using Deep Learning with Medium-Resolution Satellite Imagery: A Case Study of Wunbaik Mangrove Forest in Myanmar. *Remote Sensing*, 16(21). <https://doi.org/10.3390/rs16214077>
- Xiao, X., Zou, Y., Huang, J., Luo, X., Yang, L., Li, M., Yang, P., Ji, X., & Li, Y. (2024). An Interpretable Model for Landslide Susceptibility Assessment based on Optuna Hyperparameter Optimization and Random Forest. *Geomatics, Natural Hazards and Risk*, 15(1). <https://doi.org/10.1080/19475705.2024.2347421>
- Yi, X., Xu, Y., Hu, Q., Krishnamoorthy, S., Li, W., & Tang, Z. (2022). ASN-SMOTE: a synthetic minority oversampling method with adaptive qualified synthesizer selection. *Complex and Intelligent Systems*, 8(3). <https://doi.org/10.1007/s40747-021-00638-w>
- Zhai, Y., Qu, Z., & Hao, L. (2018). Land cover classification using integrated spectral, temporal, and spatial features derived from remotely sensed images. *Remote Sensing*, 10(3). <https://doi.org/10.3390/rs10030383>