

**KLASIFIKASI TUTUPAN LAHAN MENGGUNAKAN CITRA SATELIT
RESOLUSI TINGGI DENGAN METODE *RANDOM FOREST***

SKRIPSI

YEFTHA EL IMANI PARDEDE

201402075



**PROGRAM STUDI S1 TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA
MEDAN
2024**

KLASIFIKASI TUTUPAN LAHAN MENGGUNAKAN CITRA SATELIT
RESOLUSI TINGGI DENGAN METODE *RANDOM FOREST*

SKRIPSI

Diajukan untuk melengkapi tugas dan memenuhi syarat memperoleh ijazah
Sarjana Teknologi Informasi

YEFTHA EL IMANI PARDEDE

201402075



**PROGRAM STUDI S1 TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA
MEDAN
2024**

PERSETUJUAN

Judul : Klasifikasi Tutupan Lahan Menggunakan Citra Satelit
Resolusi Tinggi Dengan Metode *Random Forest*
Kategori : Skripsi
Nama Mahasiswa : Yeftha El Imani Pardede
Nomor Induk Mahasiswa : 201402075
Program Studi : Sarjana (S-1) Teknologi Informasi
Fakultas : Ilmu Komputer Dan Teknologi
Informasi Universitas Sumatera Utara

Medan, 04 Juli 2024

Komisi Pembimbing:

Pembimbing 2,

Pembimbing 1,



Fanindia Purnamasari, S.TI, M.IT.

NIP. 198908172019032023



Sarah Purnamawati, S.T., M.Sc.

NIP. 198302262010122003

Diketahui/disetujui oleh

Program Studi S1 Teknologi Informasi

Ketua,



Dedy Arisandi, S.T., M.Kom.

NIP. 197908312009121002

PERNYATAAN

KLASIFIKASI TUTUPAN LAHAN MENGGUNAKAN CITRA SATELIT RESOLUSI TINGGI DENGAN METODE *RANDOM FOREST*

SKRIPSI

Saya mengakui bahwa skripsi ini merupakan hasil karya saya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing telah disebutkan sumbernya.

Medan, 04 Juli 2024



Yeftha El Imani Pardede
201402075

UCAPAN TERIMA KASIH

Segala puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas rahmat dan karunia-Nya yang tak terhingga, yang telah menuntun penulis dalam menyelesaikan skripsi berjudul "Klasifikasi Tutupan Lahan Menggunakan Citra Satelit Resolusi Tinggi Dengan Metode *Random Forest*" sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer di Program Studi Teknologi Informasi Universitas Sumatera Utara. Penulis menyampaikan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan, bantuan, dan motivasi selama proses ini. Secara khusus, penulis berterima kasih kepada dosen pembimbing, yang telah memberikan bimbingan, nasihat, dan waktu berharga dalam penyusunan skripsi ini. Tak lupa, ucapan terima kasih juga ditujukan kepada orang tua dan keluarga, yang selalu memberikan doa, cinta, dan semangat. Terima kasih pula kepada teman-teman yang telah memberikan dukungan moral dan inspirasi. Semoga skripsi ini dapat memberikan manfaat bagi semua pihak.

1. Tuhan Yesus Kristus yang selalu memberikan kemampuan dan kemudahan kepada penulis dalam proses penyelesaian hingga tahap akhir penulisan.
2. Keluarga penulis yang selalu memberikan doa dan dukungan serta semangat hingga akhirnya penulis dapat menyelesaikan skripsi ini.
3. Ibu Dr. Maya Silvi Lydia, M.Sc selaku Dekan Fasilkom-TI USU.
4. Bapak Dedy Arisandi ST., M.Kom., selaku Ketua Program Studi S1 Teknologi Informasi Universitas Sumatera Utara dan Bapak Ivan Jaya S.Si., M.Kom., selaku Sekretaris Program Studi S1 Teknologi Informasi Universitas Sumatera Utara.
5. Ibu Sarah Purnamawati, S.T., M.Sc., selaku Dosen Pembimbing I dan Ibu Fanindia Purnamasari, S.TI, M.IT., selaku Dosen Pembimbing II yang telah membimbing penulis hingga akhir penulisan dengan memberikan saran dan masukan. Selaku dosen Pembimbing penulis yang telah banyak memberikan penulis.
6. Dosen Pengajar di Program Studi S1 Teknologi Informasi Universitas Sumatera Utara.

7. Staff dan pegawai di Program Studi S1 Teknologi Informasi Universitas Sumatera Utara.
8. Orang Tua Rohani yang sudah membimbing dan melatih penulis Ps. Johnson dan Ps. Lisbet, beserta saudara-saudara dalam rumah rohani yang menjadi rekan sekerja Sofi Ginting, Saor Sirait, Alfionita Manalu, John Silalahi, David Simbolon, Cherie Aritonang, Reza Surbakti, Yemima Panjaitan, Yemima Angelina, Nur Cahaya, Sevanya Gwen, serta saudara-saudari lain yang tidak bisa saya sebutkan satu persatu.
9. Teman-teman angkatan 2020 Teknologi Informasi Universitas Sumatera Utara yang memberikan dukungan dan motivasi kepada penulis.
10. Teman-teman seperjuangan penulis dalam masa perkuliahan Pretty Ohara Hutasoit, Vicky Natanael Sinaga, Felix Christian Oshwalt Rumahorbo, Albert Talupan Pangaribuan, Kevin Tulus Silitonga, Kelvin Lumbanraja, Frans Mayandro, M Arief Fadhlhan, Syavira Nur Khairani, Fildzah Zata Amani, Monica Juliana, dan Retno Wulan Sari.
11. Sahabat penulis sekaligus kakak abang, yang selalu memberikan dukungan dalam menyelesaikan tugas akhir ini, Sarah Theodora Sinurat, Vania Miranda Emmanuella Siahaan, Dhea Novianty Sitompul, Geylfedra Mathew.
12. Semua pihak yang tidak dapat penulis sebut satu per satu yang telah memberikan bantuan, perhatian, dan doa kepada penulis dalam menyelesaikan skripsi ini. Penulis mengucapkan terima kasih.

Medan, 04 Juli 2024

Penulis,



Yeftha El Imani Pardede

201402075

ABSTRAK

Pertumbuhan perkotaan yang cepat telah menyebabkan transformasi signifikan dalam karakteristik geografisnya. Khususnya di wilayah perkotaan, perubahan dalam penggunaan lahan dan infrastruktur telah menjadi sangat mencolok. Perubahan yang terjadi pada kondisi geografis, terutama penggunaan lahan, membuat pentingnya dilakukan pemetaan. Pemetaan menggunakan kemajuan teknologi dapat dilakukan terhadap perubahan lingkungan kota Medan guna mendukung upaya pembangunan yang berkelanjutan. Metode penelitian yang digunakan adalah algoritma Random Forest yang mampu menangani variabilitas spektral dalam data citra satelit. Data yang digunakan adalah citra satelit Sentinel-2 yang mencakup Band 2, Band 3, Band 4, dan Band 8 karena memiliki resolusi spasial 10 meter, yang sangat membantu untuk menghasilkan klasifikasi dengan resolusi tinggi. Dataset yang digunakan dalam penelitian ini mencakup 167 poligon yang dipecah menjadi 12.755 piksel. Data ini kemudian dibagi menjadi data train sebanyak 70%, data test sebanyak 20%, dan data validasi sebanyak 10% dari total data. Model *Random Forest* dilatih dan dievaluasi menggunakan data tersebut untuk mengklasifikasikan citra ke dalam enam jenis tutupan lahan: vegetasi, permukiman, perairan, lahan terbuka, pertanian, dan jalan. Hasil akhir menunjukkan bahwa model Random Forest memiliki akurasi keseluruhan mencapai 97.33%. *Precision*, *recall*, dan *F1 score* yang tinggi untuk masing-masing kelas menunjukkan kemampuan model dalam mengenali berbagai jenis tutupan lahan dengan baik.

Kata kunci: klasifikasi tutupan lahan, citra satelit, Sentinel-2, *Random Forest*, pemantauan lingkungan, resolusi spasial tinggi

HIGH-RESOLUTION SATELLITE IMAGERY LAND COVER CLASSIFICATION USING THE RANDOM FOREST METHOD

ABSTRACT

Rapid urban growth has significantly altered its geographical characteristics, particularly in urban areas where changes in land use and infrastructure are highly noticeable. These geographical changes, especially in land use, necessitate continuous monitoring. Leveraging technological advancements, environmental changes in Medan city can be monitored to support sustainable development efforts. This study employs the Random Forest algorithm, known for handling spectral variability in satellite imagery data. The data used consists of Sentinel-2 satellite images, specifically Band 2, Band 3, Band 4, and Band 8, chosen for their 10-meter spatial resolution, which is instrumental in producing high-resolution classifications. The dataset comprises 167 polygons, broken down into thousands of pixels. This data is divided into training data (70%), test data (20%), and validation data (10%). The Random Forest model is trained and evaluated using this data to classify the imagery into six land cover types: vegetation, urban areas, water bodies, open land, agricultural land, and roads. The final results reveal that the Random Forest model achieved an overall accuracy of 97.33%. High precision, recall, and F1 scores for each class demonstrate the model's capability to accurately identify various types of land cover.

Keywords: *land cover classification, Sentinel-2 satellite imagery, Random Forest, environmental monitoring, high-resolution spatial data*

DAFTAR ISI

PERSETUJUAN	i
PERNYATAAN	ii
UCAPAN TERIMA KASIH	iii
ABSTRAK	v
ABSTRACT	vi
DAFTAR ISI	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR PSEUDOCODE	xiii
BAB 1 PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan Penelitian	3
1.4. Batasan Masalah	3
1.5. Manfaat Penelitian	3
1.6. Metodologi Penelitian	4
1.7. Sistematika Penulisan	5
BAB 2 LANDASAN TEORI	7
2.1. Tutupan Lahan	7
2.2. Citra Satelit Sentinel-2A	7
2.3. Klasifikasi Tutupan Lahan	10
2.3.1. Lahan Terbuka	10
2.3.2. Permukiman	11
2.3.3. Pertanian	11

2.3.4. Perairan	11
2.3.5. Vegetasi	12
2.3.6. Jalan	12
2.4. Algoritma Random Forest	13
2.5. Simple Ratio (SR)	14
2.6. Normalized Difference Vegetation Index (NDVI)	15
2.7. Normalized Difference Water Index (NDWI)	15
2.8. Confusion Matrix	16
2.9. Website	17
2.10. HTML (HyperText Markup Language)	17
2.11. Bootstrap	18
2.12. JavaScript	18
2.13. Python	19
2.14. Flask	20
2.15. Penelitian Terdahulu	20
2.16. Perbedaan Penelitian	27
BAB 3 ANALISIS DAN PERANCANGAN SISTEM	29
3.1. Dataset	29
3.2. Arsitektur Umum	31
3.2.1. Pengambilan data citra satelit	32
3.2.2. Pre-Processing	32
3.2.2.1. Area of Interest (AOI)	32
3.2.2.2. Filtering Date	34
3.2.2.3. Filtering Cloud	35
3.2.2.4. Band Composit	36
3.2.2.5. Indeks (Simple Ratio, NDVI, dan NDWI)	37
3.2.3 Titik Sampel	40
3.2.4. Random Forest Algorithm	41

3.2.5. Learned Model	42
3.2.6. Klasifikasi Tutupan Lahan	44
3.2.7. Evaluasi	45
3.3. Perancangan Sistem	45
3.3.1. Desain Tampilan Halaman Depan	45
3.3.2. Desain Tampilan Training-Testing Data	46
3.3.3. Desain Tampilan Halaman Klasifikasi Citra	48
BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM	50
4.1. Implementasi Sistem	50
4.1.1. Spesifikasi Perangkat Keras dan Perangkat Lunak	50
4.1.2. Implementasi Perancangan Antarmuka	50
4.2. Implementasi Model	55
4.2.1. Hasil Pelatihan Model	55
4.2.2. Hasil Pengujian Model	61
4.3. Analisis Hasil	64
4.3.1. Analisis Hasil Vegetasi	65
4.3.2. Analisis Hasil Permukiman	66
4.3.3. Analisis Hasil Perairan	67
4.3.4. Analisis Hasil Lahan Terbuka	68
4.3.5. Analisis Hasil Pertanian	69
4.3.6. Analisis Hasil Jalan	70
4.3.7. Perbandingan Citra Raster	71
BAB 5 KESIMPULAN DAN SARAN	74
5.1. Kesimpulan	74
5.2. Saran	75
DAFTAR PUSTAKA	76

DAFTAR TABEL

Tabel 2. 1 Tabel <i>spectral bands</i> Sentinel-2A	9
Tabel 2. 2 Penelitian Terdahulu	23
Tabel 3. 1 Dataset (.shp)	30
Tabel 3. 2 Dataset dalam piksel	31
Tabel 3. 3 Band Composit Citra Raster	36
Tabel 3. 4 Band Composit <i>beserta Perhitungan Indeks</i>	37
Tabel 3. 5 Split Titik Sampel	41
Tabel 4. 1 Daftar Parameter	55
Tabel 4. 2 Hasil <i>RandomizedSearchCV</i>	55
Tabel 4. 3 Nilai TP, TN, FP, dan FN	59
Tabel 4. 4 Nilai <i>precision</i> , <i>recall</i> , dan <i>f1-score</i>	60
Tabel 4. 5 Hasil prediksi benar dan salah	63
Tabel 4. 6 Hasil prediksi benar dan salah (lanjutan)	64
Tabel 4. 7 Perhitungan <i>Precision</i> , <i>Recall</i> , dan <i>F1 Score</i> citra raster 13 band	73

DAFTAR GAMBAR

Gambar 2. 1 Satelit Sentinel-2A (ESA and AIRBUS Defence & Space)	8
Gambar 2. 2 Contoh citra satelit (dokumentasi pribadi)	10
Gambar 2. 3 Contoh citra lahan terbuka	10
Gambar 2. 4 Contoh citra permukiman	11
Gambar 2. 5 Contoh citra pertanian	11
Gambar 2. 6 Contoh citra perairan	11
Gambar 2. 7 Contoh citra vegetasi	12
Gambar 2. 8 Contoh citra jalan	12
Gambar 2. 9 Diagram Pohon Keputusan Random Forest (medium.com)	13
Gambar 3. 1 Citra raster daerah penelitian	29
Gambar 3. 2 Arsitektur Umum	32
Gambar 3. 3 Implementasi tahap <i>area of interest</i>	33
Gambar 3. 4 Implementasi <i>Filtering Cloud</i>	35
Gambar 3. 5 Titik sampel	40
Gambar 3. 6 Halaman Depan	46
Gambar 3. 7 Halaman Training-Testing	47
Gambar 3. 8 Halaman Klasifikasi Citra	49
Gambar 4. 1 Tampilan halaman depan	51
Gambar 4. 2 Tampilan halaman training-testing	51
Gambar 4. 3 Tampilan hasil training-testing	52
Gambar 4. 4 Tampilan hasil training-testing (lanjutan)	52
Gambar 4. 5 Tampilan hasil training-testing (lanjutan)	52
Gambar 4. 6 Tampilan hasil training-testing (lanjutan)	53
Gambar 4. 7 Tampilan hasil training-testing (lanjutan)	53
Gambar 4. 8 Tampilan halaman klasifikasi citra	54
Gambar 4. 9 Tampilan hasil prediksi klasifikasi citra	54
Gambar 4. 10 <i>Confusion Matrix</i>	58
Gambar 4. 11 Hasil Peta Tutupan Lahan	61
Gambar 4. 12 Hasil Peta Tutupan Lahan Vegetasi	65
Gambar 4. 13 Hasil Peta Tutupan Lahan Permukiman	67
Gambar 4. 14 Hasil Peta Tutupan Lahan Perairan	67

Gambar 4. 15 Hasil Peta Tutupan Lahan Terbuka	69
Gambar 4. 16 Hasil Peta Tutupan Pertanian	70
Gambar 4. 17 Hasil Peta Tutupan Lahan Jalan	71
Gambar 4. 18 Citra Raster dengan 13 <i>Band</i>	72
Gambar 4. 19 Gambar Hasil Klasifikasi dengan 13 <i>Band</i>	72

DAFTAR PSEUDOCODE

Pseudocode 3. 1 <i>Area of Interest</i>	33
Pseudocode 3. 2 <i>Filtering Date</i>	34
Pseudocode 3. 3 <i>Filtering Cloud</i>	35
Pseudocode 3. 4 <i>Band Composit</i>	37
Pseudocode 3. 5 <i>Index (Simple Ratio, NDVI, dan NDWI)</i>	40
Pseudocode 3. 6 <i>Learned Model</i>	44
Pseudocode 3. 7 <i>Klasifikasi Tutupan Lahan</i>	45

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Tutupan lahan adalah istilah yang mengacu pada cakupan biologis permukaan bumi yang dapat diamati sebagai hasil dari pengaturan, aktivitas, dan perlakuan manusia terhadap lahan tersebut. Tutupan lahan berhubungan dengan jenis-jenis kenampakan di permukaan bumi, seperti bangunan, danau, dan vegetasi (Lillesand et al., 2004). Informasi yang diperoleh dari tutupan lahan dapat digunakan untuk keperluan pemodelan dan membantu memahami fenomena alam yang terjadi di permukaan bumi (Gie, 2008). Informasi mengenai pemetaan tutupan lahan terbaru juga bermanfaat bagi pemerintah daerah atau pemangku kepentingan dalam merencanakan pembangunan dan memantau lingkungan.

Untuk memperoleh informasi mengenai tutupan lahan, teknik penginderaan jauh menjadi salah satu metode yang dapat digunakan (Sampurno & Thoriq, 2016). Penginderaan jauh ditemukan baik dalam pemantauan tutupan lahan karena kemampuannya menyediakan data tentang variasi spasial di permukaan bumi. Data yang direkam oleh sensor penginderaan jauh mencakup berbagai panjang gelombang, sehingga diperlukan koreksi radiometrik dan geometrik sebelum informasi spektral dapat diekstraksi (Havyas et al., 2015). Penggunaan citra satelit sangat efektif dalam pendeteksian tutupan lahan karena memiliki resolusi temporal yang baik dan cakupan wilayah yang luas, menjadikannya alat penting dalam penyediaan informasi geografis.

Adapun salah satu manfaat yang diperoleh dari penggunaan citra satelit atau penginderaan jauh adalah untuk melakukan klasifikasi pada tutupan lahan yang dapat digunakan guna pemetaan tata letak wilayah tertentu. Klasifikasi citra akan membantu pemangku kepentingan dalam kebijakan lingkungan dengan adanya informasi mengenai tutupan lahan sebagai dasar interpretasi, analisis, dan pemodelan data spasial (Jensen & Cowen, 1999).

Penelitian terkait yang sudah dilakukan sebelumnya dilakukan oleh Veriana et al. (2018) dengan judul *Identifikasi Dan Klasifikasi Tutupan Lahan Melalui Pengolahan Citra Google Earth Dengan Metode Singular Value Decomposition dan Klasifikasi K-Nearest Neighbor*. Penelitian ini menggunakan citra satelit yang berasal dari *Google Earth* sebagai sampel yang akan dideteksi dengan format JPEG dengan ukuran

256x256pixel. Akurasi yang diperoleh menggunakan metode SVD dan KNN menghasilkan nilai akurasi sebesar 87.14%.

Afasel et al. (2022) mengadakan studi berjudul *Klasifikasi Tutupan Lahan dengan Menggunakan Supervised Machine Learning pada Citra Satelit melalui Google Earth Engine*. Penelitian ini memanfaatkan citra satelit Sentinel-2 yang memiliki resolusi sedang. Klasifikasi tutupan lahan dibagi menjadi empat jenis: badan air, area terbangun, lahan terbuka, dan vegetasi. Tahap pra-pemrosesan melibatkan persiapan data, klasifikasi, dan pembuatan matriks kebingungan. Setelah pra-pemrosesan selesai, citra kemudian melalui tahap klasifikasi. Hasil menunjukkan akurasi keseluruhan sebesar 96,66% dengan algoritma *minimum distance*, 91,66% dengan algoritma *CART*, dan 90% dengan algoritma *gradient boost*.

Zhang et al. (2021) melakukan penelitian berjudul *Sentinel-2 Satellite Imagery for Urban Land Cover Classification by Optimized Random Forest Classifier*. Data citra satelit diambil dari satelit sentinel-2A/B pada daerah Beijing, Tiongkok. Penelitian ini dilakukan guna mengetahui tutupan lahan pada daerah kota dengan urbanisasi yang cepat. Hasilnya menunjukkan bahwa pengklasifikasi *Random Forest* yang dioptimalkan mencapai kinerja terbaik dengan akurasi keseluruhan (OA: 0,879), koefisien kappa (0,8210), sedangkan *SVM* mencapai nilai *OA* dan *kappa* yang rendah masing-masing sebesar 0,461 dan 0,2526.

Berdasarkan latar belakang dan penelitian terdahulu di atas, maka penulis mengajukan penelitian yang berjudul “Klasifikasi Tutupan Lahan menggunakan Citra Satelit Resolusi Tinggi dengan Metode Random Forest.”

1.2. Rumusan Masalah

Pertumbuhan perkotaan yang cepat telah menyebabkan transformasi signifikan dalam karakteristik geografisnya. Khususnya di wilayah perkotaan, perubahan dalam penggunaan lahan dan infrastruktur telah menjadi sangat mencolok. Perubahan yang terjadi pada kondisi geografis terkhusus penggunaan lahan membuat pentingnya dilakukan pemetaan. Pemetaan menggunakan kemajuan teknologi dapat dilakukan terhadap perubahan lingkungan kota medan guna mendukung upaya pembangunan yang berkelanjutan.

1.3. Tujuan Penelitian

Penelitian ini dilakukan dengan tujuan mengklasifikasi tutupan lahan dengan metode *Random Forest* pada citra satelit beresolusi tinggi.

1.4. Batasan Masalah

Dalam penelitian ini terdapat beberapa batasan masalah agar penelitian ini tidak menyimpang dari tujuan yang akan dicapai. Batasan yang disebut adalah :

1. Daerah penelitian dilakukan di sekitar 5 wilayah kecamatan kota medan.
2. Daerah penelitian tersebut akan digunakan dalam penentuan *sample*.
3. Kelas tutupan lahan yang diklasifikasikan sejumlah 6 yaitu, Pertanian, Vegetasi, Lahan terbuka, Permukiman, Jalan, dan Perairan.
4. Citra satelit resolusi tinggi yang dimaksud adalah citra dengan resolusi spasial 10 meter.
5. Keluaran yang dihasilkan berupa gambar peta dengan klasifikasi tutupan lahan.

1.5. Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Hasil penelitian diharapkan dapat membantu untuk perencanaan pembangunan dan pemantauan lingkungan.
2. Mengetahui performa *random forest* dalam melakukan klasifikasi tutupan lahan pada citra satelit resolusi tinggi.
3. Hasil penelitian diharapkan dapat menjadi sumber rujukan dalam pengembangan klasifikasi tutupan lahan menggunakan citra satelit resolusi tinggi dengan menggunakan metode *random forest*.

1.6. Metodologi Penelitian

Tahapan-tahapan yang dilakukan dalam penelitian ini yaitu :

1. Studi Literatur

Di dalam tahap ini, penulis mencari dan mengumpulkan informasi melalui publikasi penelitian, artikel, jurnal, dan materi bacaan lain mengenai tutupan lahan, penggunaan citra satelit, dan metode *machine learning random forest*.

2. Analisis Permasalahan

Dari informasi yang telah dikumpulkan, penulis melakukan analisis guna memahami konsep penggunaan metode *random forest* dalam penelitian untuk mengklasifikasi tutupan lahan dengan citra satelit.

3. Perancangan Sistem

Dalam tahap ini perancangan sistem dilakukan untuk membangun sistem yang dapat mengklasifikasi tutupan lahan menggunakan citra satelit resolusi tinggi dengan metode *Random Forest*. Perancangan sistem yang dilakukan melibatkan arsitektur sistem, mulai dari pengumpulan data, pengolahan data, pembuatan model, sampai dengan validasi model.

4. Implementasi

Sistem yang telah dirancang akan dibangun dan diimplementasikan untuk mengklasifikasikan tutupan lahan pada citra satelit resolusi tinggi. Implementasi ini melibatkan penggunaan model *Random Forest* yang telah divalidasi untuk memprediksi jenis tutupan lahan pada citra baru.

5. Pengujian

Pengujian sistem dilakukan untuk memastikan bahwa sistem klasifikasi tutupan lahan yang telah dirancang dan diimplementasikan dapat berfungsi dengan baik dan memberikan hasil.

6. Penyusunan Laporan

Pada tahap ini peneliti akan melakukan penyusunan laporan serta dokumentasi dari hasil penelitian yang dilakukan.

1.7. Sistematika Penulisan

Sistematika penulisan yang dilakukan pada penelitian ini dibagi menjadi 5 (lima) bagian, yaitu.

Bab 1: Pendahuluan

Bab ini berisi penjelasan mengenai latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, metodologi penelitian, dan sistematika penulisan.

Bab 2: Landasan Teori

Landasan teori yang dimuat dalam bab ini adalah teori-teori yang digunakan dalam penerapan memahami permasalahan yang terdapat selama penelitian. Pada bab ini juga akan dimuat penjelasan mengenai tutupan lahan, klasifikasi tutupan lahan, citra satelit, metode dan cara kerja *Random Forest*. Setelah itu penelitian terdahulu yang dijadikan referensi dalam penelitian juga turut disertakan sebagai referensi ataupun rujukan dalam melakukan penelitian.

Bab 3: Analisis dan Perancangan Sistem

Bab ini berisi analisis yang dilakukan terhadap permasalahan yang diangkat dan sistem yang dirancang dalam mengklasifikasikan tutupan lahan dengan penerapan *Random Forest*. Bab ini menguraikan urutan penyelesaian masalah serta membahas arsitektur umum pada sistem.

Bab 4: Implementasi dan Pengujian Sistem

Bab ini menjelaskan implementasi dan pengujian sistem klasifikasi tutupan lahan menggunakan citra satelit resolusi tinggi dengan metode *Random Forest*. Penjelasan dari bagaimana analisis dan desain yang telah dirancang sebelumnya diimplementasikan, serta hasil akhir sistem dan pelaksanaan evaluasi terhadap sistem akan diuraikan pada bab ini.

Bab 5: Kesimpulan dan Saran

Bab ini merupakan bagian akhir dari penelitian ini. Di dalamnya terdapat kesimpulan yang diambil berdasarkan uraian keseluruhan dari bab-bab sebelumnya. Selain itu, saran untuk penelitian dan pengembangan lebih lanjut juga disertakan untuk memberikan panduan bagi penelitian di masa mendatang.

BAB 2

LANDASAN TEORI

2.1. Tutupan Lahan

Tutupan lahan merujuk pada cakupan biologis yang terlihat pada permukaan bumi sebagai hasil dari pengaturan, aktivitas, dan perlakuan manusia terhadap lahan tersebut. Tutupan lahan merepresentasikan wajah bumi yang teramati, meliputi berbagai kenampakan fisik seperti hamparan air, batuan, dan area perkotaan. Tutupan lahan berkaitan dengan tipe lahan fisik dan manifestasi objek-objek yang meliputi suatu wilayah tanpa mempertimbangkan aktivitas manusia terhadap objek-objek tersebut (Lillesand et al., 2004). Sebaliknya, penggunaan lahan berkaitan dengan aktivitas manusia di area tertentu. Dalam konteks pemukiman, tutupan lahan mencakup elemen-elemen seperti atap, permukaan keras, rumput, dan pepohonan.

Tutupan lahan memiliki berbagai kegunaan yang penting, seperti analisis perkembangan hutan yang melibatkan proses degradasi, deforestasi, dan reforestasi, juga untuk menghitung cadangan dan emisi karbon. Selain itu, tutupan lahan digunakan dalam perencanaan dan pengembangan wilayah, pengawasan izin usaha di kawasan hutan, termasuk pemantauan aktivitas penebangan, penanaman, dan pembukaan tambang. Tutupan lahan juga diperlukan untuk memonitor area konservasi dan pengelolaan hutan guna mencegah perambahan, pembalakan liar, serta kebakaran lahan dan hutan (Badan Standardisasi Nasional, 2010).

2.2. Citra Satelit Sentinel-2A

Gambar adalah representasi visual atau foto dari kondisi permukaan bumi yang dihasilkan melalui penginderaan spektrum elektromagnetik. Citra satelit adalah gambaran permukaan bumi atau planet lain yang ditangkap oleh satelit pengamat bumi milik pemerintah atau perusahaan di berbagai negara. Gambar-gambar ini dihasilkan oleh sensor yang dipasang pada satelit yang mengorbit ratusan kilometer di atas permukaan bumi. Citra satelit memiliki manfaat yang luas dalam berbagai bidang, seperti meteorologi, oseanografi, perikanan, pertanian, konservasi keanekaragaman hayati, kehutanan, lanskap, geologi, kartografi, perencanaan wilayah, pendidikan, dan intelijen. Dalam konteks penginderaan jauh, objek atau area yang dipelajari dapat berada di permukaan bumi, atmosfer, atau planet lain di luar angkasa (Lillesand &

Kiefer, 1994). Khusus untuk perencanaan wilayah, citra satelit dapat memfasilitasi pemetaan dan analisis area yang sedang diteliti melalui gambar yang diambil dari satelit.

Dalam penelitian ini, citra yang digunakan berasal dari satelit Sentinel-2A. Sentinel-2A memiliki 13 band, dengan 4 band beresolusi 10 meter, 6 band beresolusi 20 meter, dan 3 band beresolusi spasial 60 meter, serta memiliki area cakupan seluas 290 km.



Gambar 2. 1 Satelit Sentinel-2A (ESA and AIRBUS Defence & Space)

Citra satelit Sentinel-2 dapat digunakan untuk berbagai keperluan monitoring lahan dan menyediakan data dasar yang berharga untuk aplikasi pemantauan dan perencanaan lingkungan (Kawamuna et al., 2017). Keunggulan ini ditegaskan oleh penelitian oleh Andiko et al. (2019) dengan judul "Efisiensi Penggunaan Citra Multisensor untuk Pemetaan Tutupan Lahan." Penelitian ini membandingkan efektivitas dan efisiensi beberapa jenis citra satelit, termasuk SPOT 7, Landsat-8, dan Sentinel-2, dalam pembuatan peta tutupan lahan di kawasan Hutan Tahura WAR. Metode yang digunakan adalah *Analytical Hierarchy Process (AHP)* dan perbandingan parameter untuk menilai berbagai aspek citra satelit yang digunakan.

Hasil penelitian mereka menunjukkan bahwa citra Sentinel-2 menonjol sebagai yang paling efisien. Hal ini terutama dilihat dari segi biaya operasional yang lebih rendah, tingkat akurasi yang tinggi, proses pengolahan data yang lebih efisien, serta resolusi citra yang sangat baik. Dengan demikian, citra Sentinel-2 dianggap sangat unggul dan menjadi pilihan yang lebih baik dibandingkan dengan citra satelit lainnya

untuk pemetaan tutupan lahan. Temuan ini menggarisbawahi pentingnya penggunaan teknologi canggih seperti citra Sentinel-2 dalam berbagai aplikasi lingkungan dan perencanaan wilayah untuk mencapai hasil yang optimal.

Tabel 2. 1 Tabel *spectral bands* Sentinel-2A

Band	Nama Band	Panjang Gelombang (nm)	Resolusi Spasial (m)
1	Coastal Aerosol	433 - 453	60
2	Blue	458 - 523	10
3	Green	543 - 578	10
4	Red	650 - 680	10
5	Red Edge 1	698 - 713	20
6	Red Edge 2	733 - 748	20
7	Red Edge 3	773 - 793	20
8	NIR (Near Infrared)	785 - 899	10
8A	Narrow NIR	855 - 875	20
9	Water Vapor	935 - 955	60
10	SWIR - Cirrus	1360 - 1390	60
11	SWIR 1	1565 - 1655	20
12	SWIR 2	2100 - 2280	20

Dalam penelitian ini, citra yang digunakan adalah citra dengan resolusi spasial 10 meter, yaitu Band 2 (Blue), Band 3 (Green), Band 4 (Red), dan Band 8 (Near Infrared) dari satelit Sentinel-2A. Pemilihan band-band ini didasarkan pada kemampuan resolusi tinggi yang dimilikinya, yang memungkinkan deteksi detail-detail kecil pada tutupan lahan. Resolusi spasial 10 meter dianggap sebagai resolusi tinggi karena mampu memberikan informasi yang sangat rinci mengenai fitur-fitur permukaan bumi. Menurut Roy et al. (2014), citra dengan resolusi spasial 10 meter atau lebih kecil dikategorikan sebagai resolusi tinggi karena mampu menangkap detail yang sangat baik dari objek di permukaan bumi.



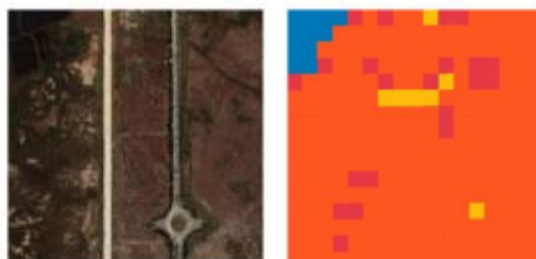
Gambar 2. 2 Contoh citra satelit (dokumentasi pribadi)

2.3. Klasifikasi Tutupan Lahan

Dalam upaya pengelompokan jenis tutupan lahan, klasifikasi tutupan lahan dapat dilakukan dalam proses pengelompokan tersebut. Menurut Badan Standarisasi Nasional (2010), dalam skala nasional, terdapat 22 kelas tutupan lahan, yang terbagi menjadi 7 kelas tutupan lahan hutan dan 15 kelas tutupan lahan bukan hutan. Penelitian ini menggunakan beberapa kelas tutupan lahan, yaitu lahan terbuka, permukiman, pertanian, perairan, vegetasi, dan jalan.

2.3.1. Lahan Terbuka

Lahan terbuka adalah area yang tidak memiliki tutupan, baik yang bersifat alami, semi alami, maupun buatan. Lahan terbuka ini dapat diklasifikasikan menjadi permukaan terkonsolidasi dan permukaan tidak terkonsolidasi (Badan Standardisasi Nasional, 2010).



Gambar 2. 3 Contoh citra lahan terbuka

2.3.2. Permukiman

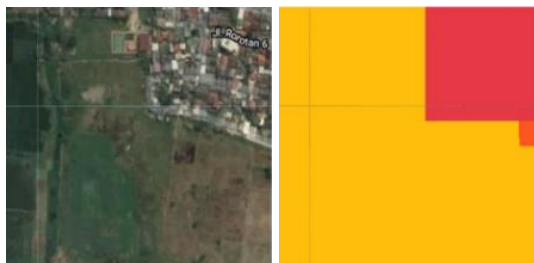
Permukiman adalah area atau lahan yang digunakan sebagai tempat tinggal atau hunian serta tempat untuk aktivitas yang mendukung kehidupan (Badan Standardisasi Nasional, 2010).



Gambar 2. 4 Contoh citra permukiman

2.3.3. Pertanian

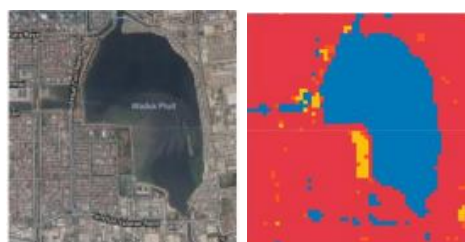
Pertanian adalah area yang digunakan untuk budidaya tanaman pangan, perkebunan, dan hortikultura. Pada periode antara musim tanam, area ini sering kali tidak memiliki tutupan vegetasi. Seluruh vegetasi yang ditanam dengan tujuan untuk dipanen termasuk dalam kategori ini (Badan Standardisasi Nasional, 2010).



Gambar 2. 5 Contoh citra pertanian

2.3.4. Perairan

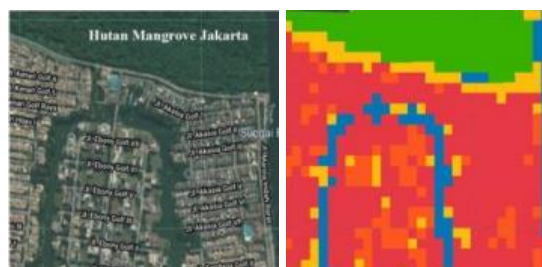
Perairan mencakup semua bentuk kenampakan air, termasuk laut, waduk, terumbu karang, dan padang lamun (Badan Standardisasi Nasional, 2010).



Gambar 2. 6 Contoh citra perairan

2.3.5. Vegetasi

Vegetasi terdiri dari hutan lahan kering dan hutan lahan basah. Hutan lahan kering tumbuh dan berkembang di habitat lahan kering seperti dataran rendah, perbukitan, pegunungan, atau hutan tropis dataran tinggi. Sementara itu, hutan lahan basah tumbuh di habitat lahan basah seperti rawa, termasuk rawa payau dan rawa gambut. Wilayah lahan basah memiliki karakteristik unik seperti: (1) dataran rendah yang membentang sepanjang pesisir, (2) wilayah berelevasi rendah, (3) area yang dipengaruhi oleh pasang-surut dekat pantai, (4) wilayah yang dipengaruhi oleh musim jauh dari pantai, dan (5) sebagian besar wilayah tertutup gambut (Badan Standardisasi Nasional, 2010).



Gambar 2. 7 Contoh citra vegetasi

2.3.6. Jalan

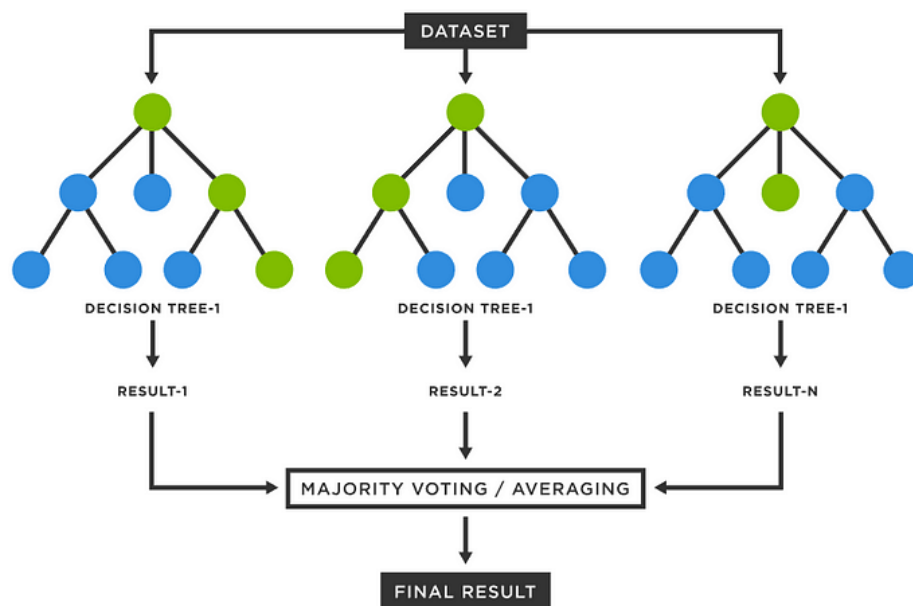
Jalan adalah jaringan infrastruktur transportasi yang dirancang untuk lalu lintas kendaraan (Badan Standardisasi Nasional, 2010).



Gambar 2. 8 Contoh citra jalan

2.4. Algoritma Random Forest

Random Forest merupakan algoritma supervised learning yang diperkenalkan oleh Leo Breiman. Algoritma Random Forest menggabungkan beberapa Decision Tree untuk mendapatkan model yang lebih baik. Random Forest akan membentuk gabungan dari beberapa *Trees* dari varian acak dari algoritma induksi *tree-based* (Gilles, 2014). Tree yang ada akan bergantung pada nilai piksel yang diambil pada setiap vektor secara acak dan independen (Breiman, 2001).



Gambar 2. 9 Diagram Pohon Keputusan Random Forest (medium.com)

Langkah-langkah utama dalam cara kerja Random Forest adalah sebagai berikut:

1. Pemilihan Sampel: Dari dataset asli, beberapa subset dibuat melalui bootstrapping (sampling dengan penggantian).
2. Pembuatan Pohon: Pohon keputusan dibangun untuk setiap subset. Pada setiap simpul pohon, hanya subset dari fitur yang dipertimbangkan untuk pembelahan.
3. Penggabungan Hasil: Untuk klasifikasi, setiap pohon memberikan "suara" untuk kelas dan kelas dengan suara terbanyak dipilih sebagai hasil akhir. Untuk regresi, hasil akhir adalah rata-rata dari semua hasil pohon.

Adapun kelebihan dari algoritma *Random Forest* dalam Klasifikasi Tutupan Lahan adalah :

1. Akurasi tinggi: *Random Forest* menghasilkan akurasi tinggi karena memanfaatkan kekuatan dari banyak pohon keputusan dan mengurangi overfitting (Breiman, 2001).
2. Kemampuan mengatasi *overfitting*: Algoritma ini dirancang untuk mengatasi overfitting dengan menggabungkan hasil dari beberapa pohon keputusan yang dibuat dari subset data (Breiman, 2001) .
3. Skalabilitas: *Random Forest* dapat diterapkan pada dataset yang besar dan kompleks seperti citra satelit dengan efisiensi tinggi (Belgiu & Drăgu, 2016) .
4. *Robust* terhadap *noise*: Algoritma ini robust terhadap data yang noise karena pengambilan keputusan didasarkan pada hasil mayoritas dari banyak pohon (Immitzer et al., 2012).
5. Kemampuan *handle* fitur: *Random Forest* dapat menangani data dengan banyak fitur tanpa perlu melakukan pengurangan dimensi secara eksplisit, yang sangat berguna dalam analisis citra satelit (Belgiu & Drăgu, 2016) .

2.5. Simple Ratio (SR)

Simple Ratio (SR) adalah salah satu indeks vegetasi yang digunakan untuk mengukur dan memantau kondisi vegetasi menggunakan citra satelit. Indeks ini diperkenalkan oleh Tucker (1979) dan diperoleh dengan membandingkan reflektansi di pita near-infrared (NIR) dengan pita merah (Red) .

Simple Ratio digunakan dalam berbagai aplikasi seperti pemantauan pertanian, deteksi kekeringan, dan studi perubahan lahan. Indeks ini membantu dalam menentukan tingkat kerapatan vegetasi, kesehatan tanaman, dan tingkat fotosintesis. Menurut (Huete, 1988), SR sangat berguna dalam mengidentifikasi daerah dengan vegetasi yang lebat dan membantu dalam pengelolaan sumber daya alam. Baret & Guyot, (1991), menunjukkan bahwa SR dapat memberikan informasi yang berharga tentang perubahan vegetasi musiman dan tahunan. Liang (2004), menyatakan bahwa SR merupakan metode yang efektif untuk mendeteksi kesehatan vegetasi dan digunakan dalam berbagai studi pemantauan vegetasi. *Simple Ratio* digunakan pada penelitian ini untuk mendorong pengenalan vegetasi agar semakin akurat.

$$SR = \frac{NIR}{RED} \quad (2.1)$$

Dengan :

NIR = Nilai spectral saluran Near Infrared (Band 8)

RED = Nilai spectral saluran Red (Band 4)

2.6. Normalized Difference Vegetation Index (NDVI)

Normalized Difference Vegetation Index (NDVI) merupakan indeks untuk mengukur dan memantau kondisi dari vegetasi menggunakan citra satelit (Rouse et al., 1974). Penggunaan NDVI adalah dengan menghitung menggunakan rasio perbedaan normalisasi antara reflektansi pita near-infrared (NIR) dan pita merah (Red).

NDVI sering digunakan dalam berbagai aplikasi yang berhubungan dengan vegetasi, termasuk pemantauan kesehatan tanaman, deteksi kekeringan, dan analisis perubahan lahan. Indeks ini efektif untuk memprediksi karakteristik permukaan ketika kanopi vegetasi tidak terlalu padat maupun jarang (Liang, 2004).

Krisnayanti & Wiryawan (2018) menunjukkan bahwa NDVI dapat digunakan untuk pemantauan vegetasi di daerah rawan bencana. Dengan memantau kondisi vegetasi secara berkala, pemerintah dan pihak terkait dapat melakukan tindakan pencegahan lebih dini. Ardiansyah & Wibowo (2019) juga menemukan bahwa NDVI efektif dalam mengidentifikasi perubahan tutupan lahan di kawasan rawan banjir, yang membantu dalam pengelolaan risiko bencana. NDVI digunakan juga pada penelitian ini untuk membuat sistem semakin dapat mengerti membedakan vegetasi yang ada ada citra satelit yang akan dilakukan klasifikasi.

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (2.2)$$

Dengan :

NIR = Nilai spectral saluran Near Infrared (Band 8)

RED = Nilai spectral saluran Red (Band 4)

2.7. Normalized Difference Water Index (NDWI)

Normalized Difference Water Index (NDWI) adalah indeks yang digunakan untuk mendeteksi dan memantau keberadaan air permukaan menggunakan citra satelit. Indeks ini pertama kali diperkenalkan oleh McFeeters (1996) dan dirancang untuk

meningkatkan visibilitas badan air pada peta dengan mengurangi pengaruh tanah dan vegetasi.

NDWI banyak digunakan dalam berbagai aplikasi, seperti pemantauan kualitas air, deteksi perubahan luasan badan air, dan manajemen sumber daya air. Indeks ini membantu mengidentifikasi area dengan kandungan air tinggi dan memantau perubahan dalam waktu yang cepat dan efisien. Menurut McFeeters (1996), NDWI sangat efektif dalam meningkatkan kontras antara badan air dan lahan kering serta vegetasi. Xu (2006) menunjukkan bahwa NDWI dapat digunakan untuk memetakan badan air dengan akurasi yang tinggi, bahkan di daerah yang terpengaruh oleh vegetasi.

Pratama & Sugiarto (2018) menyatakan bahwa NDWI dapat digunakan untuk memantau perubahan luasan badan air di wilayah perkotaan dengan baik. Sementara itu, penelitian yang dilakukan oleh Wahyuni & Hadi (2020) menunjukkan bahwa NDWI efektif dalam pemetaan distribusi air di daerah pertanian.

$$NDWI = \frac{GREEN - NIR}{GREEN + NIR} \quad (2.3)$$

Dengan :

NIR = Nilai spectral saluran Near Infrared (Band 8)

GREEN = Nilai spectral saluran Red (Band 3)

2.8. Confusion Matrix

Confusion Matrix adalah alat yang digunakan dalam machine learning dan statistik untuk mengevaluasi kinerja model klasifikasi. Matriks ini menyajikan perbandingan antara hasil klasifikasi yang sebenarnya dan hasil prediksi yang dihasilkan oleh model.

Confusion Matrix terdiri dari empat komponen utama:

1. True Positive (TP): Jumlah kasus yang sebenarnya positif dan diprediksi positif oleh model.
2. True Negative (TN): Jumlah kasus yang sebenarnya negatif dan diprediksi negatif oleh model.
3. False Positive (FP): Jumlah kasus yang sebenarnya negatif tetapi diprediksi positif oleh model.
4. False Negative (FN): Jumlah kasus yang sebenarnya positif tetapi diprediksi negatif oleh model.

Confusion Matrix sangat berguna dalam menganalisis kinerja model klasifikasi tutupan lahan menggunakan citra satelit. Model seperti *Random Forest* dapat digunakan untuk mengklasifikasikan berbagai jenis tutupan lahan, seperti hutan, perkotaan, dan pertanian. Dengan menggunakan Confusion Matrix, kita dapat mengukur seberapa baik model tersebut dalam mengidentifikasi masing-masing kelas tutupan lahan. Penggunaan Confusion Matrix dalam penelitian semacam ini memungkinkan untuk evaluasi yang lebih mendalam dan peningkatan akurasi model melalui teknik-teknik seperti augmentasi data dan tuning parameter (Mishra et al., 2022). Dari matriks ini, berbagai metrik kinerja model dapat dihitung, seperti akurasi, presisi, recall, dan F1-score :

$$Akurasi = \frac{True\ Positives + True\ Negatives}{Total\ Population} \quad (2.4)$$

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (2.5)$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (2.6)$$

$$F1_{ham} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.7)$$

2.9. Website

Website merupakan sekumpulan halaman web yang saling terhubung dan dapat diakses melalui internet. Setiap halaman web dapat memuat berbagai informasi seperti teks, gambar, video, dan yang lainnya. Pengembangan website melibatkan penggunaan teknologi seperti *HTML*, *CSS*, dan *JavaScript* yang memungkinkan interaksi dinamis dan presentasi konten. Desain dan pengembangan website memerlukan pemahaman tentang struktur halaman web serta bagaimana elemen-elemen tersebut bekerja sama untuk memberikan pengalaman pengguna yang optimal (Sharma, 2021).

2.10. HTML (HyperText Markup Language)

HTML (HyperText Markup Language) adalah bahasa markup standar yang digunakan untuk membangun dan menyusun konten di halaman web. HTML memungkinkan pengembang menentukan struktur dasar halaman web dengan menggunakan elemen seperti heading, paragraf, tautan, gambar, dan form. Setiap elemen HTML didefinisikan

oleh tag yang membantu browser web menampilkan konten sesuai keinginan (Sharma & Aakanksha, 2018)

2.11. Bootstrap

Bootstrap merupakan framework front-end yang banyak digunakan untuk mengembangkan website yang responsif dan mobile-first. Bootstrap menawarkan berbagai komponen CSS dan JavaScript yang siap digunakan, termasuk sistem grid, navigasi, tombol, dan modal. Dengan menggunakan Bootstrap, pengembang dapat dengan cepat membuat layout responsif tanpa harus menulis banyak kode CSS dari awal. Framework ini juga mendukung berbagai browser dan perangkat, memastikan tampilan yang konsisten di berbagai platform (Johnson & Brown, 2020).

2.12. JavaScript

JavaScript adalah bahasa pemrograman yang digunakan untuk menciptakan halaman web yang interaktif. Dengan JavaScript, pengembang dapat menambahkan fungsi dinamis ke website, seperti validasi formulir, animasi, dan manipulasi DOM. JavaScript bekerja bersama HTML dan CSS untuk meningkatkan pengalaman pengguna dengan memungkinkan interaksi yang lebih responsif dan kaya fitur. JavaScript juga dapat digunakan di sisi server dengan menggunakan lingkungan runtime seperti Node.js, memungkinkan pengembangan aplikasi web full-stack dengan satu bahasa pemrograman (Lee, 2022).

Keunggulan lain dari JavaScript termasuk kemampuannya untuk meningkatkan keterlibatan pengguna melalui elemen-elemen interaktif seperti menu dropdown, slider gambar, dan efek animasi. Selain itu, JavaScript sangat kompatibel dengan berbagai browser dan platform, menjadikannya bahasa yang fleksibel dan serbaguna untuk pengembangan web (Flanagan, 2020). Dengan JavaScript, pengembang dapat dengan mudah mengimplementasikan fitur-fitur canggih seperti drag-and-drop, carousel, dan modal popup, yang semuanya meningkatkan kegunaan dan daya tarik visual dari website (Brown, 2019).

2.13. Python

Python adalah bahasa pemrograman yang dikenal karena sintaksisnya yang sederhana dan mudah dipelajari. Python sering digunakan dalam pengembangan web, data science, machine learning, dan otomatisasi tugas. Dalam konteks pengembangan web, Python dapat digunakan bersama framework seperti Django dan Flask untuk membangun aplikasi web yang kuat dan dapat diskalakan. Keunggulan Python termasuk ekosistem pustaka yang kaya dan komunitas yang aktif, menjadikannya pilihan populer di kalangan pengembang (O'Reilly, 2021).

Selain itu, Python juga menunjukkan performa yang sangat baik dalam mengolah data. Berbagai pustaka seperti Pandas, NumPy, dan SciPy memungkinkan pengembang untuk melakukan manipulasi data, analisis statistik, dan komputasi ilmiah dengan mudah. Wes McKinney (2017) menunjukkan bahwa Python menjadi pilihan utama untuk analisis data karena kemampuannya untuk menangani dataset besar dan kompleks dengan efisiensi yang tinggi. Penggunaan Python dalam data science telah membantu banyak perusahaan dalam pengambilan keputusan berbasis data yang lebih baik.

Harris et al. (2020) juga menyatakan bahwa Python sangat efisien dalam menangani data skala besar dan kompleks, menjadikannya alat yang sangat efektif dalam analisis data. Penggunaan pustaka seperti NumPy memungkinkan pemrosesan array dengan performa tinggi, yang mendukung berbagai aplikasi ilmiah dan rekayasa. Python dalam data science telah memperkuat kemampuan perusahaan dalam mengambil keputusan yang lebih baik dan tepat waktu berdasarkan analisis data yang mendalam.

Pemanfaatan Python juga meluas ke bidang pengolahan citra satelit. Dengan pustaka seperti Rasterio, GDAL, dan OpenCV, Python memungkinkan pengolahan dan analisis data citra satelit secara efisien. Penelitian terbaru menunjukkan bahwa Python dapat digunakan untuk berbagai aplikasi penginderaan jauh, termasuk klasifikasi tutupan lahan, deteksi perubahan, dan pemantauan lingkungan. Python sangat berguna dalam pengolahan data citra satelit karena kemampuannya untuk mengelola dan menganalisis data besar dengan presisi dan kecepatan tinggi. Keunggulan ini menjadikan Python sebagai alat yang sangat berharga dalam pengolahan data citra satelit (Gorelick et al., 2017).

2.14. Flask

Flask adalah microframework web yang ditulis dalam bahasa Python. Flask dirancang untuk memberikan fleksibilitas kepada pengembang dengan menyediakan kerangka kerja minimalis namun dapat diperluas. Flask mendukung berbagai ekstensi yang dapat menambahkan fungsionalitas seperti autentikasi, ORM (Object-Relational Mapping), dan manajemen formulir. Kelebihan Flask termasuk kemudahan penggunaan dan kemampuan untuk mengembangkan prototipe dengan cepat, menjadikannya pilihan populer untuk pengembangan aplikasi web kecil hingga menengah (Thompson, 2020).

Dalam hal performa, Flask dan Python bekerja sangat baik bersama-sama, memberikan kinerja yang optimal untuk aplikasi web. Kombinasi ini memungkinkan pengembang untuk membangun aplikasi yang efisien dan responsif. Flask menyediakan routing yang cepat dan sederhana serta mendukung berbagai alat pengembang yang kuat, sementara Python menawarkan ekosistem pustaka yang kaya untuk penanganan data, manipulasi string, dan operasi matematis. Menurut Grinberg (2018), penggunaan Flask dengan Python dapat mengurangi waktu pengembangan dan meningkatkan kinerja aplikasi web secara keseluruhan. Gurnani (2021) menambahkan bahwa performa Flask dalam menangani permintaan HTTP dan integrasi dengan pustaka Python lainnya menjadikannya pilihan unggul bagi pengembang yang membutuhkan solusi yang cepat dan ringan.

2.15. Penelitian Terdahulu

Penelitian terkait yang sudah dilakukan sebelumnya dilakukan oleh Maulana (2022). Penelitian ini berjudul Klasifikasi Jenis Lahan Menggunakan CNN Berbasis Citra Satelit. Dataset yang digunakan untuk penelitian tersebut adalah EuroSAT. Model CNN menggunakan dataset dengan rasio pembagian 80% sebagai data latih dan 20% sebagai data uji. Hasil penelitian menunjukkan bahwa Overall Accuracy (OA) terbaik dimiliki oleh ResNet-50 sebesar 96,93% disusul oleh VGG-16 sebesar 95,22%, sedangkan EfficientNet-B0 memiliki akurasi cukup rendah dengan nilai 31,96%.

Penelitian lainnya yang sudah dilakukan oleh Veriana et al. (2018) dengan judul Identifikasi Dan Klasifikasi Tutupan Lahan Melalui Pengolahan Citra Google Earth Dengan Metode Singular Value Decomposition Dan Klasifikasi K-Nearest Neighbor. Dalam penelitian ini, citra satelit yang diambil dari Google Earth digunakan sebagai sampel yang akan diidentifikasi, dengan format JPEG dan ukuran 256x256 pixel.

Metode yang digunakan melibatkan Singular Value Decomposition (SVD) untuk pengolahan citra dan K-Nearest Neighbor (KNN) untuk klasifikasi. Hasil akurasi dari penelitian ini mencapai 87.14%, menunjukkan keberhasilan metode SVD dan KNN dalam mengidentifikasi dan mengklasifikasikan tutupan lahan pada citra satelit tersebut.

Penelitian lainnya dilakukan oleh Afasel et al. (2022) dengan judul Klasifikasi Tutupan Lahan Menggunakan Supervised Machine Learning Pada Citra Satelit Menggunakan Google Earth Engine. Citra satelit yang digunakan adalah citra satelit sentinel-2 dengan resolusi sedang. Klasifikasi tutupan lahan yang diterapkan pada penelitian ini terdiri dari empat kelas yaitu badan air, lahan terbangun, lahan terbuka, dan vegetasi. Tahap pre-processing terdiri dari preparasi data, klasifikasi, dan confusion matrix. Setelah tahap pre-processing selesai maka citra tersebut akan masuk ke dalam tahap klasifikasi. Hasil overall accuracy pada algoritma minimum distance yaitu 96.66%, pada algoritma CART yaitu 91.66%, dan pada algoritma gradient boost 90%.

Zulfajria et al. (2021) melakukan penelitian dengan judul Klasifikasi Tutupan Lahan Data Landsat-8 Oli Menggunakan Metode Random Forest. Penelitian ini menggunakan penginderaan jauh dengan resolusi spasial multispektral 30 m dan resolusi temporal 16 hari pada Kabupaten Pidie. Penentuan sampel lapangan atau training point dilakukan dengan menggunakan metode stratified random sampling dengan strata berupa luas tutupan lahan. Terdapat 86 titik sampel lapangan atau training point yang digunakan untuk survei dan validasi hasil klasifikasi yang tersebar secara random dengan mewakili semua jenis tutupan lahan.. Dengan proporsi 70% untuk data training dan 30% untuk data testing, random forest menghasilkan nilai akurasi keseluruhan sebesar 89,53% dan nilai kappa 0,91 dengan metode confusion matrix.

Marlina (2022) melakukan penelitian yang berjudul Klasifikasi Tutupan Lahan Pada Citra Sentinel-2 Kabupaten Kuningan Dengan NDVI dan Algoritme Random Forest. Penelitian ini menggunakan citra sentinel-2A untuk data yang digunakan. Sentinel 2 memiliki 13 band untuk informasi spektral dan penelitian ini menggunakan adalah bands 2, 3, 4, 5, 6, 7, 8, 8A, 11 dan 12. Hasil klasifikasi tutupan lahan menunjukkan tutupan lahan didominasi oleh pertanian 32,66% seluas 398,98 km² dan vegetasi 38,59% seluas 471,47 km². Hasil dari penelitian yang dilakukan, nilai NDVI yang dihasilkan adalah -0,3 – 0,91 dan akurasi algoritme random forest 91,39%, dan kappa 0,88.

Zhang et al. (2021) melakukan penelitian berjudul *Sentinel-2 Satellite Imagery for Urban Land Cover Classification by Optimized Random Forest Classifier*. Data citra satelit diambil dari satelit Sentinel-2A/B yang melibatkan daerah Beijing, Tiongkok, dengan tujuan untuk mengklasifikasikan tutupan lahan di daerah kota yang mengalami urbanisasi cepat. Hasil penelitian menunjukkan bahwa penggunaan pengklasifikasi Random Forest yang dioptimalkan menghasilkan kinerja terbaik dengan akurasi keseluruhan (OA: 0,879) dan koefisien kappa (0,8210). Sebaliknya, SVM menunjukkan nilai akurasi keseluruhan dan koefisien kappa yang rendah, masing-masing sebesar 0,461 dan 0,2526. Dalam eksperimen perbandingan, terlihat bahwa pengklasifikasi Random Forest yang dioptimalkan mampu mengungguli SVM konvensional dan Random Forest dengan hyperparameter default dalam hal akurasi, presisi, dan perolehan hasil.

(Gunawansa et al., 2022) melakukan penelitian berjudul *Application of Sentinel-2 Satellite Data to Map Forest Cover in Southeast Sri Lanka through the Random Forest Classifier*. Lahan pertanian, hutan termasuk taman nasional Yala, dan desa-desa dengan tanaman tahunan merupakan bagian dari wilayah yang digunakan pada penelitian ini. Data citra satelit diambil dari gambar Google Earth sebagai referensi. Keakuratan klasifikasi dihitung menggunakan confusion matrix (error matrix) melalui 100 titik sampel yang dipilih secara acak. Keakuratan keseluruhan peta tutupan lahan adalah 85%, dengan akurasi 96% untuk identifikasi tutupan hutan.

Tabel 2. 2 Penelitian Terdahulu

No	Nama	Tahun	Metode	Keterangan
1	Afasel, et. al.	2022	<i>Supervised Machine Learning</i>	Penelitian ini melakukan klasifikasi tutupan lahan pada citra satelit menggunakan Google Earth Engine. Sistem dibuat dengan menggunakan metode Supervised Machine Learning dengan metode klasifikasi Minimum Distance, CART, dan Gradient Boost pada daerah Kec. Cihampelas, Kabupaten Bandung Barat, Jawa Barat. Data citra satelit diambil dari satelit Sentinel 2 lalu mengklasifikasi menjadi empat kelas yaitu badan air, lahan terbangun, lahan terbuka, dan vegetasi. Hasil overall accuracy pada algoritma minimum distance yaitu 96.66%, pada algoritma CART yaitu 91.66%, dan pada algoritma gradient boost 90%.
2	Fuad Maulana	2022	<i>Convolutional Neural Network</i>	Penelitian ini melakukan klasifikasi jenis lahan menggunakan citra satelit dengan 3 arsitektur CNN, yaitu VGG-16, ResNet-50, dan EfficientNet-B0. Dataset yang digunakan untuk penelitian tersebut adalah EuroSAT. Model CNN menggunakan dataset dengan rasio pembagian 80% sebagai data latih

Tabel 2.2 Penelitian Terdahulu (lanjutan)

No	Nama	Tahun	Metode	Keterangan
				<p>dan 20% sebagai data uji. Hasil penelitian menunjukkan bahwa Overall Accuracy (OA) terbaik dimiliki oleh ResNet-50 sebesar 96,93% disusul oleh VGG-16 sebesar 95,22%, sedangkan EfficientNet-B0 memiliki akurasi cukup rendah dengan nilai 31,96% dan 20% sebagai data uji. Hasil penelitian menunjukkan bahwa Overall Accuracy (OA) terbaik dimiliki oleh ResNet-50 sebesar 96,93% disusul oleh VGG-16 sebesar 95,22%, sedangkan EfficientNet-B0 memiliki akurasi cukup rendah dengan nilai 31,96%.</p>
3	Dwi Marlina	2022	<i>NDVI</i> dan <i>Random Forest</i> .	<p>Penelitian ini menggunakan citra sentinel-2A untuk data yang digunakan. Hasil klasifikasi tutupan lahan menunjukkan tutupan lahan didominasi oleh pertanian 32,66% seluas 398,98 km² dan vegetasi 38,59% seluas 471,47 km². Hasil dari penelitian yang dilakukan, nilai <i>NDVI</i> yang dihasilkan adalah -0,3 – 0,91 dan akurasi algoritme random forest 91,39%, dan kappa 0,88.</p>

Tabel 2.2 Penelitian Terdahulu (lanjutan)

No	Nama	Tahun	Metode	Keterangan
4	T. Gunawansa, <i>et. al</i>	2022	<i>Random Forest</i>	Penelitian ini melakukan klasifikasi pemetaan tutupan hutan pada daerah Sri Lanka Tenggara. Penelitian ini menggunakan gambar Level-2A Sentinel-2 dengan tutupan awan kurang dari 10 persen digunakan dalam perangkat lunak European Space Agency's (ESA) SNAP versi 8.0.0 untuk pemrosesan gambar dan tutupan hutan di wilayah studi dipetakan melalui pengklasifikasi <i>Random Forest</i> . Data lapangan dikumpulkan melalui referensi data resolusi tinggi Google Earth. Keakuratan klasifikasi dihitung menggunakan <i>confusion matrix</i> (<i>error matrix</i>) melalui 100 titik sampel yang dipilih secara acak. Keakuratan keseluruhan peta tutupan lahan adalah 85%, dengan akurasi 96% untuk identifikasi tutupan hutan.
5	T. Zhang, <i>et. al</i>	2021	<i>Support Vector Machine</i> dan <i>Random Forest</i>	Penelitian ini melakukan klasifikasi tutupan lahan pada perkotaan di Beijing, Tiongkok. Penelitian ini dilakukan guna mengetahui tutupan lahan pada daerah kota dengan urbanisasi yang cepat. Citra satelit yang digunakan adalah citra satelit

Tabel 2.1 Penelitian Terdahulu (lanjutan)

No	Nama	Tahun	Metode	Keterangan
				Sentinel-2A/B. Hasilnya menunjukkan bahwa pengklasifikasi <i>Random Forest</i> yang dioptimalkan mencapai kinerja terbaik dengan akurasi keseluruhan (OA: 0,879), koefisien kappa (0,8210), sedangkan <i>SVM</i> mencapai nilai OA dan kappa yang rendah masing-masing sebesar 0,461 dan 0,2526.
6	Zulfajri <i>et. al.</i>	2021	<i>Random Forest</i>	Penelitian ini menggunakan penginderaan jauh dengan resolusi spasial multispektral 30 m dan resolusi temporal 16 hari pada Kabupaten Pidie. Penentuan sampel lapangan atau <i>training point</i> dilakukan dengan menggunakan metode <i>stratified random sampling</i> dengan strata berupa luas tutupan lahan. Terdapat 86 titik sampel lapangan atau <i>training point</i> yang digunakan untuk survei dan validasi hasil klasifikasi yang tersebar secara random dengan mewakili semua jenis tutupan lahan. Dengan proporsi 70% untuk data training dan 30% untuk data testing, <i>random forest</i> menghasilkan nilai akurasi keseluruhan sebesar

Tabel 2.2 Penelitian Terdahulu (lanjutan)

No	Nama	Tahun	Metode	Keterangan
				89,53% dan nilai kappa 0,91 dengan metode <i>confusion matrix</i> .
7	Veriana, et. al.	2018	<i>Singular Value Decomposition</i> dan <i>K-Nearest Neighbor</i>	Penelitian ini mendentifikasi dan mengklasifikasi tutupan lahan melalui pengolahan citra google earth dengan metode <i>singular value decomposition</i> dan <i>klasifikasi k-nearest neighbor</i> . Citra satelit yang diambil dari Google Earth digunakan sebagai sampel yang akan diidentifikasi. Metode yang digunakan melibatkan <i>Singular Value Decomposition (SVD)</i> untuk pengolahan citra dan KNN untuk klasifikasi. Hasil akurasi dari penelitian ini mencapai 87.14%, menunjukkan keberhasilan metode SVD dan KNN dalam mengidentifikasi dan mengklasifikasikan tutupan lahan pada citra satelit tersebut.

2.16. Perbedaan Penelitian

Penelitian ini memiliki beberapa perbedaan dengan penelitian sebelumnya yaitu penelitian ini akan mengklasifikasi citra menjadi 6 klasifikasi yaitu lahan terbuka, permukiman, pertanian, perairan, vegetasi, dan jalan. Daerah yang akan digunakan sebagai daerah untuk *sample* dan penelitian adalah citra daerah kota medan dengan 5 kecamatan di dalamnya. Lalu, pada penelitian ini juga akan menggunakan citra dari *Sentinel-2* dan mengambil 4 *band* yang memiliki resolusi spasial 10 meter, yaitu Band 2 (Blue), Band 3 (Green), Band 4 (Red), dan Band 8 (Near Infrared). Keempat band

tersebut akan digabung dengan *Simpe Ratio (SR)*, *Normalized difference Vegetation index (NDVI)*, dan *Normalized difference Water index (NDWI)* untuk mendapatkan klasifikasi yang baik untuk vegetasi dan air.

Pada penelitian terdahulu, Gunawansa et al. (2022) ,menggunakan 100 titik sampel dan Zulfajria et al. (2021) sebanyak 87 titik sampel. Penelitian ini akan menggunakan 12.755 titik sampel berbentuk piksel yang nantinya akan dibagi menjadi dengan proporsi 70% untuk data training, 20% untuk data testing, dan 10% untuk data validation. Titik sampel tersebut akan diambil 166 *polygon* sampel yang dipecah menjadi piksel dan digunakan sebagai titik sampel.

BAB 3

ANALISIS DAN PERANCANGAN SISTEM

3.1. Dataset

Penelitian ini menggunakan data citra raster (.tif) yang diambil menggunakan *google earth engine*. Pemilihan daerah untuk citra yang diambil adalah kota Medan dan sekitar. Citra tersebut berasal dari satelit sentinel-2 yang dapat di akses melalui *google earth engine*. Data raster yang diambil adalah dari tahun 2023. Data raster yang dipilih memiliki tingkat persentase awan sebesar 5% agar tidak mengganggu dalam pembelajaran model Random Forest. Data tersebut memiliki koordinat EPSG 32647, yang merupakan kode referensi untuk sistem koordinat *Universal Transverse Mercator (UTM)* zona 47N berbasis *WGS 84 (World Geodetic System 1984)*. Sistem koordinat ini digunakan untuk memproyeksikan peta bumi dalam zona UTM tertentu, yang memungkinkan pengukuran dan representasi posisi geografis dengan akurasi tinggi serta meminimalkan distorsi di area yang lebih kecil, seperti kota Medan dan sekitarnya (Spatial Reference, 2020).

Citra yang digunakan mengambil 4 *band* dari 12 *band* yang dimiliki citra satelit sentinel-2 yaitu Band 2 (*Blue*), Band 3 (*Green*), Band 4 (*Red*), dan Band 8 (*Near Infrared*), karena memiliki resolusi spasial 10 meter.



Gambar 3. 1 Citra raster daerah penelitian

Penentuan *Training Point* atau titik sampel (.shp) ditentukan berdasarkan prinsip interpretasi citra. Prinsip tersebut adalah teknik untuk mengenali dan memahami ciri-

ciri pada gambar dari satelit atau udara. Teknik ini melibatkan banyak elemen kunci, seperti tampilan, dimensi, desain, siluet, nada atau hue, tekstur, lokasi, dan hubungan antara objek dalam gambar. Dengan menggunakan prinsip-prinsip tersebut, analis dapat mengidentifikasi serta mengelompokkan elemen-elemen di permukaan bumi secara lebih tepat.

Titik sample (*training point*) yang telah ditentukan memuat 166 label berbentuk polygon. Data tersebut terdiri dari 13 polygon vegetasi, 11 polygon permukiman, 76 polygon perairan, 6 polygon lahan terbuka, 15 polygon pertanian, dan 45 polygon jalan. Pada tabel 3.1 menunjukkan berbagai kategori penggunaan lahan yang telah diberi label tertentu. Label 0 mewakili vegetasi, yang mencakup area dengan tutupan tanaman hijau yang lebat. Label 1 menunjukkan area permukiman, yaitu kawasan yang dipenuhi oleh bangunan dan infrastruktur lainnya. Label 2 digunakan untuk mengidentifikasi perairan, seperti danau, sungai, dan waduk. Label 3 mewakili lahan terbuka, yang terdiri dari area tanpa tutupan vegetasi atau bangunan. Label 4 merujuk pada area pertanian, di mana aktivitas bercocok tanam dilakukan. Terakhir, label 5 menunjukkan jalan, yang mencakup semua jenis infrastruktur jalan.

Tabel 3. 1 Dataset (.shp)

Label land cover	geometry
0	Polygon ((98.7124329116096 3.6033984430442842,...
0	Polygon ((98.72463746850639 3.6007681908134828,...
1	Polygon ((98.70109933882593 3.5955428805569047,...
1	Polygon ((98.6827411308841 3.5898137176692444,...
2	Polygon ((98.65220608809989 3.6210753139556573,...
2	Polygon ((98.72345676707806 3.5769534301625185,...
3	Polygon ((98.67025410937022 3.588402868099871,...
3	Polygon ((98.70850040496003 3.6015293621952083,...
4	Polygon ((98.642858770645 3.565292456691526, 9...
4	Polygon ((98.61961380292539 3.6107794348761026,...
5	Polygon ((98.68884802888329 3.58839510655514, ...
5	Polygon ((98.68335217647012 3.5862764109390355,...

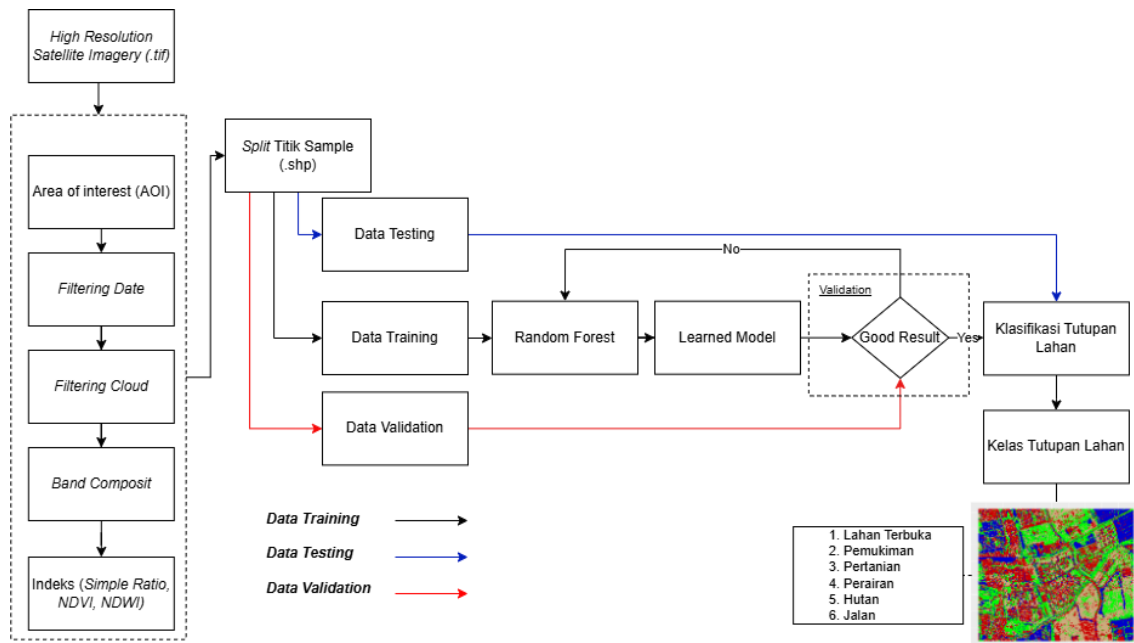
Setiap polygon memiliki ukuran luas yang berbeda-beda sehingga data polygon tersebut kemudian diubah menjadi bentuk piksel dan diperoleh 12.755 titik pelatihan. Perincian dari piksel tersebut adalah 2.230 piksel vegetasi, 2.178 piksel permukiman, 2.052 piksel perairan, 2.025 piksel lahan terbuka, 2.181 piksel pertanian, dan 2.089 piksel jalan.

Tabel 3. 2 Dataset dalam piksel

Kelas	Piksel
vegetasi	2.230
pertanian	2.181
permukiman	2.052
jalan	2.089
perairan	2.052
lahanterbuka	2.025

3.2. Arsitektur Umum

Terdapat tahapan-tahapan yang dilakukan dalam penelitian ini yaitu, mengambil data citra raster yang akan digunakan dan data citra raster akan masuk tahap *pre-processing* yang terdiri dari *area of interest (AOI)*, *filtering date*, *filtering cloud*, dan *band composit*. Kemudian, data tersebut akan dihitung dengan index *Simple Ratio*, *NDVI*, dan *NDWI* untuk membantu dalam mengenali tutupan lahan vegetasi dan air. Setelah itu, titik sampel akan digabung dengan citra raster dan data *training point* akan dibagi menjadi data *training*, *testing*, dan *validation*. Berikutnya, algoritma *Random Forest* akan melakukan *training* dengan data titik sampel. Hasil dari model yang sudah dilatih dengan baik akan diuji dengan *data testing*. Setelah itu, model akan melakukan prediksi terhadap piksel lain pada citra raster dengan mengklasifikasikan tutupan lahan menjadi 6 jenis yaitu, vegetasi, permukiman, perairan, lahan terbuka, pertanian, dan jalan. Arsitektur umum dalam penelitian ini digambarkan dalam Gambar 3.2 sebagai berikut.



Gambar 3. 2 Arsitektur Umum

3.2.1. Pengambilan data citra satelit

Pada tahap awal, dilakukan pengambilan data berupa gambar citra raster satelit beresolusi tinggi yang di unduh dari *Google Earth Engine* (<https://code.earthengine.google.com/>). Data yang diambil adalah data citra dari satelit Sentinel-2 yang memiliki resolusi spasial 10 meter. Pemilihan daerah untuk citra yang diambil adalah kota Medan dan sekitar.

3.2.2. Pre-Processing

Tahap ini dilakukan dengan tujuan untuk mengolah data citra satelit agar bisa digunakan dalam meng-klasifikasikan tutupan lahan nantinya. Data citra tersebut akan melalui tahap-tahap *pre-processing* sebagai berikut.

3.2.2.1. Area of Interest (AOI)

Area of Interest (AOI) merupakan kawasan geografis spesifik yang menjadi fokus penelitian. Tahap ini akan menentukan data citra yang akan digunakan dalam penelitian yang dilakukan. AOI dipilih berdasarkan tujuan penelitian, ketersediaan data, dan karakteristik wilayah yang relevan dengan studi. Area yang dipilih pada penelitian ini adalah daerah sekita kota Medan. Hal ini dikarenakan peneliti tinggal di wilayah kota

Medan sehingga lebih mudah dalam menganalisis tutupan lahan yang ada di kota tersebut. Area sekitar kota Medan juga mencakup wilayah yang relevan dengan tujuan penelitian. Dimana pada kota Medan terdapat 6 jenis tutupan lahan yang akan diteliti dan diklasifikasikan nantinya. Berikut pseudocode dari proses *AOI* seperti *pseudocode* 3.1

```

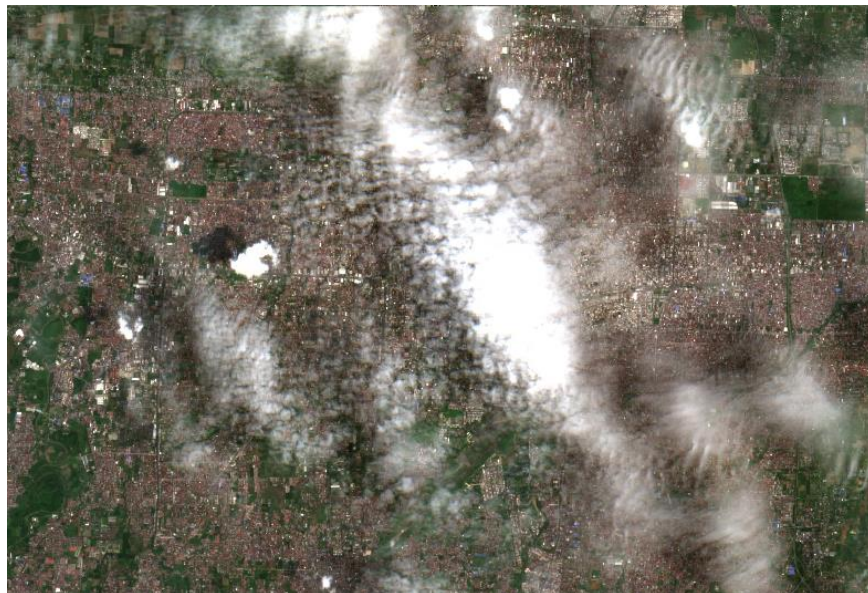
FUNCTION selectAOI(dataset, AOI):
    # Inisialisasi dataset yang difilter
    aoi_dataset = []

    # Iterasi melalui setiap citra dalam dataset
    FOR each image IN dataset:
        # Periksa apakah citra berada dalam AOI
        IF image.bounds INTERSECTS AOI:
            # Tambahkan citra ke dataset yang difilter
            ADD image TO aoi_dataset
        END IF
    END FOR

    RETURN aoi_dataset
END FUNCTION

```

Pseudocode 3. 1 *Area of Interest*



Gambar 3. 3 Implementasi tahap *area of interest*

3.2.2.2. *Filtering Date*

Tahap ini merupakan proses dalam pengolahan citra satelit di mana citra yang diambil dalam rentang waktu tertentu dipilih untuk analisis lebih lanjut. Tujuan utama dari tahap ini adalah untuk memastikan bahwa data yang digunakan sesuai dengan periode waktu yang relevan dengan penelitian yang dilakukan dan juga melihat ketersediaan datanya. Pada penelitian ini, rentang waktu yang ditentukan adalah pada tahun 2023. Berikut pseudocode dari proses *filtering date* seperti pada pseudocode 3.2

```

FUNCTION filteringDate(dataset, start_date, end_date):
    # Inisialisasi dataset yang difilter
    filtered_dataset = []

    # Iterasi melalui setiap citra dalam dataset
    FOR each image IN dataset:
        # Periksa tanggal citra
        IF image.date >= start_date AND image.date <=
end_date:
            # Tambahkan citra ke dataset yang difilter
            ADD image TO filtered_dataset
        END IF
    END FOR

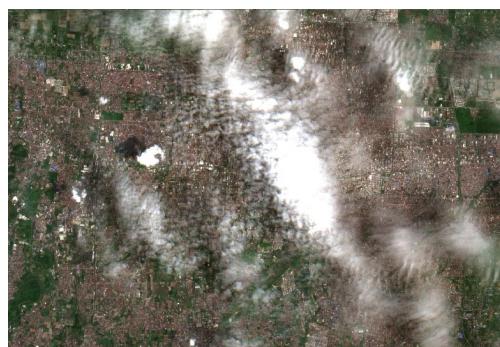
    RETURN filtered_dataset
END FUNCTION

```

Pseudocode 3. 2 *Filtering Date*



Gambar 3. 4 Citra sebelum *filtering date*



Gambar 3. 5 Citra setelah *filtering date*

3.2.2.3. *Filtering Cloud*

Tahap ini akan memilah data citra dengan persentase awan yang rendah dipilih untuk diteliti lebih lanjut. Tujuan utama dari tahap ini adalah untuk memastikan bahwa data yang digunakan memiliki kualitas yang baik dan minim gangguan dari tutupan awan, sehingga klasifikasi menjadi lebih akurat. Pada penelitian ini, persentase awan yang dipilih ada 5%. Berikut pseudocode dari proses *filtering data* seperti pada pseudocode 3.3

```
FUNCTION filteringCloud(dataset, cloud_percentage):
    # Inisialisasi dataset yang difilter
    filtered_dataset = []

    # Iterasi melalui setiap citra dalam dataset
    FOR each image IN dataset:
        # Periksa persentase awan pada citra
        IF image.cloud_coverage < cloud_percentage:
            # Tambahkan citra ke dataset yang difilter
            ADD image TO filtered_dataset
        END IF
    END FOR

    RETURN filtered_dataset
END FUNCTION
```

Pseudocode 3. 3 *Filtering Cloud*



Gambar 3. 6 Implementasi *Filtering Cloud*

3.2.2.4. Band Composit

Tahap ini adalah proses dalam pengolahan citra satelit di mana beberapa band spektral dari citra akan digabungkan sehingga menciptakan representasi visual. Pada penelitian ini, digunakan Band 2 (*Blue*), Band 3 (*Green*), Band 4 (*Red*), dan Band 8 (*Near Infrared*) dari citra Sentinel-2, yang memiliki resolusi spasial 10 meter. Kombinasi band ini memungkinkan visualisasi yang lebih baik.

Tabel 3. 3 Band Composit Citra Raster

No	Band 2	Band 3	Band 4	Band 8
1	0.21200	0.21040	0.21640	0.20400
2	0.12460	0.12120	0.11520	0.11060
3	0.17540	0.18140	0.19520	0.17920
4	0.11500	0.12000	0.11020	0.10180
5	0.36600	0.35000	0.31360	0.25980
6	0.05620	0.07280	0.13240	0.19320
7	0.07940	0.08880	0.09560	0.18540
8	0.09720	0.11880	0.12060	0.15240
9	0.02410	0.06360	0.02950	0.35280
10	0.10500	0.13160	0.15040	0.17100

Berikut merupakan pseudocode dari proses *filtering data* seperti pada pseudocode 3.4

```

FUNCTION createBandComposite(dataset):
    # Inisialisasi dataset komposit
    composite_dataset = []

    # Iterasi melalui setiap citra dalam dataset
    FOR each image IN dataset:
        # Ekstrak band yang diperlukan
        blue_band = image['Band_2']
        green_band = image['Band_3']
        red_band = image['Band_4']
        nir_band = image['Band_8']

        # Gabungkan band menjadi komposit
        composite_image = COMBINE(blue_band, green_band,
red_band, nir_band)

        # Tambahkan citra komposit ke dataset komposit
        ADD composite_image TO composite_dataset

```

```

END FOR

RETURN composite_dataset
END FUNCTION

```

Pseudocode 3. 4 Band Composit

3.2.2.5. Indeks (*Simple Ratio*, *NDVI*, dan *NDWI*)

Indeks *Simple Ratio*, *NDVI*, dan *NDWI* digunakan untuk membantu dalam mengidentifikasi dan mengklasifikasikan berbagai jenis tutupan lahan, khususnya vegetasi dan perairan. Perhitungan indeks akan dilakukan kepada setiap piksel dalam citra satelit menggunakan band yang relevan. Rumus dari perhitungan indeks yang akan dilakukan seperti yang diuraikan pada bab 2.

Indeks yang dihitung akan digabung ke dalam citra raster bersama dengan 4 *band* yang sudah ditentukan sebelumnya. Dengan adanya nilai indeks, model klasifikasi dapat lebih akurat dalam membedakan antara area vegetasi, perairan, dan jenis tutupan lahan lainnya.

Tabel 3. 4 Band Composit beserta Perhitungan Indeks

No	Band 2	Band 3	Band 4	Band 8	SR	NDVI	NDWI
1	0.212	0.2104	0.2164	0.204	0.9427	-0.0295	-0.01544
2	0.1246	0.1212	0.1152	0.1106	0.96007	-0.02037	-0.04573
3	0.1754	0.1814	0.1952	0.1792	0.91803	-0.04274	-0.0061
4	0.115	0.12	0.1102	0.1018	0.92377	-0.03962	-0.08206
5	0.366	0.35	0.3136	0.2598	0.82844	-0.09383	-0.14792
6	0.0562	0.0728	0.1324	0.1932	1.45921	0.18673	0.45263
7	0.0794	0.0888	0.0956	0.1854	1.93933	0.31957	0.3523
8	0.0972	0.1188	0.1206	0.1524	1.26368	0.11648	0.12389
9	0.0241	0.0636	0.0295	0.3528	11.95932	0.84567	0.69452
10	0.105	0.1316	0.1504	0.171	1.13697	0.06409	0.1302

Berikut adalah detail perhitungan indeks yang dilakukan dengan rumus *Simple Ratio*, *NDVI*, dan *NDWI*.

I. Baris 1

$$SR = \text{Band 8} / \text{Band 4} = 0.204 / 0.2164 = 0.94270$$

$$\text{NDVI} = (\text{Band 8} - \text{Band 4}) / (\text{Band 8} + \text{Band 4}) = (0.204 - 0.2164) / (0.204 + 0.2164) = -0.02950$$

$$\text{NDWI} = (\text{Band 3} - \text{Band 8}) / (\text{Band 3} + \text{Band 8}) = (0.2104 - 0.204) / (0.2104 + 0.204) = -0.01544$$

II. Baris 2

$$\text{SR} = \text{Band 8} / \text{Band 4} = 0.1106 / 0.1152 = 0.96007$$

$$\text{NDVI} = (\text{Band 8} - \text{Band 4}) / (\text{Band 8} + \text{Band 4}) = (0.1106 - 0.1152) / (0.1106 + 0.1152) = -0.02037$$

$$\text{NDWI} = (\text{Band 3} - \text{Band 8}) / (\text{Band 3} + \text{Band 8}) = (0.1212 - 0.1106) / (0.1212 + 0.1106) = -0.04573$$

III. Baris 3

$$\text{SR} = \text{Band 8} / \text{Band 4} = 0.1792 / 0.1952 = 0.91803$$

$$\text{NDVI} = (\text{Band 8} - \text{Band 4}) / (\text{Band 8} + \text{Band 4}) = (0.1792 - 0.1952) / (0.1792 + 0.1952) = -0.04274$$

$$\text{NDWI} = (\text{Band 3} - \text{Band 8}) / (\text{Band 3} + \text{Band 8}) = (0.1814 - 0.1792) / (0.1814 + 0.1792) = -0.00610$$

IV. Baris 4

$$\text{SR} = \text{Band 8} / \text{Band 4} = 0.1018 / 0.1102 = 0.92377$$

$$\text{NDVI} = (\text{Band 8} - \text{Band 4}) / (\text{Band 8} + \text{Band 4}) = (0.1018 - 0.1102) / (0.1018 + 0.1102) = -0.03962$$

$$\text{NDWI} = (\text{Band 3} - \text{Band 8}) / (\text{Band 3} + \text{Band 8}) = (0.12 - 0.1018) / (0.12 + 0.1018) = -0.08206$$

V. Baris 5

$$\text{SR} = \text{Band 8} / \text{Band 4} = 0.2598 / 0.3136 = 0.82844$$

$$\text{NDVI} = (\text{Band 8} - \text{Band 4}) / (\text{Band 8} + \text{Band 4}) = (0.2598 - 0.3136) / (0.2598 + 0.3136) = -0.09383$$

Normalized Difference Water Index (NDWI):

$$\text{NDWI} = (\text{Band 3} - \text{Band 8}) / (\text{Band 3} + \text{Band 8}) = (0.35 - 0.2598) / (0.35 + 0.2598) = -0.14792$$

VI. Baris 6

$$\text{SR} = \text{Band 8} / \text{Band 4} = 0.1932 / 0.1324 = 1.45921$$

$$\text{NDVI} = (\text{Band 8} - \text{Band 4}) / (\text{Band 8} + \text{Band 4}) = (0.1932 - 0.1324) / (0.1932 + 0.1324) = 0.18673$$

$$NDWI = (Band\ 3 - Band\ 8) / (Band\ 3 + Band\ 8) = (0.0728 - 0.1932) / (0.0728 + 0.1932) = 0.45263$$

VII. Baris 7

$$SR = Band\ 8 / Band\ 4 = 0.1854 / 0.0956 = 1.93933$$

$$NDVI = (Band\ 8 - Band\ 4) / (Band\ 8 + Band\ 4) = (0.1854 - 0.0956) / (0.1854 + 0.0956) = 0.31957$$

$$NDWI = (Band\ 3 - Band\ 8) / (Band\ 3 + Band\ 8) = (0.0888 - 0.1854) / (0.0888 + 0.1854) = 0.35230$$

VIII. Baris 8

$$SR = Band\ 8 / Band\ 4 = 0.1524 / 0.1206 = 1.26368$$

$$NDVI = (Band\ 8 - Band\ 4) / (Band\ 8 + Band\ 4) = (0.1524 - 0.1206) / (0.1524 + 0.1206) = 0.11648$$

$$NDWI = (Band\ 3 - Band\ 8) / (Band\ 3 + Band\ 8) = (0.1188 - 0.1524) / (0.1188 + 0.1524) = 0.12389$$

IX. Baris 9

$$SR = Band\ 8 / Band\ 4 = 0.3528 / 0.0295 = 11.95932$$

$$NDVI = (Band\ 8 - Band\ 4) / (Band\ 8 + Band\ 4) = (0.3528 - 0.0295) / (0.3528 + 0.0295) = 0.84567$$

$$NDWI = (Band\ 3 - Band\ 8) / (Band\ 3 + Band\ 8) = (0.0636 - 0.3528) / (0.0636 + 0.3528) = 0.69452$$

X. Baris 10

$$SR = Band\ 8 / Band\ 4 = 0.171 / 0.1504 = 1.13697$$

$$NDVI = (Band\ 8 - Band\ 4) / (Band\ 8 + Band\ 4) = (0.171 - 0.1504) / (0.171 + 0.1504) = 0.06409$$

$$NDWI = (Band\ 3 - Band\ 8) / (Band\ 3 + Band\ 8) = (0.1316 - 0.171) / (0.1316 + 0.171) = 0.13020$$

Berikut merupakan pseudocode dari proses index (*Simple Ration*, *NDVI*, *NDWI*) yang dilakukan seperti pada pseudocode 3.5

```
FUNCTION calculate_SR(band_8, band_4) :
    // Rumus Simple Ratio (SR)
    sr = band_8 / band_4
    RETURN sr
```

```

END FUNCTION

FUNCTION calculate_NDVI(band_8, band_4):
    // Rumus Normalized Difference Vegetation Index (NDVI)
    ndvi = (band_8 - band_4) / (band_8 + band_4)
    RETURN ndvi
END FUNCTION

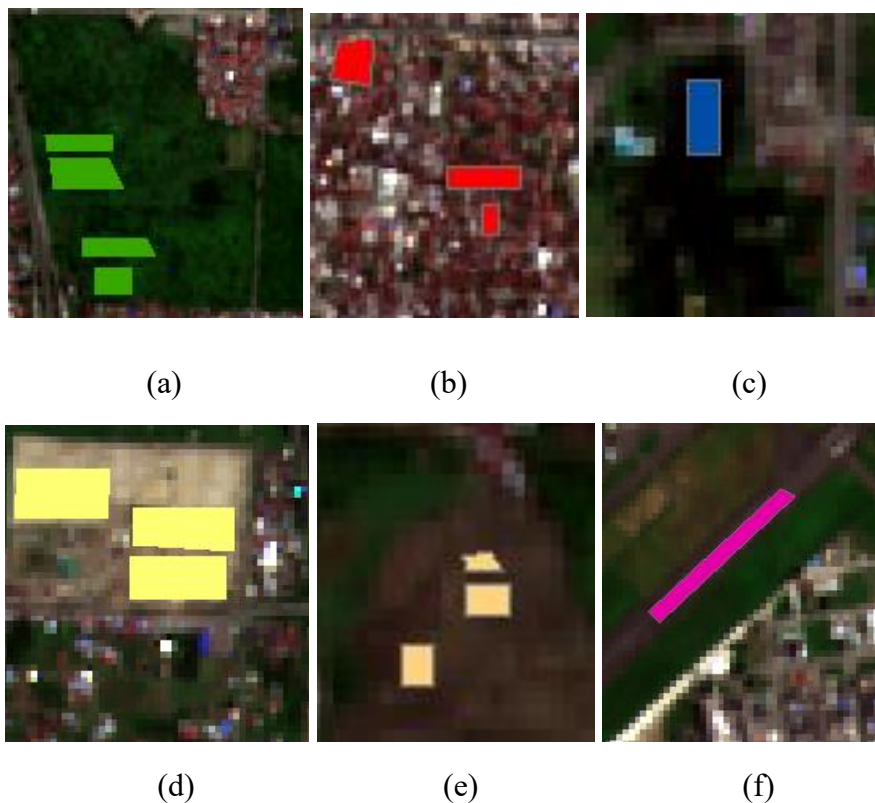
FUNCTION calculate_NDWI(band_3, band_8):
    // Rumus Normalized Difference Water Index (NDWI)
    ndwi = (band_3 - band_8) / (band_3 + band_8)
    RETURN ndwi
END FUNCTION

```

Pseudocode 3. 5 *Index (Simple Ratio, NDVI, dan NDWI)*

3.2.3 Titik Sampel

Titik sampel atau *Training Point* dimuat dalam ekstensi ShapeFile (.shp). Pengambilan titik sampel ini berdasarkan prinsip interpretasi citra. Titik sampel terdiri dari 166 polygon dan diubah menjadi 12.755 piksel sebagai titik sampel.



Gambar 3. 7 Titik sampel (a) vegetasi, (b) permukiman, (c) perairan, (d) lahan terbuka, (e) pertanian, (f) jalan

Perincian dari piksel tersebut adalah 2.230 piksel vegetasi, 2.178 piksel permukiman, 2.052 piksel perairan, 2.025 piksel lahan terbuka, 2.181 piksel pertanian, dan 2.089 piksel jalan. Selanjutnya, data piksel tadi dibagi dengan rasio 70% *data training*, 20% *data testing*, dan 10% *data validation*.

Tabel 3. 5 Split Titik Sampel

Kelas	<i>Training</i>	<i>Testing</i>	<i>Validation</i>
vegetasi	1.561	446	223
pertanian	1.527	436	218
permukiman	1.525	435	218
jalan	1.462	418	209
perairan	1.436	411	205
lahanterbuka	1.417	405	203

3.2.4. Random Forest Algorithm

Dalam penelitian ini terdapat data citra raster (.tif) dan titik sample (.shp). Data citra raster (.tif) merupakan representasi gambar yang terdiri dari piksel. Setiap piksel memiliki nilai yang mewakili informasi tertentu, seperti reflektansi spektral pada berbagai panjang gelombang. Titik sampel (.shp) mengandung informasi geografis dan kelas target yang digunakan sebagai label untuk melatih model. Setiap titik sampel dalam shapefile akan diekstrak nilai dari setiap band raster pada lokasi titik tersebut lalu di gabungkan nilai-nilai ini dengan label kelas dari titik sampel untuk membentuk dataset pelatihan.

Setelah dataset pelatihan terbentuk, model *Random Forest* akan digunakan untuk melakukan klasifikasi terhadap piksel yang ada pada titik sampel atau shapefile. Proses dalam pembentukan pohon pada *Random Forest* dimulai dari *Bootstrap Sampling*. Proses ini merupakan metode untuk menghasilkan beberapa subset dari dataset asli dengan penggantian (*replacement*). Ini berarti bahwa beberapa contoh data mungkin muncul lebih dari sekali dalam satu subset, sementara beberapa contoh mungkin tidak muncul sama sekali. Setelah itu akan dilakukan pemilihan fitur acak dimana pada setiap node dalam pohon keputusan yang dibangun, subset acak dari fitur-fitur yang tersedia dipilih untuk menentukan pemisahan terbaik. Pohon keputusan kemudian dibentuk berdasarkan subset data dan subset fitur yang berbeda-beda ini.

Setelah semua pohon dalam hutan acak dilatih menggunakan subset data dan subset fitur yang berbeda, hasil prediksi dari setiap pohon digabungkan menggunakan voting mayoritas untuk menentukan kelas akhir dari setiap piksel. Voting mayoritas ini memastikan bahwa keputusan akhir lebih tahan terhadap variasi dalam data dan lebih akurat dibandingkan dengan hanya menggunakan satu pohon keputusan.

Untuk lebih mengoptimalkan model, *Random Forest* akan menggunakan *RandomizedSearchCV* untuk mencari hyperparameter terbaik. Dengan *RandomizedSearchCV*, kita dapat mengoptimalkan parameter seperti jumlah pohon dalam hutan (*n_estimators*), kedalaman maksimum pohon (*max_depth*), dan jumlah fitur yang dipertimbangkan untuk pemisahan terbaik (*max_features*).

Dengan menggunakan *RandomizedSearchCV*, model dapat di-tuning untuk mendapatkan set parameter terbaik yang memaksimalkan kinerja prediktifnya pada data validasi. Setelah model terbaik ditemukan, model tersebut dapat digunakan untuk mengklasifikasikan piksel pada data raster yang lebih luas, memberikan peta klasifikasi tutupan lahan yang akurat berdasarkan titik sampel yang telah dilabeli.

3.2.5. Learned Model

Setelah melalui proses pelatihan, model *Random Forest* yang telah dibangun dan dilatih pada data pelatihan dapat dianggap sebagai model yang dipelajari (*learned model*). Model ini terdiri dari sejumlah pohon keputusan yang masing-masing dilatih menggunakan subset acak dari data pelatihan dan fitur. Dengan menggabungkan prediksi dari setiap pohon menggunakan voting mayoritas, model *Random Forest* mampu menghasilkan prediksi pada piksel lainnya. Model disimpan dengan nama *random_forest_model.pkl*. Penyimpanan model ini memungkinkan kita untuk memanfaatkan hasil pelatihan yang telah dilakukan tanpa perlu melatih model dari awal setiap kali kita ingin melakukan prediksi.

```
FUNCTION      trainAndEvaluateRandomForest(train_data,
test_data, validation_data, columns):
    # Inisialisasi data train, test, dan validation
    X_train = SELECT features FROM train_data EXCLUDING
last column
    y_train = SELECT labels FROM train_data LAST column
    X_test = SELECT features FROM test_data EXCLUDING
last column
    y_test = SELECT labels FROM test_data LAST column
```



```

X_validation = SELECT features FROM validation_data
EXCLUDING last column
y_validation = SELECT labels FROM validation_data
LAST column

# Definisikan parameter grid untuk RandomizedSearch
param_dist = {
    'n_estimators': [50, 100, 200],
    'max_features': ['sqrt', 'log2', None], # None
akan mengambil semua fitur
    'max_depth': [None, 10, 20, 30],
    'criterion': ['gini', 'entropy']
}
# Siapkan RandomizedSearchCV
random_search = INITIALIZE RandomizedSearchCV:
    estimator                                     =
RandomForestClassifier(random_state=42)
    param_distributions = param_dist
    n_iter = 50 # Jumlah parameter settings yang
diambil secara acak
    cv = 5 # 5-fold cross-validation
    n_jobs = 1 # Nonaktifkan parallel processing
    random_state = 42
    verbose = 2
    scoring = 'accuracy' # Metode scoring yang
digunakan

# Lakukan pencarian hyperparameter dengan data
pelatihan
random_search.FIT(X_train, y_train)

# Dapatkan model terbaik dari hasil pencarian
best_classifier = random_search.BEST_ESTIMATOR_

# Simpan model terbaik yang telah dilatih ke dalam
file
model_filename = 'best_random_forest_model.pkl'
OPEN file model_filename FOR writing
WRITE best_classifier TO file USING pickle
CLOSE file

# Muat model terbaik yang telah dilatih dari file
OPEN file model_filename FOR reading
loaded_classifier = READ FROM file USING pickle
CLOSE file

# Prediksi data pengujian menggunakan model yang
telah dimuat

```

```

y_test_pred = loaded_classifier.PREDICT(X_test)

# Prediksi data validasi menggunakan model yang telah
dimuat
y_validation_pred =
loaded_classifier.PREDICT(X_validation)

# Evaluasi akurasi pada data pengujian
test_accuracy = CALCULATE accuracy_score(y_test,
y_test_pred)
PRINT 'Testing overall accuracy RF: ' + test_accuracy

# Evaluasi akurasi pada data validasi
validation_accuracy = CALCULATE
accuracy_score(y_validation, y_validation_pred)
PRINT 'Validation overall accuracy RF: ' +
validation_accuracy

END FUNCTION

```

Pseudocode 3. 6 *Learned Model*

3.2.6. Klasifikasi Tutupan Lahan

Dari model yang sudah bagus maka selanjutnya model tersebut akan memprediksi piksel yang berada pada citra raster. Hasil yang ditampilkan adalah peta dengan keterangan warna tutupan lahan. Dimana warna hijau menandakan vegetasi, warna merah menandakan permukiman, warna biru menandakan perairan, warna coklat menandakan lahan terbuka, warna kuning menandakan pertanian, dan warna hitam menandakan jalan.

```

FUNCTION displayLandCoverPrediction(img, classifier,
cmap):
    # Ubah nilai NaN dalam citra menjadi nol
    predict_img = NAN_TO_NUM(img)

    # Ubah bentuk citra untuk prediksi
    predict_img = RESHAPE(predict_img, (img.shape[0] *
img.shape[1], img.shape[2]))

    # Prediksi menggunakan model yang telah dilatih
    array_pred = classifier.PREDICT(predict_img)

    # Kembalikan bentuk prediksi ke bentuk citra asli
    array_pred = RESHAPE(array_pred, (img.shape[0],
img.shape[1]))

```

```

# Cetak bentuk citra prediksi (opsional)
PRINT(predict_img.shape)

# Buat plot untuk menampilkan peta tutupan lahan
CREATE_FIGURE(size=(20, 20))
SHOW_IMAGE(array_pred, cmap)
TURN_OFF_AXIS

# Definisikan nama kelas
class_names = {
    0: 'vegetasi',
    1: 'permukiman',
    2: 'perairan',
    3: 'lahanterbuka',
    4: 'pertanian',
    5: 'jalan'
}

END FUNCTION

```

Pseudocode 3. 7 Klasifikasi Tutupan Lahan

3.2.7. Evaluasi

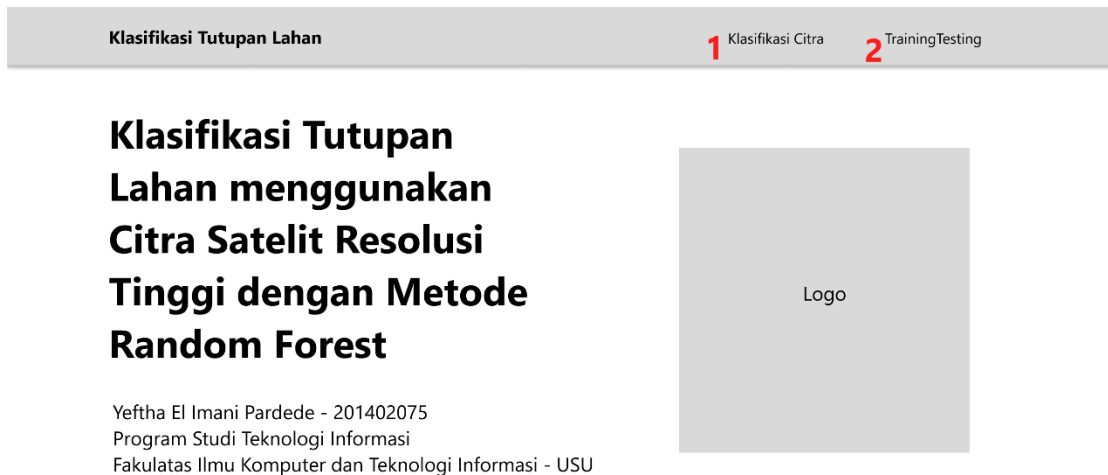
Setelah melatih model Random Forest dan melakukan prediksi pada data pengujian serta validasi, langkah selanjutnya adalah mengevaluasi kinerja model menggunakan *confusion matrix*. Matrix akan menampilkan jumlah prediksi yang benar dan salah untuk setiap kelas, memungkinkan kita untuk melihat kesalahan spesifik yang dibuat oleh model. Dengan analisis *confusion matrix*, kita dapat menghitung metrik evaluasi lainnya seperti akurasi, presisi, *recall*, dan *F1-score*, yang memberikan kinerja model.

3.3. Perancangan Sistem

Tahap ini akan menjelaskan rancangan antarmuka dari sistem untuk klasifikasi tutupan lahan. Sistem akan dikembangkan dalam bentuk website. Dengan sistem ini pengguna dapat memasukkan file citra raster dan gambar peta akan terklasifikasi.

3.3.1 Desain Tampilan Halaman Depan

Pada rancangan tampilan depan terdapat informasi mengenai judul penelitian, peneliti, *navigation bar* yang menghubungkan halaman untuk training-testing dan klasifikasi citra. Berikut rancangan tampilan halaman depan sistem pada web dalam gambar 3.8.



Gambar 3. 8 Halaman Depan

Keterangan:

1. Fungsi label 1 adalah untuk masuk ke halaman klasifikasi citra yang nanti nya pengguna akan memasukkan file raster (.tif).
2. Fungsi label 2 adalah untuk masuk ke halaman training-testing dimana data akan dilatih dan diuji.

3.3.2 Desain Tampilan Training-Testing Data

Pada halaman ini terdapat *input file field* untuk input file citra raster (.tif) dan juga *shapefile* (.shp). Dalam memasukkan file *shapefile*, terdapat file lainnya yang diperlukan yaitu file *database file* (.dbf) dan file *shape file index* (.shx). Ketiga file ini bekerja bersama untuk menyediakan representasi lengkap dari data citra raster geospasial vektor. Proses training dan testing akan berjalan dan hasilnya akan ditampilkan di halaman ini. Berikut rancangan tampilan training-testing data pada web dalam gambar 3.9.

Klasifikasi Tutupan Lahan	Klasifikasi Citra	TrainingTesting
---------------------------	-------------------	-----------------

Training dan Testing Data

Upload Files

- 1 Input raster file TIF

Choose File
- 2 Input Train files (shp, shx, dbf)

Choose File
- 3 Upload
- 4

Tabel Metadata Citra
- 5 Raster Image
- 6 Raster dan SHP Image
- 7

Tabel Split Sampel
- 8 Prediction Image
- 9

Persentase Tutupan Lahan
- 10 Hasil Klasifikasi Random Forest

Confusion Matrix

Classification Report

Gambar 3. 9 Halaman Training-Testing

Keterangan:

1. Fungsi label 1 adalah *input file field* untuk memasukkan file citra raster dengan ekstensi (.tif).
2. Fungsi label 2 adalah *input file field* untuk memasukkan multi file train dengan ekstensi (.shp, .shx, dan .dbx).
3. Label 3 merupakan tombol upload untuk memulai proses training-testing pada file yang sudah di upload.
4. Label 4 merupakan table yang akan menampilkan metadata dari citra raster.

5. Label 5 merupakan gambar dari citra raster.
6. Label 6 merupakan gambar dari citra raster yang digabung dengan file train.
7. Label 7 akan menampilkan table dari pembagian sampel data.
8. Label 8 merupakan gambar citra *testing* yang diprediksi dengan model yang diperoleh dari *training*.
9. Label 9 akan menampilkan persentase dari setiap jenis tutupan lahan yang telah diklasifikasikan.
10. Label 10 akan menampilkan *validation report* dan *confusion matrix* dari model yang ada.

3.3.3. Desain Tampilan Halaman Klasifikasi Citra

Halaman klasifikasi citra merupakan tampilan yang akan menerima data dari citra raster dari pengguna. Halaman ini berisi *input file field* untuk file citra raster (.tif). Model yang sudah dilatih sebelumnya akan digunakan untuk mengklasifikasi jenis tutupan lahan pada file citra raster yang dimasukkan. Berikut rancangan tampilan halaman klasifikasi citra pada web dalam gambar 3.10.

Klasifikasi Tutupan Lahan
Klasifikasi Citra
TrainingTesting

Input User

1. Pastikan file TIF yang anda masukkan memiliki Bands ('B2', 'B3', 'B4', 'B8', 'SR', 'NDVI', 'NDWI')

Choose File
2.

Tabel Metadata Citra
3. Raster Image
4. Prediction Image
5.

Persentase Tutupan Lahan

Gambar 3. 10 Halaman Klasifikasi Citra

Keterangan:

1. Fungsi label 1 adalah *input file field* untuk memasukkan file citra raster dengan ekstensi (.tif).
2. Label 2 merupakan table yang akan menampilkan metadata dari citra raster.
3. Label 3 merupakan gambar dari citra raster.
4. Label 4 merupakan gambar citra yang telah diklasifikasi dengan model yang ada.
5. Label 5 akan menampilkan persentase dari setiap jenis tutupan lahan yang telah diklasifikasikan.

BAB 4

IMPLEMENTASI DAN PENGUJIAN SISTEM

4.1. Implementasi Sistem

4.1.1. Spesifikasi Perangkat Keras dan Perangkat Lunak

Spesifikasi perangkat keras yang dipakai untuk pengimplementasian sistem pada penelitian ini adalah sebagai berikut:

1. Processor : Intel(R) Core(TM) i5-7200U CPU @ 2.50GHz (4 CPUs), ~2.7GHz
2. Memory : 8 GB RAM
3. SSD : 512 GB

Berikut adalah spesifikasi perangkat lunak yang dipakai untuk pengimplementasian sistem pada penelitian ini:

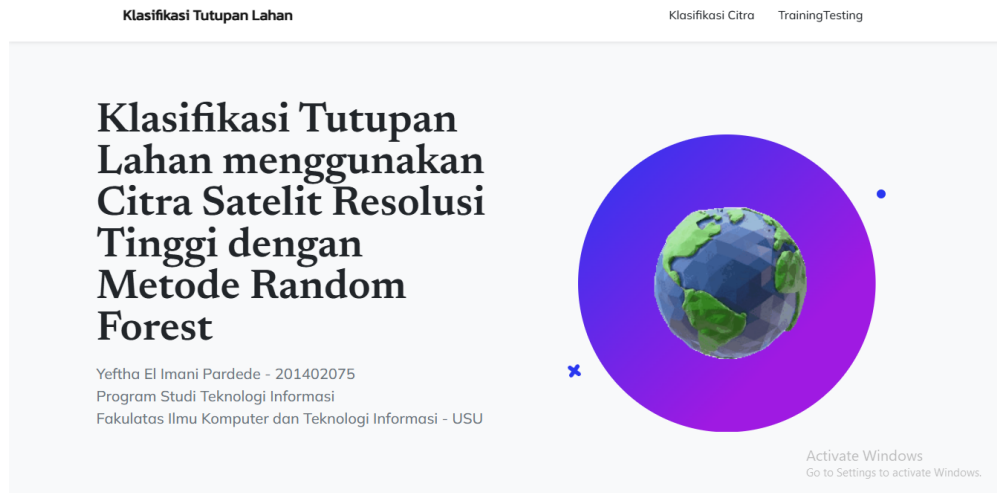
1. Sistem Operasi : Windows 10
2. Microsoft Visual Studio Code
3. Python Versi 3.10.0.
4. *Library* yang digunakan *RandomForestClassifier*, *numpy*, *rasterio*, *pandas*, *geopandas*, *matplotlib*, *seaborn*, *flask*, *scikit-learn*, *pickle*, *joblib*.

4.1.2. Implementasi Perancangan Antarmuka

Dari perancangan yang dilakukan pada bab 3, maka berikut ini merupakan implementasi dari rancangan tersebut.

1. Tampilan Halaman Depan

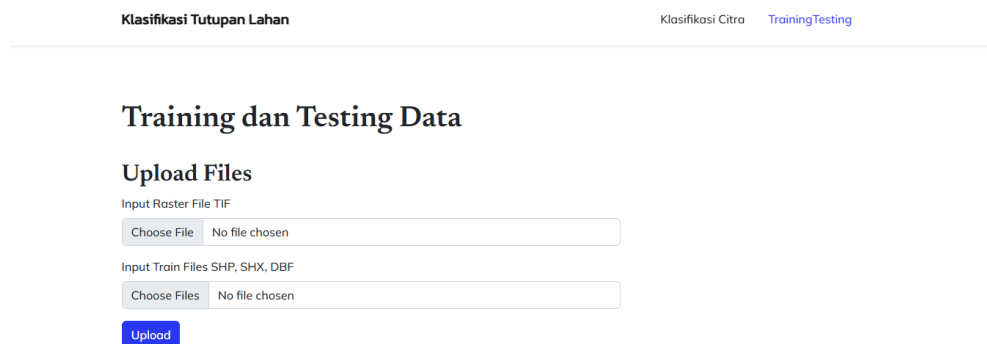
Tampilan halaman depan merupakan tampilan awal yang akan pertama kali dilihat oleh pengguna. Pada halaman tersebut berisi informasi mengenai judul penelitian dan identitas penulis.



Gambar 4. 1 Tampilan halaman depan

2. Tampilan Halaman Training Data

Tampilan ini berisi *input file field* untuk mengupload file yang dibutuhkan dalam tahap *training* dan *testing*. File yang akan diupload adalah file citra raster dengan ekstensi .tif dan juga data *train* dengan ekstensi .shp, .dbf, dan .shx.



Gambar 4. 2 Tampilan halaman training-testing

Setelah file data diunggah maka berikutnya data tersebut akan dilatih dan diuji. Hasil dari proses latih dan uji tersebut akan ditampilkan. Informasi yang akan ditampilkan adalah metadata citra raster, gambar citra raster, gambar citra raster dengan gabungan file train (.shp), tabel pembagian sampel, hasil testing citra raster, persentase jenis tutupan lahan, *validation report*, dan *confusion matrix*.

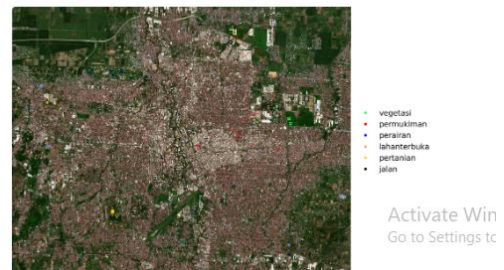
Image Metadata

Key	Value
driver	GTiff
dtype	float32
nodata	None
width	3426
height	2723
count	7
crs	EPSG:32647

Raster Image



Raster and SHP Image



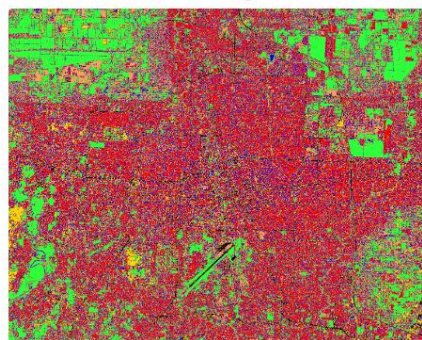
Gambar 4. 3 Tampilan hasil training-testing

Sample Split Information

Class	Train Count	Test Count	Validation Count
vegetasi	1561	446	223
pertanian	1527	436	218
permukiman	1525	435	218
jalan	1462	418	209
perairan	1436	411	205
lahanterbuka	1417	405	203

Gambar 4. 4 Tampilan hasil training-testing (lanjutan)

Prediction Raster Image



Class Percentages



Gambar 4. 5 Tampilan hasil training-testing (lanjutan)

Random Forest Classification Results

Validation Accuracy: 0.9749216300940439

Testing Accuracy: 0.973343786750294

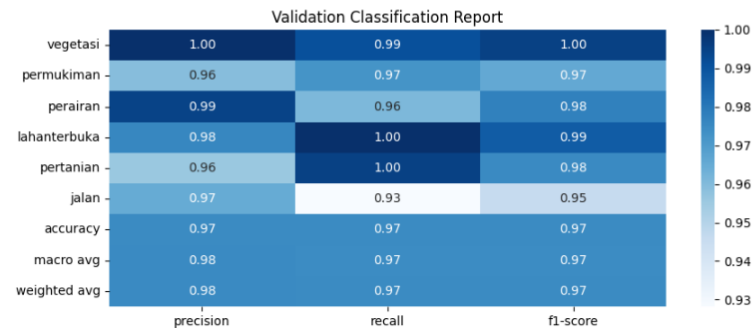
Validation Report

Validation Matrix

Test Report

Test Matrix

Validation Classification Report



Gambar 4. 6 Tampilan hasil training-testing (lanjutan)

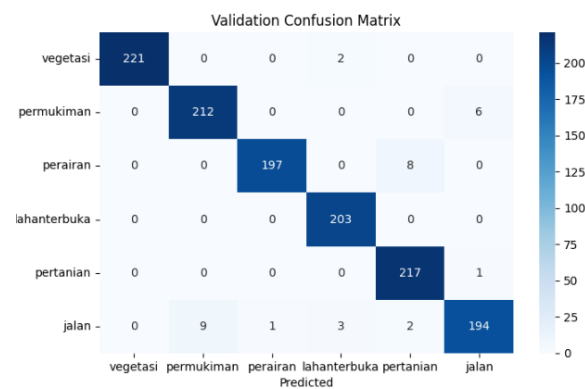
Validation Report

Validation Matrix

Test Report

Test Matrix

Validation Confusion Matrix



Gambar 4. 7 Tampilan hasil training-testing (lanjutan)

3. Tampilan Halaman Klasifikasi Citra

Halaman ini berisi *file input field* citra raster (.tif) yang dapat diunggah oleh pengguna. Citra raster yang diunggah dapat dari peta mana saja asal memiliki 7 *bands* yang diperlukan yaitu *Band 2*, *Band 3*, *Band 4*, *Band 8*, *SR*, *NDVI*, dan *NDWI*. Tampilan halaman ini dapat dilihat pada gambar 4.8 di bawah ini.

Klasifikasi Tutupan Lahan Home Klasifikasi Citra TrainingTesting

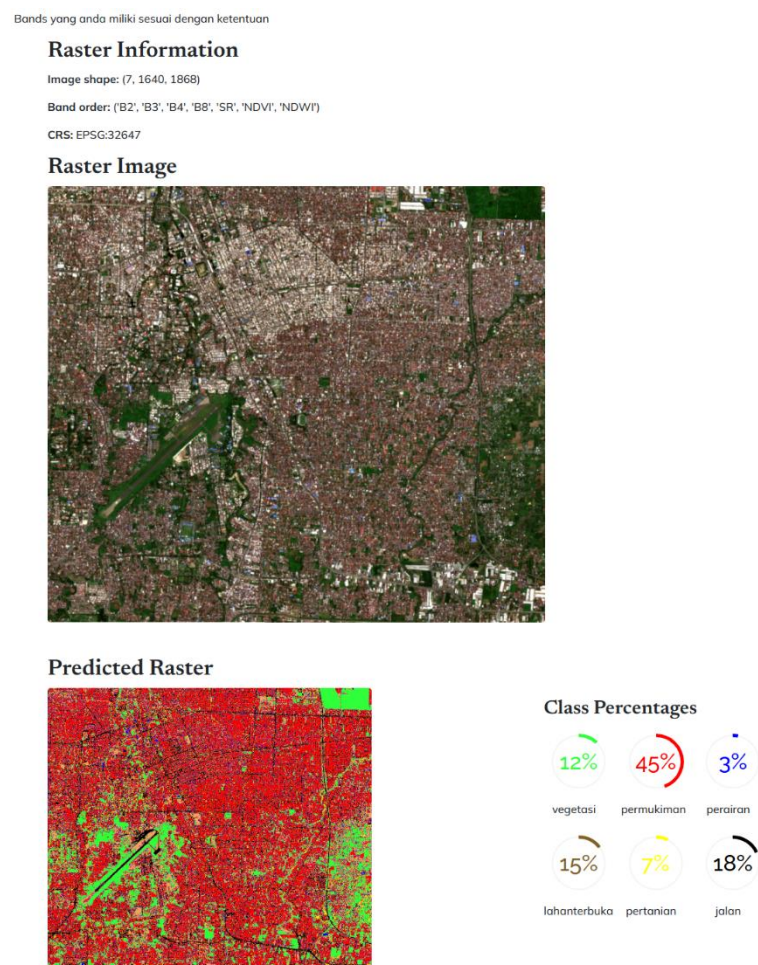
Input User

Pastikan file TIF yang anda masukkan memiliki Bands ('B2', 'B3', 'B4', 'B8', 'SR', 'NDVI', 'NDWI')

No file chosen

Gambar 4. 8 Tampilan halaman klasifikasi citra

Setelah itu data akan diklasifikasikan menggunakan model yang sudah ada. Hasil dari proses tersebut akan menampilkan metadata citra raster, gambar citra raster, hasil klasifikasi citra raster, dan persentase jenis tutupan lahan citra raster. Hasil prediksi dapat dilihat pada gambar 4.9.



Gambar 4. 9 Tampilan hasil prediksi klasifikasi citra

4.2. Implementasi Model

4.2.1. Hasil Pelatihan Model

Bab ini menyajikan hasil penelitian yang diperoleh dari penerapan model *Random Forest* dalam analisis data citra raster yang digunakan. Untuk memastikan bahwa model *Random Forest* bekerja dengan optimal, digunakan teknik *RandomizedSearchCV* untuk mencari kombinasi parameter yang paling sesuai. *RandomizedSearchCV* adalah metode pencarian hyperparameter yang melakukan pencarian secara acak pada ruang hyperparameter yang ditentukan. Teknik ini memungkinkan evaluasi sejumlah kombinasi parameter yang lebih luas dalam waktu yang lebih singkat dibandingkan dengan *GridSearchCV*, sehingga dapat menemukan parameter optimal dengan lebih efisien (Bergstra & Bengio, 2012). Adapun parameter yang digunakan sebagai berikut:

Tabel 4. 1 Daftar Parameter

Parameter	Value
N_estimators	[50, 100, 200]
Max_features	['sqrt', 'log2', None]
Max_depth	[None, 10, 20, 30]
Criterion	['gini', 'entropy']

Melalui parameter pada tabel 4.1, dilakukan kombinasi untuk menentukan *best parameters* yang dapat dilihat pada Tabel 4.2 sebagai berikut

Tabel 4. 2 Hasil *RandomizedSearchCV*

criterion	n_estimators	max_depth	max_features	mean_test_score
Entropy	50	20	sqrt	0,96460513
	50	10	sqrt	0,926186053
	50	none	log2	0,965501049
	50	30	none	0,965612592
	50	none	none	0,965612592
	50	20	log2	0,96460513
	50	none	sqrt	0,965501049
	50	30	log2	0,965501049

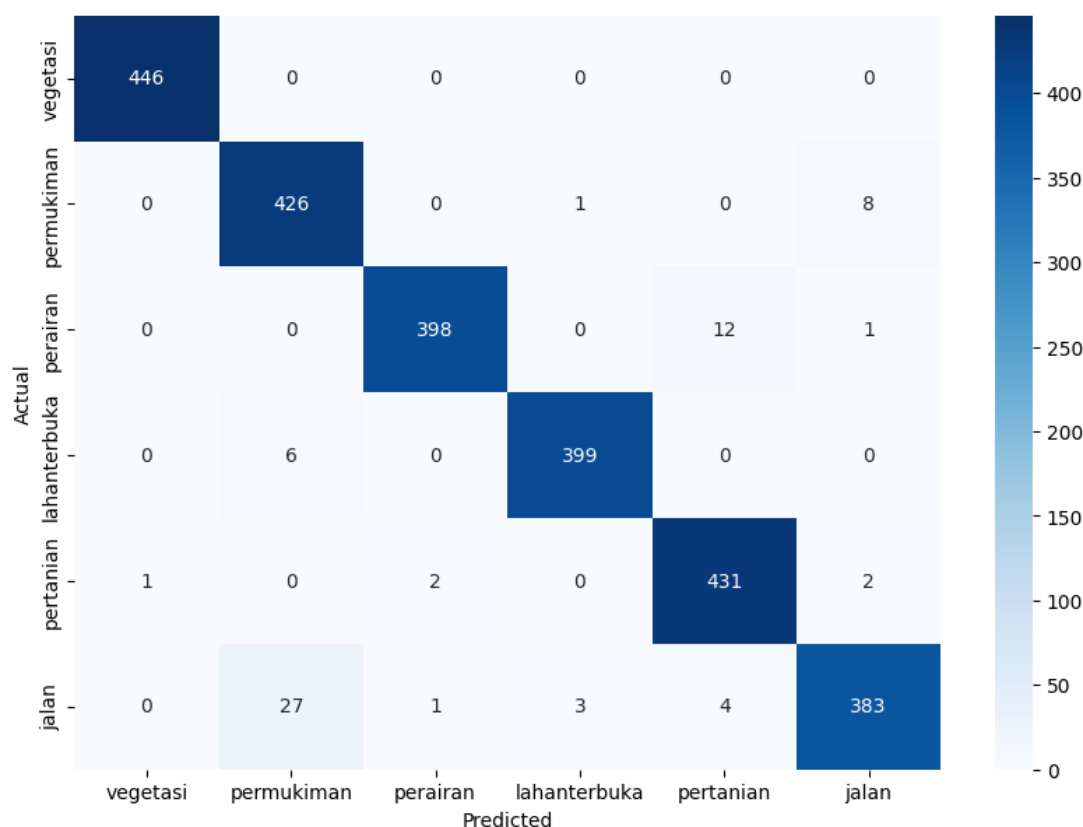
Tabel 4. 2 Hasil *RandomizedSearchCV* (lanjutan)

criterion	n_estimators	max_depth	max_features	mean_test_score
	100	10	log2	0,926409704
	100	20	sqrt	0,966397031
	100	10	sqrt	0,926409704
	100	20	log2	0,966397031
	100	none	log2	0,96673304
	100	30	sqrt	0,96673304
	100	30	none	0,96628461
	200	20	none	0,966508951
	200	10	log2	0,929210197
	200	10	sqrt	0,929210197
	200	20	sqrt	0,96684496
	200	20	log2	0.973343786
	200	none	none	0,966508888
	200	30	none	0,966508888
	200	none	sqrt	0,966172879
	200	10	none	0,945227336
	200	none	log2	0,966172879
Gini	50	20	sqrt	0,964940825
	50	none	sqrt	0,965388942
	50	10	log2	0,91532078
	50	10	sqrt	0,91532078
	50	30	none	0,964604502
	50	30	log2	0,965276897
	50	none	log2	0,965388942
	50	none	none	0,964604502
	50	10	none	0,920473775
	50	30	sqrt	0,965276897
	100	30	sqrt	0,966172691
	100	10	sqrt	0,918793542
	100	none	log2	0,966172691

Tabel 4. 2 Hasil *RandomizedSearchCV* (lanjutan)

criterion	n_estimators	max_depth	max_features	mean_test_score
	100	30	none	0,964828843
	100	30	log2	0,966172691
	100	20	log2	0,964940951
	100	10	none	0,921033623
	100	none	none	0,964828843
	100	20	sqrt	0,964940951
	100	20	none	0,96404478
	100	10	log2	0,918793542
	200	none	log2	0,966284798
	200	30	none	0,965388753
	200	10	none	0,920361919
	200	none	none	0,965388753

Berdasarkan Tabel 4.2, ditemukan *best parameters* dengan skor tertinggi adalah dari kombinasi *Criterion* = *entropy*, *n_estimators* = 200, *max_dept* = 20, dan *max_features* = 20. Dengan kombinasi ini maka model mendapatkan akurasi rata-rata 97.73%. Dalam artian model dapat mengklasifikasikan jenis tutupan lahan dengan akurasi sebesar 97.73%. Adapun *confusion matrix* yang dihasilkan oleh kombinasi tersebut adalah sebagai berikut.



Gambar 4. 10 *Confusion Matrix*

Gambar 4.10 menunjukkan hasil klasifikasi tutupan lahan menggunakan metode *Random Forest*. Matriks ini menggambarkan performa model dalam mengklasifikasikan enam kelas tutupan lahan: vegetasi, permukiman, perairan, lahan terbuka, pertanian, dan jalan. Untuk kelas vegetasi, dari 446 piksel yang sebenarnya adalah vegetasi, seluruhnya diklasifikasikan dengan benar sebagai vegetasi oleh model. Pada kelas permukiman, dari 435 piksel yang sebenarnya adalah permukiman, model berhasil mengklasifikasikan 426 piksel dengan benar sebagai permukiman. Namun, terdapat 1 piksel yang salah diklasifikasikan sebagai perairan dan 8 piksel sebagai jalan. Untuk kelas perairan, model mengklasifikasikan 398 dari 411 piksel yang sebenarnya adalah perairan dengan benar. Namun, terdapat 12 piksel yang salah diklasifikasikan sebagai lahan terbuka dan 1 piksel sebagai jalan.

Selanjutnya, untuk kelas lahan terbuka, dari 405 piksel yang sebenarnya adalah lahan terbuka, model berhasil mengklasifikasikan 399 piksel dengan benar sebagai lahan terbuka. Namun, terdapat 6 piksel yang salah diklasifikasikan sebagai permukiman. Pada kelas pertanian, dari 434 piksel yang sebenarnya adalah pertanian, model berhasil mengklasifikasikan 431 piksel dengan benar sebagai pertanian. Namun,

terdapat 2 piksel yang salah diklasifikasikan sebagai jalan. Terakhir, untuk kelas jalan, dari 418 piksel yang sebenarnya adalah jalan, model berhasil mengklasifikasikan 383 piksel dengan benar sebagai jalan. Namun, terdapat beberapa kesalahan klasifikasi, yaitu 27 piksel diklasifikasikan sebagai permukiman, 1 piksel sebagai perairan, dan 3 piksel sebagai lahan terbuka.

Secara keseluruhan, matriks menunjukkan bahwa model *Random Forest* memiliki performa yang baik dalam mengklasifikasikan sebagian besar kelas dengan tingkat kesalahan yang relatif rendah. Kelas dengan kesalahan terbesar adalah jalan, di mana beberapa piksel salah diklasifikasikan sebagai permukiman. Penjelasan ini menggambarkan performa model dalam klasifikasi tutupan lahan menggunakan citra satelit Sentinel-2.

Tabel 4. 3 Nilai TP, TN, FP, dan FN

Label	TP	TN	FP	FN
Vegetasi	446	2.104	1	0
Permukiman	426	2.083	33	9
Perairan	398	2.137	3	13
Lahan Terbuka	399	2.142	4	6
Pertanian	431	2.099	6	5
Jalan	383	2.122	11	35

Dengan demikian, selanjutnya dapat dilakukan perhitungan nilai *accuracy*, *precision*, *recall*, dan *f1-score* yang merupakan metode evaluasi dari *confussion matrix*. Berikut merupakan perhitungannya.

a. Vegetasi

$$Precision = \frac{TP}{TP+FP} \times 100\% = \frac{446}{446+1} \times 100\% = 99.7\%$$

$$Recall = \frac{TP}{TP+FN} \times 100\% = \frac{446}{446+0} \times 100\% = 100\%$$

$$F1-score = \frac{2 \times precision \times recall}{precision + recall} \times 100\% = \frac{2 \times 0.997 \times 1}{0.997+1} \times 100\% = 99.8\%$$

b. Permukiman

$$Precision = \frac{TP}{TP+FP} \times 100\% = \frac{426}{426+23} \times 100\% = 92.8\%$$

$$Recall = \frac{TP}{TP+FN} \times 100\% = \frac{426}{426+9} \times 100\% = 97.9\%$$

$$F1\text{-score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \times 100\% = \frac{2 \times 0.928 \times 0.979}{0.928 + 0.979} \times 100\% = 97.9\%$$

c. Perairan

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% = \frac{398}{398+3} \times 100\% = 99.2\%$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% = \frac{398}{398+13} \times 100\% = 96.8\%$$

$$F1\text{-score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \times 100\% = \frac{2 \times 0.992 \times 0.968}{0.992 + 0.968} \times 100\% = 98\%$$

d. Lahan Terbuka

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% = \frac{399}{399+4} \times 100\% = 99\%$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% = \frac{399}{399+6} \times 100\% = 98.5\%$$

$$F1\text{-score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \times 100\% = \frac{2 \times 0.99 \times 0.985}{0.99 + 0.985} \times 100\% = 98.7\%$$

e. Pertanian

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% = \frac{431}{431+16} \times 100\% = 96.4\%$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% = \frac{431}{431+5} \times 100\% = 98.8\%$$

$$F1\text{-score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \times 100\% = \frac{2 \times 0.964 \times 0.988}{0.964 + 0.988} \times 100\% = 97.6\%$$

f. Jalan

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% = \frac{383}{383+11} \times 100\% = 97.2\%$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% = \frac{383}{383+35} \times 100\% = 91.6\%$$

$$F1\text{-score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \times 100\% = \frac{2 \times 0.972 \times 0.916}{0.972 + 0.916} \times 100\% = 94.3\%$$

g. Accuracy

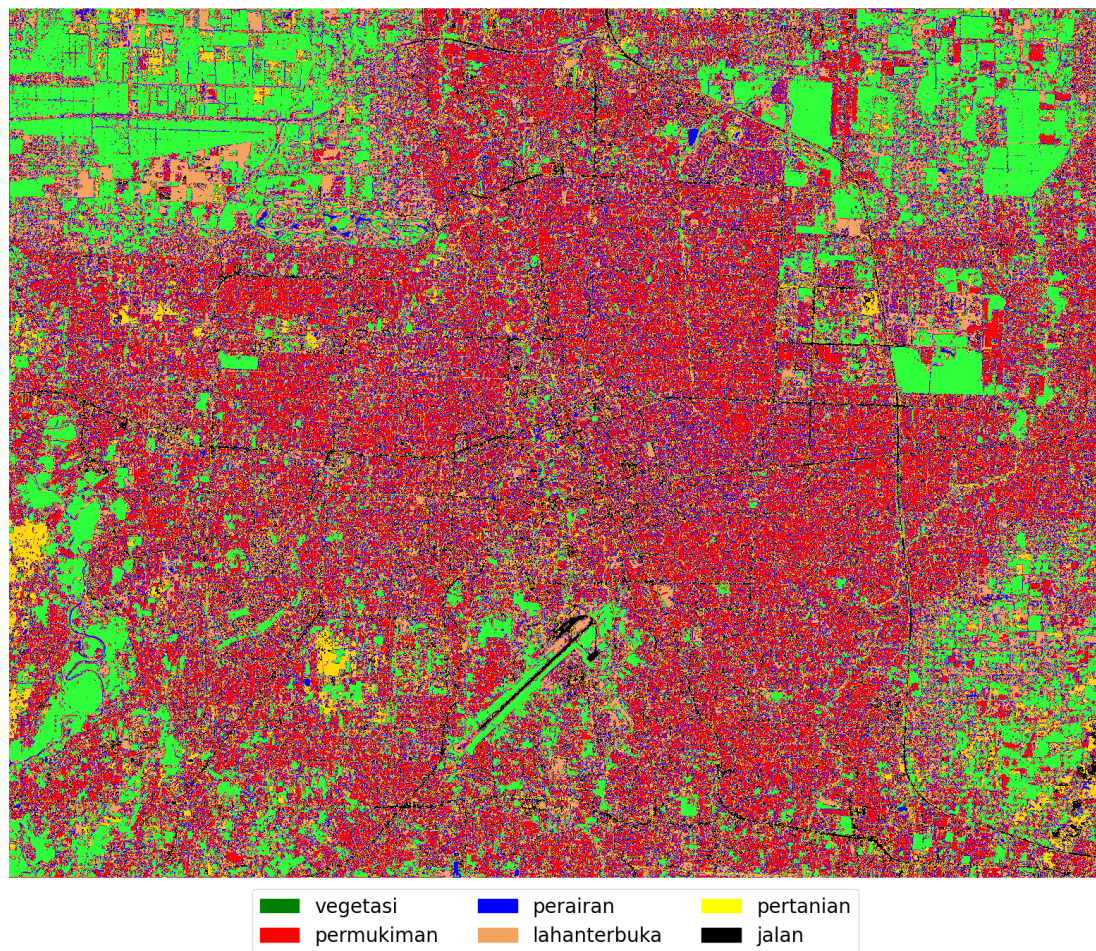
$$\text{Accuracy} = \frac{TP}{TP+FP+TN+FN} \times 100\% = \frac{2483}{2551} = 97.3\%$$

Tabel 4. 4 Nilai *precision*, *recall*, dan *f1- score*

Label	Precision	Recall	F1 Score
Vegetasi	99.7%	100%	99.8%
Permukiman	92.8%	97.9%	97.9%
Perairan	99.2%	96.8%	98%
Lahan Terbuka	99%	98.5%	98.7%
Pertanian	96.4%	98.8%	97.6%
Jalan	97.2%	91.6%	94.3%

4.2.2. Hasil Pengujian Model

Dalam penelitian ini, kami menggunakan model Random Forest untuk mengklasifikasikan tutupan lahan dari citra satelit Sentinel-2. Hasil klasifikasi tersebut divisualisasikan dalam peta yang menunjukkan distribusi berbagai jenis tutupan lahan di wilayah studi. Klasifikasi ini mencakup enam kelas tutupan lahan: vegetasi, permukiman, perairan, lahan terbuka, pertanian, dan jalan. Setiap kelas diwakili oleh warna tertentu pada peta, yaitu hijau untuk vegetasi, merah untuk permukiman, biru untuk perairan, coklat untuk lahan terbuka, kuning untuk pertanian, dan hitam untuk jalan. Gambar 4.11 adalah hasil peta tutupan lahan dari model yang sudah dilatih dan diuji pada citra raster (.tif).



Gambar 4. 11 Hasil Peta Tutupan Lahan

Peta hasil klasifikasi tutupan lahan menggunakan model Random Forest menunjukkan distribusi berbagai jenis tutupan lahan di wilayah studi. Berdasarkan hasil klasifikasi, berikut adalah penjelasan mengenai setiap kelas tutupan lahan yang diidentifikasi dalam peta:

a. Vegetasi (Hijau)

Area yang ditandai dengan warna hijau menunjukkan tutupan lahan berupa vegetasi. Pada peta ini, vegetasi meliputi 21.91% dari total area. Wilayah vegetasi tersebar di beberapa bagian wilayah studi, terutama di area yang mungkin merupakan kawasan hijau atau hutan.

b. Permukiman (Merah)

Area yang ditandai dengan warna merah menunjukkan tutupan lahan berupa permukiman. Kelas ini mendominasi wilayah studi dengan persentase sebesar 35.55%. Permukiman tersebar luas di seluruh peta, menunjukkan area perkotaan dan permukiman penduduk yang padat.

c. Perairan (Biru)

Area yang ditandai dengan warna biru menunjukkan tutupan lahan berupa perairan. Perairan mencakup 3.39% dari total area. Lokasi perairan biasanya ditemukan di sekitar sungai, danau, atau badan air lainnya.

d. Lahan Terbuka (Coklat)

Area yang ditandai dengan warna coklat menunjukkan tutupan lahan berupa lahan terbuka. Lahan terbuka meliputi 17.32% dari total area. Wilayah ini kemungkinan besar mencakup area tanpa tutupan vegetasi yang signifikan, seperti lahan kosong atau area yang baru dikembangkan.

e. Pertanian (Kuning)
















Area yang ditandai dengan warna kuning menunjukkan tutupan lahan berupa pertanian. Pertanian mencakup 8.01% dari total area. Wilayah pertanian biasanya terletak di sekitar pinggiran area permukiman atau di daerah yang lebih terpencil.

f. Jalan (Hitam)


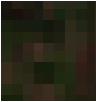





Area yang ditandai dengan warna hitam menunjukkan tutupan lahan berupa jalan. Jalan mencakup 13.82% dari total area. Jaringan jalan terlihat jelas di seluruh peta, menunjukkan infrastruktur transportasi di wilayah studi.

Hasil ini menunjukkan dominasi kelas permukiman di wilayah studi, diikuti oleh lahan terbuka dan vegetasi. Persentase ini memberikan gambaran yang jelas tentang distribusi dan proporsi setiap jenis tutupan lahan, yang dapat digunakan untuk analisis lebih lanjut dalam perencanaan tata ruang dan pengelolaan lingkungan.

Tabel 4. 5 Hasil prediksi benar dan salah

Gambar Pikel	Actual Output	Predicted Output
	Vegetasi	Vegetasi
	Vegetasi	Vegetasi
	Permukiman	Permukiman
	Permukiman	Jalan
	Permukiman	Jalan
	Permukiman	Lahan Terbuka
	Perairan	Perairan
	Perairan	Perairan
	Perairan	Lahan Terbuka
	Perairan	Lahan Terbuka
	Lahan Terbuka	Lahan Terbuka
	Lahan Terbuka	Lahan Terbuka
	Lahan Terbuka	Permukiman
	Lahan Terbuka	Permukiman
	Pertanian	Pertanian

Tabel 4. 6 Hasil prediksi benar dan salah (lanjutan)

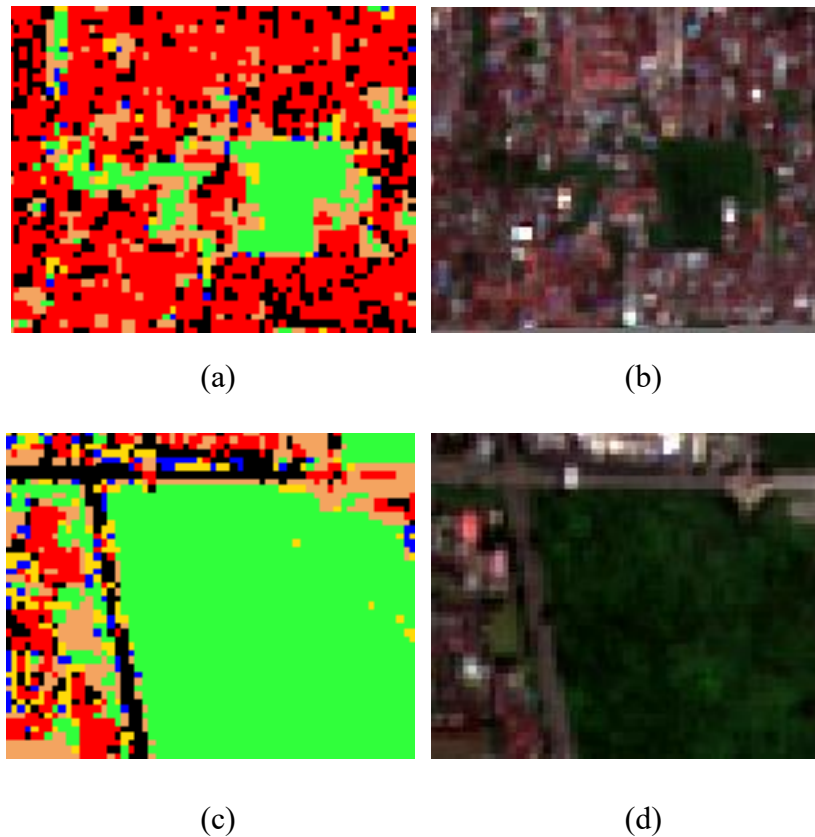
Gambar Piksel	Actual Output	Predicted Output
	Pertanian	Pertanian
	Pertanian	Vegetasi
	Pertanian	Perairan
	Jalan	Jalan
	Jalan	Jalan
	Jalan	Permukiman
	Jalan	Lahan Terbuka

4.3. Analisis Hasil

Untuk memahami lebih lanjut hasil klasifikasi, penulis membandingkan dua visualisasi utama: citra raster RGB asli dan citra hasil klasifikasi dengan colormap (cmap) yang telah diterapkan. Citra raster RGB memberikan representasi visual yang mendekati tampilan asli permukaan bumi, sementara citra hasil klasifikasi dengan cmap memberikan interpretasi yang lebih terstruktur tentang berbagai jenis tutupan lahan di wilayah studi. Analisis ini bertujuan untuk mengevaluasi akurasi klasifikasi model dan memahami distribusi spasial dari berbagai kelas tutupan lahan. Dengan membandingkan kedua visualisasi tersebut, diharapkan nantinya dapat mengidentifikasi area di mana klasifikasi berhasil dengan baik dan area yang memerlukan perbaikan lebih lanjut.

4.3.1. Analisis Hasil Vegetasi

Hasil klasifikasi untuk jenis vegetasi akan dibandingkan dengan citra raster RGB asli untuk mengevaluasi ketepatan dan keakuratan model dalam mengidentifikasi area vegetasi.



Gambar 4. 12 Hasil Peta Tutupan Lahan Vegetasi (a) Hasil Klasifikasi Vegetasi 1; (b) Citra Raster Vegetasi 1; (c) Hasil Klasifikasi Vegetasi 2; (d) Citra Raster Vegetasi 2;

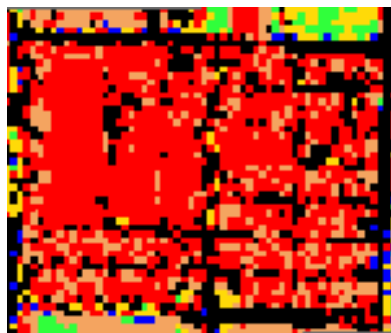
Pada gambar 4.12 menunjukkan gambar (a) dan (c) menunjukkan hasil klasifikasi tutupan lahan menggunakan model *Random Forest*, sedangkan gambar (b) dan (d) adalah citra raster RGB asli Vegetasi yang ditandai dengan warna hijau pada hasil klasifikasi menunjukkan area yang dikenali sebagai vegetasi oleh model.

Hasil klasifikasi menunjukkan bahwa model *Random Forest* mampu mengenali vegetasi dengan sangat baik. Precision untuk kelas vegetasi sangat tinggi (0.997), yang berarti hampir semua prediksi vegetasi oleh model benar-benar merupakan vegetasi. Recall sempurna (1.000) menunjukkan bahwa model mampu mengenali semua area vegetasi dalam citra. F1 Score yang hampir mendekati 1 (0.998) menunjukkan keseimbangan yang sangat baik antara precision dan recall.

Kemampuan model *Random Forest* untuk mengenali vegetasi dengan sangat baik didukung oleh penggunaan indeks spektral seperti *Simple Ratio (SR)* dan *Normalized Difference Vegetation Index (NDVI)*. Indeks ini sangat baik dalam membedakan vegetasi dari jenis tutupan lahan lainnya karena mereka mengeksploitasi karakteristik spektral khas vegetasi yang mencerminkan lebih banyak cahaya dekat-inframerah (NIR) dibandingkan dengan cahaya merah (RED). SR dan NDVI membantu meningkatkan sensitivitas model dalam mendeteksi vegetasi, karena nilai-nilai indeks ini memberikan indikasi kuat tentang kehadiran biomassa hijau yang hidup. Dengan memasukkan indeks ini sebagai fitur dalam model, *Random Forest* dapat lebih akurat mengidentifikasi dan mengklasifikasikan area vegetasi, seperti yang ditunjukkan oleh tingginya nilai *precision*, *recall*, dan *F1 score* dalam hasil klasifikasi.

4.3.2. Analisis Hasil Permukiman

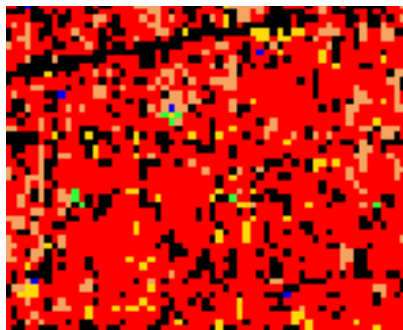
Hasil dari klasifikasi untuk jenis permukiman akan dibandingkan dengan citra raster RGB asli untuk mengevaluasi ketepatan dan keakuratan model dalam mengidentifikasi area permukiman.



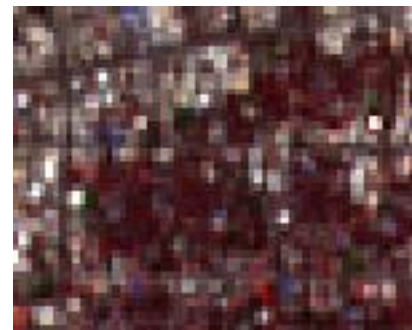
(a)



(b)



(c)



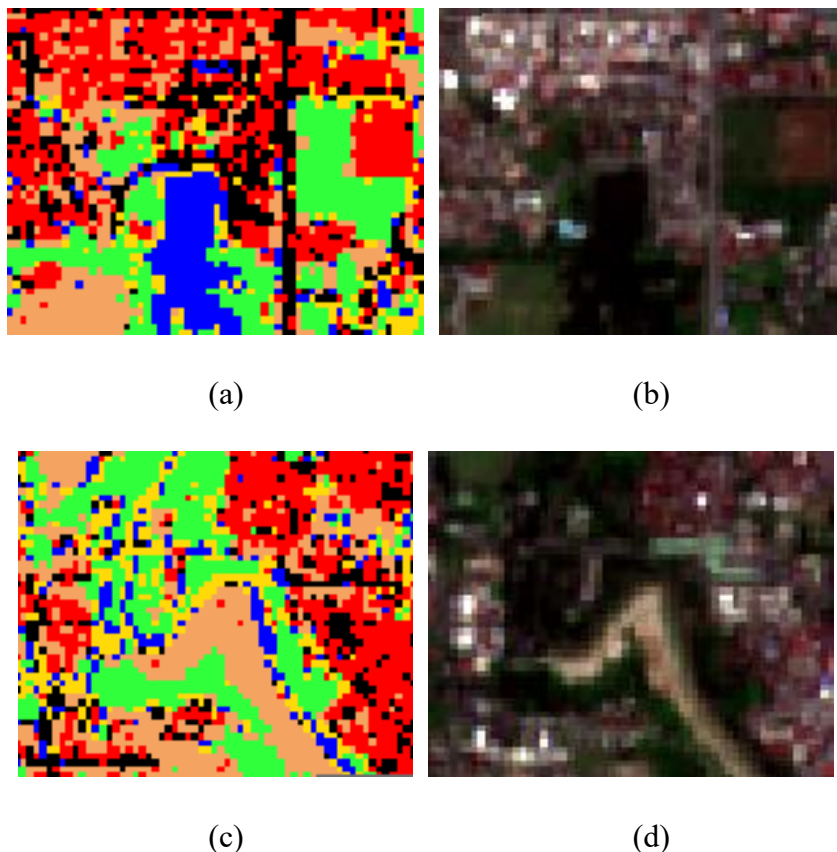
(d)

Gambar 4. 13 Hasil Peta Tutupan Lahan Permukiman (a) Hasil Klasifikasi Permukiman 1; (b) Citra Raster Permukiman 1; (c) Hasil Klasifikasi Permukiman 2; (d) Citra Raster Permukiman 2;

Hasil klasifikasi menunjukkan bahwa model Random Forest mampu mengenali permukiman dengan baik. Precision untuk kelas permukiman cukup tinggi (0.928), yang berarti sebagian besar prediksi permukiman oleh model benar-benar merupakan permukiman. Recall yang sangat tinggi (0.979) menunjukkan bahwa model mampu mengenali hampir semua area permukiman dalam citra. *F1 Score* yang juga tinggi (0.953) menunjukkan keseimbangan yang baik antara *precision* dan *recall*.

4.3.3. Analisis Hasil Perairan

Perairan ditandai dengan warna biru pada hasil klasifikasi menunjukkan area yang dikenali sebagai perairan oleh model.



Gambar 4. 14 Hasil Peta Tutupan Lahan Perairan (a) Hasil Klasifikasi Perairan 1; (b) Citra Raster Perairan 1; (c) Hasil Klasifikasi Perairan 2; (d) Citra Raster Perairan 2;

Hasil klasifikasi menunjukkan bahwa model *Random Forest* dengan *Precision* untuk kelas perairan sangat tinggi (0.992), yang berarti hampir semua prediksi perairan oleh model benar-benar merupakan perairan. *Recall* yang juga sangat tinggi (0.968) menunjukkan bahwa model mampu mengenali sebagian besar area perairan dalam citra. *F1 Score* yang tinggi (0.980) menunjukkan keseimbangan yang baik antara *precision* dan *recall*.

Pada gambar di atas, gambar 4.9 (c) menunjukkan hasil klasifikasi tutupan lahan menggunakan model *Random Forest*, sedangkan gambar (d) adalah citra raster RGB asli. Namun, pada gambar (c) dan (d), terlihat bahwa model belum mengenali perairan dengan baik di beberapa area di Kota Medan yang cenderung memiliki air yang keruh. Model mengenali itu sebagai lahan terbuka. *NDWI (Normalized Difference Water Index)* yang digunakan sebagai fitur spektral memiliki batasan karena tidak selalu dapat mengenali air yang keruh dengan baik. *NDWI* bekerja dengan baik untuk air yang jernih karena mengeksploitasi perbedaan reflektansi antara band hijau dan inframerah dekat. Namun, untuk air yang keruh, reflektansi mungkin tidak mengikuti pola yang sama, sehingga menyebabkan penurunan akurasi dalam klasifikasi.

4.3.4. Analisis Hasil Lahan Terbuka

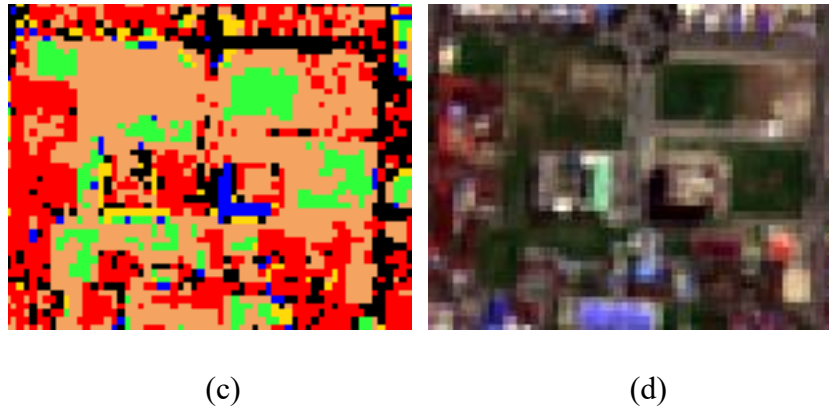
Lahan terbuka ditandai dengan coklat muda pada hasil klasifikasi menunjukkan area yang dikenali sebagai lahan terbuka oleh model.



(a)



(b)



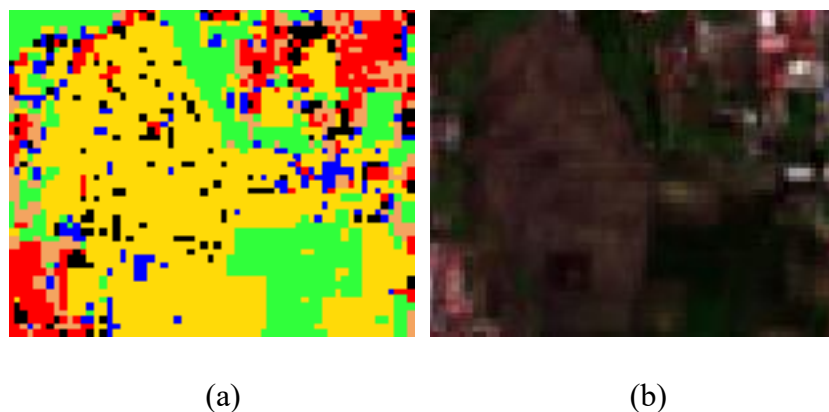
Gambar 4. 15 Hasil Peta Tutupan Lahan Terbuka (a) Hasil Klasifikasi Lahan Terbuka 1; (b) Citra Raster Lahan Terbuka 1; (c) Hasil Klasifikasi Lahan Terbuka 2; (d) Citra Raster Lahan Terbuka 2;

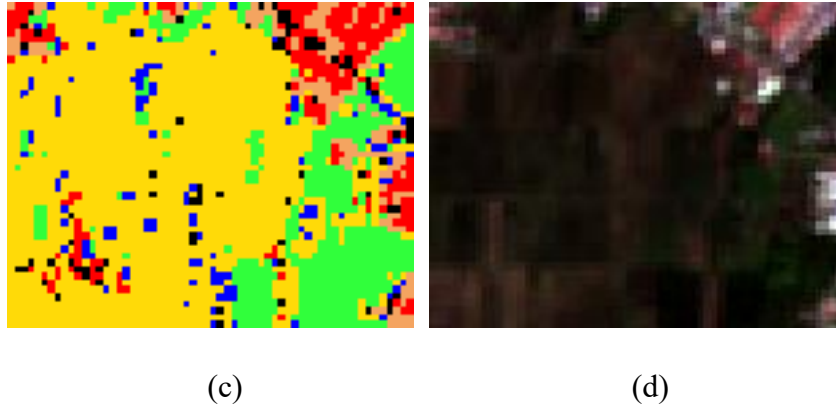
Hasil klasifikasi menunjukkan bahwa model *Random Forest* mampu mengenali lahan terbuka dengan *precision* untuk kelas lahan terbuka sangat tinggi (0.990), yang berarti hampir semua prediksi lahan terbuka oleh model benar-benar merupakan lahan terbuka. *Recall* yang juga sangat tinggi (0.985) menunjukkan bahwa model mampu mengenali sebagian besar area lahan terbuka dalam citra. *F1 Score* yang tinggi (0.987) menunjukkan keseimbangan yang baik antara *precision* dan *recall*.

Ada beberapa area di mana model mengklasifikasikan air keruh sebagai lahan terbuka. Ini menunjukkan keterbatasan dalam pengenalan lahan terbuka yang disebabkan oleh kondisi tertentu seperti air keruh. Seperti yang dijelaskan sebelumnya, air keruh dapat memiliki reflektansi spektral yang mirip dengan lahan terbuka, yang menyebabkan kesalahan klasifikasi.

4.3.5. Analisis Hasil Pertanian

Pertanian yang ditandai dengan warna kuning pada hasil klasifikasi menunjukkan area yang dikenali sebagai pertanian oleh model.





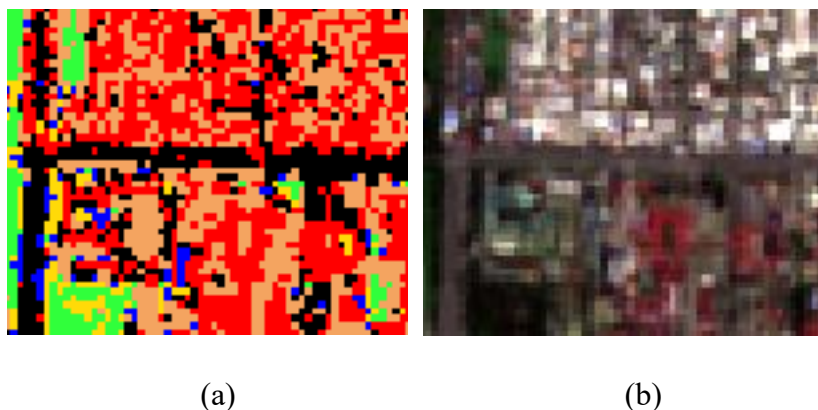
Gambar 4. 16 Hasil Peta Tutupan Pertanian (a) Hasil Klasifikasi Pertanian 1; (b) Citra Raster Pertanian 1; (c) Hasil Klasifikasi Pertanian 2; (d) Citra Raster Pertanian;

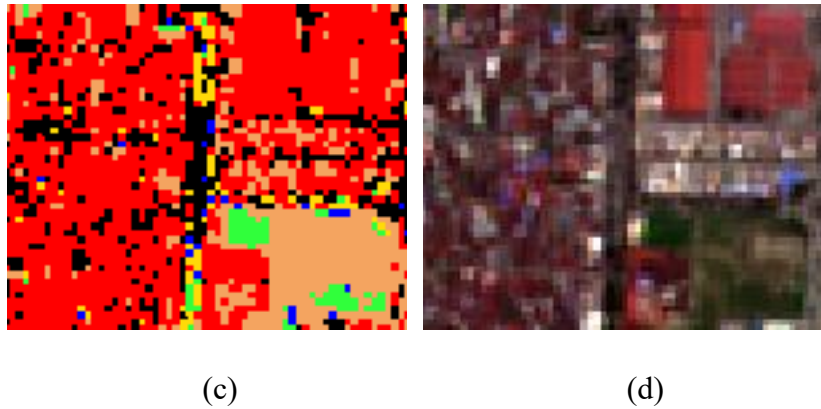
Hasil klasifikasi menunjukkan bahwa model *Random Forest* mampu mengenali pertanian dengan sangat baik. *Precision* untuk kelas pertanian sangat tinggi (0.9642), yang berarti hampir semua prediksi pertanian oleh model benar-benar merupakan pertanian. *Recall* yang tinggi (0.9885) menunjukkan bahwa model mampu mengenali sebagian besar area pertanian dalam citra. *F1 Score* yang tinggi (0.9762) menunjukkan keseimbangan yang baik antara *precision* dan *recall*.

Keterbatasan dalam pengenalan pertanian dapat disebabkan oleh kondisi tertentu seperti pola spektral yang mirip dengan lahan terbuka dan vegetasi. Dalam kasus ini, model *Random Forest* mengklasifikasikan beberapa area lahan terbuka sebagai pertanian karena pola spektral yang mirip. Selain itu, ketika tanaman di lahan pertanian mulai tumbuh dan berwarna hijau, model mungkin salah mengklasifikasikannya sebagai vegetasi alami.

4.3.6. Analisis Hasil Jalan

Jalan yang ditandai dengan warna hitam pada hasil klasifikasi menunjukkan area yang dikenali sebagai jalan oleh model.





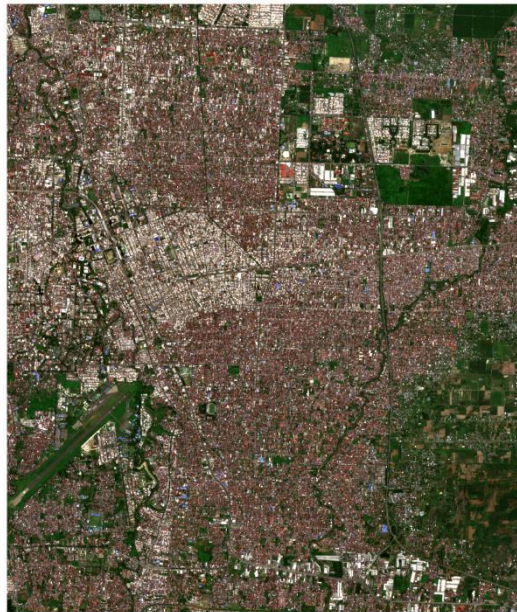
Gambar 4. 17 Hasil Peta Tutupan Lahan Jalan (a) Hasil Klasifikasi Jalan 1; (b) Citra Raster Jalan 1; (c) Hasil Klasifikasi Jalan 2; (d) Citra Raster Jalan 2;

Hasil klasifikasi menunjukkan bahwa model *Random Forest* mampu mengenali jalan dengan baik. *Precision* untuk kelas jalan sangat tinggi (0.972), yang berarti hampir semua prediksi jalan oleh model benar-benar merupakan jalan. *Recall* yang cukup tinggi (0.916) menunjukkan bahwa model mampu mengenali sebagian besar area jalan dalam citra. *F1 Score* yang juga tinggi (0.943) menunjukkan keseimbangan yang baik antara *precision* dan *recall*.

Namun, ada beberapa area di mana model mengklasifikasikan objek lain sebagai jalan. Hal ini dapat terjadi karena beberapa jalan mungkin tertutup bayangan atau vegetasi, yang menyebabkan model kesulitan mengenalnya dengan akurasi tinggi. Selain itu, tekstur dan pola spektral jalan yang mungkin mirip dengan struktur buatan manusia lainnya, seperti atap bangunan atau area parkir, dapat menyebabkan kesalahan klasifikasi

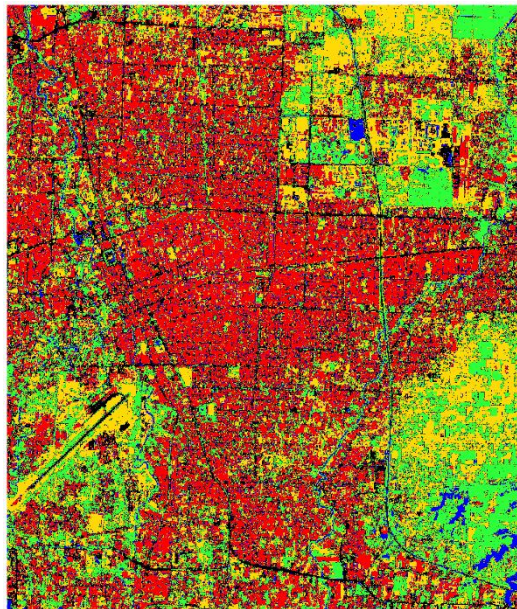
4.3.7. Perbandingan Citra Raster

Dalam melakukan penelitian ini, peneliti mencoba juga untuk mengklasifikasi citra raster dengan memiliki resolusi spasial 10 meter, 20 meter, dan 60 meter. Resolusi spasial ini diperoleh juga dari citra raster sentinel-2 dengan mengambil Band 1, Band 2, Band 3, Band 4, Band 5, Band 6, Band 7, Band 8, Band 8A, Band 9, Band 10, Band 11, dan Band 12.



Gambar 4. 18 Citra Raster dengan 13 *Band*

Dalam melakukan klasifikasi, ditemui akurasi yang rendah diakibatkan karena penggunaan citra raster yang memiliki resolusi diatas 10 meter. Ketika klasifikasi dilakukan model tidak dapat mengenali titik sample dengan baik karena terdapat band yang memiliki resolusi 20 meter dan 60 meter sehingga tidak dapat dikenali dengan baik.



Gambar 4. 19 Gambar Hasil Klasifikasi dengan 13 *Band*

Hasil klasifikasi dengan citra raster yang memiliki 13 *band* dengan resolusi spasial 10 meter, 20 meter, dan 60 meter mendapati hasil akurasi sebesar 82%. Hasil menunjukkan model sulit membedakan lahan terbuka dengan pertanian. Lalu, terdapat beberapa vegetasi juga yang diprediksi sebagai perairan. Hal ini diakibatkan karena resolusi spasial yang digunakan tidak merata dan terdapat band dengan resolusi spasial yang jauh yaitu 60 meter.

Tabel 4. 7 Perhitungan *Precision*, *Recall*, dan *F1 Score* citra raster 13 band

Label	Precision	Recall	F1 Score
Vegetasi	87%	81%	84%
Permukiman	80%	92%	86%
Perairan	82%	90%	86%
Lahan Terbuka	67%	67%	67%
Pertanian	83%	62%	71%
Jalan	85%	80%	82%

BAB 5

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Kesimpulan dari penelitian klasifikasi tutupan lahan menggunakan citra satelit resolusi tinggi dengan metode *random forest* adalah sebagai berikut:

1. Hasil klasifikasi menunjukkan bahwa model *Random Forest* memiliki akurasi yang sangat baik dalam mengenali berbagai jenis tutupan lahan, dengan akurasi keseluruhan mencapai 97.33%.
2. Titik sampel atau *Training Point* dimuat dalam ekstensi ShapeFile (.shp). Pengambilan titik sampel ini berdasarkan prinsip interpretasi citra. Titik sampel terdiri dari 166 polygon dan diubah menjadi 12.755 piksel sebagai titik sampel. Perincian piksel tersebut adalah 2.230 piksel vegetasi, 2.178 piksel permukiman, 2.052 piksel perairan, 2.025 piksel lahan terbuka, 2.181 piksel pertanian, dan 2.089 piksel jalan. Lalu, data tersebut dibagi menjadi 70% *data training*, 20% *data testing*, dan 10% *data validation*.
3. Dalam penelitian ini, data citra raster yang digunakan adalah Band 2, Band 3, Band 4, dan Band 8 dari citra satelit Sentinel-2. Pemilihan band-band ini didasarkan pada resolusi spasialnya yang tinggi, yaitu 10 meter, yang sangat membantu dalam menghasilkan klasifikasi tutupan lahan dengan resolusi yang tinggi. Resolusi spasial 10 meter memungkinkan untuk menangkap detail yang lebih baik dalam citra, sehingga meningkatkan ketepatan dan keakuratan model dalam mengenali berbagai jenis tutupan lahan. Penggunaan band-band ini memberikan informasi spektral yang kaya dan relevan, yang sangat penting untuk tugas klasifikasi tutupan lahan menggunakan model Random Forest.
4. Penggunaan indeks spektral seperti *Simple Ratio (SR)* dan *Normalized Difference Vegetation Index (NDVI)* terbukti sangat membantu dalam meningkatkan tingkat pengenalan model. Indeks-indeks ini mampu menangkap karakteristik spektral khas dari vegetasi, seperti tingginya reflektansi cahaya dekat-inframerah dan

rendahnya reflektansi cahaya merah, sehingga meningkatkan kemampuan model dalam mengenali area vegetasi dengan lebih akurat.

5. Normalized Difference Water Index (NDWI) digunakan dalam model ini untuk meningkatkan kemampuan pengenalan area perairan. NDWI efektif dalam mendeteksi air jernih dengan memanfaatkan perbedaan reflektansi antara band hijau dan inframerah dekat. Namun, dalam kondisi di mana air keruh mendominasi, seperti di beberapa bagian Kota Medan, NDWI kurang efektif karena pola reflektansi spektral air keruh berbeda dari air jernih. Hal ini mengakibatkan beberapa area air keruh diklasifikasikan sebagai lahan terbuka, menunjukkan batasan NDWI dalam kondisi tertentu.

5.2. Saran

Saran yang dapat diberikan oleh peneliti selanjutnya adalah:

1. Menggunakan indeks spektral tambahan, seperti *Enhanced Vegetation Index (EVI)* atau *Soil-Adjusted Vegetation Index (SAVI)*, untuk meningkatkan akurasi klasifikasi. Indeks-indeks ini dapat membantu mengatasi keterbatasan yang dihadapi oleh NDVI dan NDWI dalam kondisi tertentu.
2. Menggunakan data multispektral atau bahkan hiperspektral dengan resolusi spektral yang lebih tinggi dapat membantu dalam mengidentifikasi jenis tutupan lahan yang lebih kompleks dan heterogen. Penelitian lebih lanjut dapat mengeksplorasi penggunaan data dari sensor lain yang memiliki resolusi spektral lebih baik.
3. Pada penelitian berikutnya dapat mempertimbangkan untuk memasukkan fitur tekstural dalam analisis klasifikasi. Fitur tekstural, yang diperoleh dari pengolahan citra seperti *matriks co-occurrence* atau *GLCM (Gray Level Co-occurrence Matrix)*, dapat membantu dalam mengenali pola-pola yang lebih halus dan spesifik, terutama dalam mengenali jalan dan struktur buatan lainnya.

DAFTAR PUSTAKA

- Afasel, D., Purnamasari, R., & Edwar. (2022). Klasifikasi Tutupan Lahan Menggunakan Supervised Machine Learning Pada Citra Satelit Menggunakan Google Earth Engine. *E-Proceeding of Engineering*, 8(6).
- Ardiansyah, R., & Wibowo, W. (2019). Penerapan NDVI untuk Identifikasi Perubahan Tutupan Lahan di Kawasan Rawan Banjir. *Jurnal Geografi*, 21(2), 45–56.
- Badan Standardisasi Nasional. (2010). Standar Nasional Indonesia (SNI) Klasifikasi Penutup Lahan. *Sni 7645:2010*.
- Baret, F., & Guyot, G. (1991). Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of Environment*, 35(2–3). [https://doi.org/10.1016/0034-4257\(91\)90009-U](https://doi.org/10.1016/0034-4257(91)90009-U)
- Belgiu, M., & Drăgu, L. (2016). Random forest in remote sensing: A review of applications and future directions. In *ISPRS Journal of Photogrammetry and Remote Sensing* (Vol. 114). <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13(10), 281–305.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1). <https://doi.org/10.1023/A:1010933404324>
- Brown, E. (2019). *Add Sparkle and Life to Your Web Pages* (3rd Edition). No Starch Press.
- Flanagan, D. (2020). *JavaScript: The Definitive Guide* (7th Edition). O'Reilly Media.
- Gie, T. L. (2008). *Efisiensi Kerja Bagi Pembangunan Negara*. Erlangga.
- Gilles, L. (2014). *Understanding Random Forests: From Theory to Practice*. Université de Liège.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202. <https://doi.org/10.1016/j.rse.2017.06.031>
- Grinberg, M. (2018). *Flask Web Development: Developing Web Applications with Python* (2nd Edition). O'Reilly Media.
- Gunawansa, T., Perera, K., Apan, A., & Hettiarachchi, N. (2022). Application of Sentinel-2 Satellite Data to Map Forest Cover in Southeast Sri Lanka through the Random Forest Classifier. In *Journal of Advances in Engineering and Technology: Vol. I* (Issue I). www.fao.org
- Gurnani, P. (2021). Performance comparison of web frameworks: Flask vs Django vs FastAPI. *Journal of Web Engineering*, 20(4), 234–250.

- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy. In *Nature* (Vol. 585, Issue 7825). <https://doi.org/10.1038/s41586-020-2649-2>
- Havyas, V. B., Choodarathnakara, A. L., Thribhuvan, R., & Chethan, K. S. (2015). Decision Tree Approach for Classification of Satellite Imagery. *International Journal of Soft Computing and Engineering (IJSCE)*, 5(2), 101–104.
- Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3). [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)
- Immitzer, M., Atzberger, C., & Koukal, T. (2012). Tree species classification with Random forest using very high spatial resolution 8-band worldView-2 satellite data. *Remote Sensing*, 4(9). <https://doi.org/10.3390/rs4092661>
- Jensen, J. R., & Cowen, D. C. (1999). Remote sensing of urban/suburban infrastructure and socio-economic attributes. In *Photogrammetric Engineering and Remote Sensing* (Vol. 65, Issue 5).
- Johnson, L., & Brown, K. (2020). Building Responsive Websites with Bootstrap. *Front-End Developer Handbook*, 5(1), 30–50.
- Kawamuna, A., Suprayogi, A., & Wijaya, A. P. (2017). ANALISIS KESEHATAN HUTAN MANGROVE BERDASARKAN METODE KLASIFIKASI NDVI PADA CITRA SENTINEL-2 (Studi Kasus: Teluk Panggang Kabupaten Banyuwangi). *Jurnal Geodesi Undip*, 6(1).
- Krisnayanti, B. D., & Wiryawan, K. G. (2018). Pemanfaatan NDVI untuk Pemantauan Vegetasi di Daerah Rawan Bencana. *Jurnal Teknologi Lingkungan*, 19(1), 55–64.
- Lee, M. (2022). JavaScript Essentials for Web Development. *Programming Today*, 15(4), 100–115.
- Liang, S. (2004). *Quantitative Remote Sensing of Land Surface* (1st Edition). John Wiley & Sons, Inc.
- Lillesand, T. M., & Kiefer, R. W. (1994). Pengindraan Jauh dan Penafsiran Citra. *Remote Sensing and Image Interpretation*.
- Lillesand, T. M., Kiefer, R. W. K., & Chipman, J. (2004). *Remote Sensing and Image Interpretation* (Fifth Edition). John Wiley & Sons, Inc.
- Marlina, D. (2022). Klasifikasi Tutupan Lahan pada Citra Sentinel-2 Kabupaten Kuningan dengan NDVI dan Algoritme Random Forest. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 7(1). <https://doi.org/10.30998/string.v7i1.12948>
- Maulana, F. (2022). *Klasifikasi Jenis Lahan Menggunakan CNN Berbasis Citra Satelit*. Universitas Sriwijaya.

- McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7). <https://doi.org/10.1080/01431169608948714>
- Mishra, S., Khashabi, D., Baral, C., Choi, Y., & Hajishirzi, H. (2022). Reframing Instructional Prompts to GPTk's Language. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. <https://doi.org/10.18653/v1/2022.findings-acl.50>
- O'Reilly, T. (2021). Python Programming: An Introduction to Computer Science. *Coding Journal*, 14(3), 200–215.
- Pratama, I. G. K., & Sugiarto, Y. (2018). Pemanfaatan NDWI untuk pemantauan perubahan luas badan air di wilayah perkotaan. *Jurnal Penginderaan Jauh Indonesia*, 4(22), 90–101.
- Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. NASA special publication. *NASA Special Publication*, 24(1).
- Roy, D. P., Wulder, M. A., Loveland, T. R., C.E., W., Allen, R. G., Anderson, M. C., Helder, D., Irons, J. R., Johnson, D. M., Kennedy, R., Scambos, T. A., Schaaf, C. B., Schott, J. R., Sheng, Y., Vermote, E. F., Belward, A. S., Bindaschadler, R., Cohen, W. B., Gao, F., ... Zhu, Z. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, 145. <https://doi.org/10.1016/j.rse.2014.02.001>
- Sampurno, R., & Thoriq, A. (2016). KLASIFIKASI TUTUPAN LAHAN MENGGUNAKAN CITRA LANDSAT 8 OPERATIONAL LAND IMAGER (OLI) DI KABUPATEN SUMEDANG. *Jurnal Teknotan*, 10(2), 61–70. <https://doi.org/10.24198/jt.vol10n2.9>
- Sharma, A., & Aakanksha. (2018). Introduction to HTML (Hyper Text Markup Language) - A Review Paper. *International Journal of Science and Research (IJSR)*, 7(5), 1337–1339.
- Spatial Reference. (n.d.). *EPSG:32647 - WGS 84 / UTM zone 47N*. 2020.
- Thompson, G. (2020). Developing Web Applications with Flask. *Python Developer Magazine*, 7(1), 55–77.
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2). [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)
- Veriana, K., Hidayat, B., & Sa'idah, S. (2018). Identifikasi dan klasifikasi tutupan lahan melalui pengolahan citra google earth dengan metode singular value decomposition dan klasifikasi k-nearest neighbor. *E-Proceeding of Engineering*, 5(3).
- Wahyuni, N., & Hadi, F. (2020). Penerapan NDWI dalam pemetaan distribusi air untuk manajemen irigasi pertanian. *Jurnal Sumber Daya Air Indonesia*, 16(1), 45–55.

- Wes McKinney, W. (2017). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython* (2nd Edition). O'Reilly Media.
- Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, 27(14). <https://doi.org/10.1080/01431160600589179>
- Zhang, T., Su, J., Xu, Z., Luo, Y., & Li, J. (2021). Sentinel-2 satellite imagery for urban land cover classification by optimized random forest classifier. *Applied Sciences (Switzerland)*, 11(2), 1–17. <https://doi.org/10.3390/app11020543>
- Zulfajria, Danoedoro, P., & Murti, S. H. (2021). Klasifikasi Tutupan Lahan Data Landsat-8 OLI Menggunakan Metode Random Forest. *Jurnal Penginderaan Jauh Indonesia*, 3(1), 1–7.