

KLASIFIKASI TUTUPAN LAHAN PADA CITRA SENTINEL-2 DI KAWASAN IKN MENGGUNAKAN GOOGLE EARTH ENGINE

Hanif Al Fathoni, Achmad Junaidi, Firza Prima Aditiawan

Informatika, Universitas Pembangunan Nasional Veteran Jawa Timur
Jl. Rungkut Madya, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur 60294
achmadjunaidi.if@upnjatim.ac.id

ABSTRAK

Pemerintah Indonesia telah meresmikan pemindahan Ibu Kota Negara (IKN) ke Nusantara melalui Undang-Undang Nomor 3 Tahun 2022. Nusantara dirancang sebagai simbol identitas nasional dan pusat pertumbuhan ekonomi dengan konsep keberlanjutan. Pemindahan ini berdampak pada tata ruang, infrastruktur, dan lingkungan, sehingga analisis tutupan lahan menjadi krusial untuk memastikan perencanaan yang efisien. Penelitian ini bertujuan untuk mengklasifikasikan tutupan lahan di kawasan IKN menggunakan citra satelit Sentinel-2 dan teknologi Google Earth Engine (GEE). Algoritma yang digunakan adalah Random Forest (RF) dan Support Vector Machine (SVM), dengan ekstraksi fitur berbasis indeks spektral NDVI, NDBI, dan NDWI. Teknik cloud masking dengan QA Band diterapkan untuk meningkatkan kualitas data sebelum analisis lebih lanjut. Tahapan penelitian meliputi pengumpulan dan pre-processing data citra Sentinel-2, ekstraksi fitur, pembuatan dataset latih dan validasi, serta proses klasifikasi menggunakan algoritma RF dan SVM. Evaluasi dilakukan dengan metrik akurasi, presisi, recall, dan F1-score untuk menentukan model terbaik. Hasil penelitian menunjukkan bahwa model RF dengan 100 pohon (RF_100trees) dan SVM dengan kernel linear (SVM_LINEAR) memiliki akurasi validasi terbaik sebesar 88%. RF unggul dalam kestabilannya dengan jumlah pohon yang besar, sementara SVM lebih sensitif terhadap pemilihan parameter kernel. Kesimpulannya, kedua model ini efektif dalam klasifikasi tutupan lahan kawasan IKN.

Kata kunci : IKN Nusantara; Klasifikasi; Tutupan lahan; Random Forest; SVM; Google Earth Engine

1. PENDAHULUAN

Pemerintah Indonesia telah meresmikan pemindahan Ibu Kota Negara (IKN) dari DKI Jakarta ke Nusantara melalui Undang-Undang Nomor 3 Tahun 2022 tentang Ibu Kota Negara. Nusantara, yang berlokasi di Kabupaten Penajam Paser Utara serta sebagian wilayah Kabupaten Kutai Kartanegara, Provinsi Kalimantan Timur, dirancang sebagai simbol identitas nasional, pusat pertumbuhan ekonomi baru, dan kota yang berorientasi pada keberlanjutan, dengan luas wilayah daratan sekitar 199.962 hektar. Pemindahan IKN ini diperkirakan akan memberikan dampak signifikan terhadap tata ruang, pembangunan infrastruktur, dan pengelolaan lingkungan di kawasan tersebut. Oleh karena itu, kajian mendalam mengenai pola penggunaan dan tutupan lahan menjadi aspek krusial untuk memastikan perencanaan yang efisien, terpadu, dan sejalan dengan prinsip keberlanjutan. Analisis terhadap penggunaan dan tutupan lahan, yang mengintegrasikan pengamatan karakteristik fisik permukaan bumi dengan interpretasi penggunaannya, menjadi pendekatan yang sangat penting untuk mendukung pengambilan keputusan terkait optimalisasi pemanfaatan lahan [1].

Proses pemetaan tutupan lahan telah menjadi metode yang krusial di berbagai bidang, seperti perencanaan kota, pelestarian sumber daya alam, hingga upaya mitigasi perubahan iklim. Pemetaan ini tidak hanya memvisualisasikan kondisi fisik suatu wilayah, tetapi juga menyediakan informasi mendalam mengenai dinamika ekosistem serta perubahan lingkungan yang terjadi [2]. Teknologi penginderaan

jauh telah menjadi metode yang diandalkan dalam analisis tutupan lahan karena kemampuannya dalam mencakup wilayah yang luas, memperoleh data secara temporal, dan mengurangi biaya [3].

Dalam penelitian ini, teknologi Google Earth Engine (GEE) dipilih sebagai platform utama untuk klasifikasi tutupan lahan di kawasan IKN. GEE merupakan platform berbasis komputasi awan yang kuat untuk mengakses, memproses, dan menganalisis data penginderaan jauh secara efisien. Dengan menggunakan platform GEE, penelitian mengenai klasifikasi tutupan lahan dapat dilakukan dengan mudah, dengan disediakannya berbagai algoritma klasifikasi supervised, seperti Random Forest (RF) dan Support Vector Machine (SVM), yang telah terbukti efektif dalam analisis tutupan lahan [4], [5].

Dengan demikian, penelitian ini akan mengambil judul “Klasifikasi Tutupan Lahan pada Citra Sentinel-2 di Kawasan IKN Menggunakan Google Earth Engine” dengan tujuan untuk menerapkan algoritma klasifikasi Random Forest dan SVM pada citra satelit Sentinel-2 dengan menggunakan Google Earth Engine untuk memperoleh hasil analisis tutupan lahan pada kawasan IKN.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian terdahulu berperan dalam menjelaskan keterkaitan antara riset yang telah dilakukan sebelumnya dengan penelitian yang sedang berlangsung. Berikut ini adalah beberapa jurnal dan

referensi yang digunakan sebagai acuan dalam penelitian ini.

Dalam penelitian pertama yang dilakukan oleh Marlina dengan judul “Klasifikasi Tutupan Lahan pada Citra Sentinel-2 Kabupaten Kuningan dengan NDVI dan Algoritme Random Forest” [3], telah berhasil mengklasifikasikan tutupan lahan menggunakan citra satelit Sentinel-2 di wilayah Kabupaten Kuningan, yang diperoleh melalui situs Earth Explorer pada 21 April 2021. Pada tahap pre-processing, penulis menerapkan komposit band 732 untuk mempermudah identifikasi objek dalam citra berwarna. Metode yang digunakan dalam penelitian ini adalah algoritma Random Forest yang diimplementasikan melalui perangkat lunak QGIS. Selain itu, indeks Normalized Difference Vegetation Index (NDVI) dimanfaatkan untuk menilai kualitas vegetasi dalam citra satelit. Dengan penerapan algoritma serta teknik tersebut, penelitian ini mencapai tingkat akurasi yang cukup tinggi, yaitu 91,39%, dengan nilai kappa sebesar 0,88.

Penelitian kedua oleh Prayogi, dkk. dengan judul “Land cover change analysis in Majalengka Regency using the pan-sharpening method and random forest machine learning algorithm” [6], penelitian tersebut telah berhasil melakukan klasifikasi serta analisis perubahan tutupan lahan di Kabupaten Majalengka, Jawa Barat, pada tahun 2014, 2018, dan 2022. Data yang digunakan berupa citra satelit Landsat-8 untuk wilayah tersebut. Untuk meningkatkan ketajaman detail citra, penelitian ini menerapkan metode pan-sharpening, sehingga menghasilkan klasifikasi yang lebih akurat menggunakan algoritma Random Forest. Hasil penelitian menunjukkan tingkat akurasi keseluruhan (overall accuracy) sebesar 0,96 pada tahun 2014, 0,94 pada tahun 2018, dan 0,93 pada tahun 2022, dengan indeks kappa masing-masing sebesar 0,95, 0,925, dan 0,91.

Penelitian ketiga yang dilakukan oleh Novianti dengan judul “Klasifikasi Tutupan Lahan Menggunakan Google Earth Engine” [7], berhasil menerapkan algoritma klasifikasi Classification and Regression Trees (CART) menggunakan platform Google Earth Engine. Citra satelit yang digunakan adalah Landsat-8, dengan data yang dikumpulkan dalam rentang waktu 1 Januari hingga 31 Desember 2020 di wilayah Kota Palembang. Pada tahap pre-processing, dilakukan pemilihan citra dengan tingkat tutupan awan rendah serta penerapan komposit band 654 untuk meningkatkan kualitas citra. Hasil penelitian menunjukkan bahwa klasifikasi tutupan lahan yang diperoleh memiliki tingkat akurasi sebesar 96,45% dengan nilai kappa accuracy mencapai 93,85%.

2.2. Analisis Tutupan Lahan

Tutupan lahan merujuk pada karakteristik fisik yang tampak di permukaan bumi, sedangkan penggunaan lahan menggambarkan cara manusia memanfaatkan karakteristik fisik tersebut. Oleh karena

itu, analisis terhadap penggunaan dan tutupan lahan merupakan metode untuk mengkaji pemanfaatan lahan berdasarkan kondisi fisik permukaan bumi [1].

Khususnya dalam pemetaan, analisis penggunaan dan tutupan lahan memiliki berbagai aplikasi di berbagai bidang, seperti pertanian, pemantauan lingkungan, dan perencanaan perkotaan. Peta yang menggambarkan tutupan lahan menjadi alat yang sangat bermanfaat karena mampu menampilkan kondisi alami dan sosial suatu wilayah secara detail. Informasi yang diperoleh dari peta ini berperan penting dalam memahami berbagai fenomena yang terjadi di permukaan bumi, termasuk perubahan iklim, perubahan tata guna lahan, serta dinamika ekosistem.

2.3. Kawasan IKN

Kawasan Ibu Kota Negara (IKN) terletak di Kabupaten Penajam Paser Utara dan Kabupaten Kutai Kartanegara, Provinsi Kalimantan Timur. Wilayah ini memiliki peran penting dalam menjaga keanekaragaman hayati, sejalan dengan Kalimantan yang dikenal sebagai "Paru-Paru Dunia" karena hutan-hutannya yang kaya akan biodiversitas dan berfungsi sebagai penyerap karbon serta penghasil oksigen [8].

IKN terbagi menjadi tiga kawasan utama: Kawasan Pengembangan IKN (KP-IKN) seluas 256.142,72 Ha, Kawasan Ibu Kota Negara (KI-IKN) seluas 56.181 Ha yang terdiri dari enam klaster pengembangan, serta Kawasan Inti Pusat Pemerintahan (KIPP) seluas 6.671 Ha. Wilayah ini mencakup berbagai jenis lahan, seperti perkebunan, hutan rimba, hutan bakau, dan hutan rawa, serta memiliki beragam status kehutanan, termasuk Hutan Konservasi di Tahura Bukit Soeharto, Hutan Produksi, dan Hutan Lindung.

2.4. Citra Satelit Sentinel-2

Satelit Sentinel-2, yang diluncurkan oleh Badan Antariksa Eropa (ESA) pada tahun 2015 sebagai bagian dari Program Copernicus, berperan penting dalam pemantauan Bumi. Citra yang dihasilkan memiliki resolusi spasial tinggi, dengan ukuran piksel bervariasi antara 10 m, 20 m, hingga 60 m, tergantung pada pita spektral yang digunakan. Satelit ini dilengkapi dengan 13 pita spektral yang dirancang untuk menangkap berbagai informasi, seperti kondisi vegetasi, permukaan tanah, perairan, dan infrastruktur buatan manusia [9].

Keunggulan utama dari misi ini adalah data yang tersedia secara gratis dan terbuka untuk publik melalui platform seperti Google Earth Engine. Aksesibilitas ini memungkinkan berbagai pihak, termasuk peneliti, pembuat kebijakan, dan masyarakat umum, untuk memanfaatkannya dalam berbagai bidang, seperti pemantauan lingkungan, perencanaan kota, pertanian, serta mitigasi bencana.

2.5. Cloud Masking

Cloud masking pada citra Sentinel-2 dengan menggunakan QA Band QA60 adalah proses

identifikasi dan penghapusan piksel yang tertutup awan dari data satelit [10]. QA60, atau Quality Assurance Band, merupakan fitur Sentinel-2 yang memberikan informasi mengenai keberadaan awan di setiap piksel citra. Band ini menggunakan representasi biner (bitwise flags) untuk mendeteksi dua jenis awan, yaitu awan tebal (opaque clouds) dan awan cirrus, yang masing-masing ditandai pada Bit 10 dan Bit 11 dalam QA60.

2.6. Google Earth Engine

Google Earth Engine (GEE) adalah platform komputasi awan yang sangat canggih, menyediakan kemampuan luas untuk pemetaan dan analisis tutupan lahan dalam skala besar. Melalui GEE, pengguna dapat mengakses beragam sumber data, termasuk citra satelit dari berbagai misi seperti Landsat, Sentinel, dan MODIS [4].

Salah satu keunggulan utama GEE adalah tersedianya berbagai algoritma klasifikasi untuk analisis supervised, seperti CART, Random Forest, Naive Bayes, dan Support Vector Machine (SVM), yang terbukti efektif dalam pemetaan serta klasifikasi tutupan lahan.

Dengan platform ini, peneliti dan praktisi dapat mengakses serta mengolah data secara lebih efisien, memungkinkan analisis yang lebih akurat untuk mendukung pengambilan keputusan berbasis data.

2.7. Klasifikasi Terbimbing

Klasifikasi terbimbing disebut demikian karena memerlukan peran ahli dalam memberikan label atau kelas pada citra satelit. Proses ini menggunakan set pelatihan (training set), yang menjadi faktor utama dalam keberhasilan klasifikasi citra satelit. Akurasi metode ini sangat bergantung pada kualitas sampel pelatihan, yang terbagi menjadi dua jenis: satu untuk klasifikasi dan satu lagi untuk evaluasi akurasi hasil klasifikasi [11].

Namun, kelemahan dari metode ini adalah keterbatasannya dalam mengidentifikasi hanya kelas-kelas yang telah ditentukan dalam sampel pelatihan, serta ketergantungannya pada kualitas data pelatihan yang digunakan. Beberapa algoritma yang umum digunakan dalam klasifikasi terbimbing meliputi Random Forest, Support Vector Machine (SVM), dan Naive Bayes.

2.8. Random Forest

Random Forest adalah algoritma klasifikasi supervised yang menerapkan konsep ensemble learning dengan menggabungkan beberapa classifier untuk menangani masalah yang kompleks [5]. Diperkenalkan oleh Leo Breiman pada tahun 2001, algoritma ini menggunakan sejumlah pohon keputusan (decision tree) yang diterapkan pada berbagai subset data. Prediksi akhir diperoleh dengan metode voting terbanyak untuk klasifikasi dan rata-rata untuk regresi, sehingga meningkatkan akurasi serta ketahanan terhadap overfitting.

Random Forest adalah metode klasifikasi yang banyak digunakan dalam analisis penginderaan jauh. Algoritma ini populer dalam bidang tersebut karena mampu menangani dataset berdimensi tinggi, yang menjadi faktor penting dalam analisis data hiperspektral dan berorientasi objek.

2.9. Support Vector Machine

Support Vector Machine (SVM) adalah algoritma klasifikasi berbasis kernel yang sering digunakan dalam analisis citra [12]. Prinsip utama SVM adalah membangun hyperplane optimal yang memisahkan data ke dalam dua kelas dengan margin terbesar antara sampel terdekat (support vectors) dan hyperplane, sehingga memungkinkan pemisahan kelas yang lebih efektif.

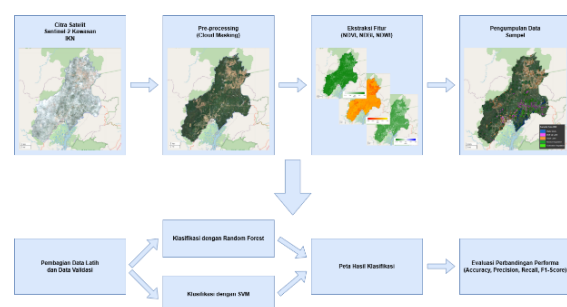
SVM terdiri dari dua jenis utama, yaitu SVM linear dan SVM non-linear. Pada SVM linear, hyperplane berbentuk garis lurus yang memisahkan data dengan margin terbesar, tetapi hanya efektif untuk data yang dapat dipisahkan secara linear, di mana tidak ada titik data yang berada di sisi yang salah dari garis pemisah.

Sementara itu, SVM non-linear menggunakan hyperplane berbentuk kurva yang ditentukan oleh fungsi kernel. Fungsi ini memetakan data ke ruang dimensi yang lebih tinggi, memungkinkan pemisahan data yang tidak dapat dipisahkan secara linear dalam ruang asli. Dalam klasifikasi citra, data umumnya bersifat non-linear, sehingga SVM non-linear lebih sering digunakan.

3. METODE PENELITIAN

3.1. Tahapan Penelitian

Tahapan penelitian adalah rangkaian langkah yang digunakan oleh peneliti sebagai pedoman untuk melaksanakan sebuah penelitian, mulai dari tahap awal hingga tahap akhir.



Gambar 1. Tahapan penelitian

Tahapan penelitian ini dimulai dengan pengumpulan data citra satelit Sentinel-2 yang mencakup kawasan Ibu Kota Negara (IKN) pada rentang waktu 1 Januari 2024 hingga 31 Juli 2024 sebagai sumber utama analisis. Data ini kemudian melalui tahap pre-processing untuk menghilangkan gangguan, seperti keberadaan awan, melalui teknik cloud masking dengan memanfaatkan QA Band dari satelit Sentinel-2. Setelah itu, dilakukan ekstraksi fitur

menggunakan beberapa indeks, yaitu NDVI (Normalized Difference Vegetation Index), NDBI (Normalized Difference Built-up Index), dan NDWI (Normalized Difference Water Index), yang bertujuan untuk mengidentifikasi jenis tutupan lahan seperti vegetasi, area terbangun, dan badan air. Selanjutnya, data sampel dikumpulkan dengan penandaan manual pada area tertentu untuk mewakili kategori tutupan lahan, yang kemudian dibagi menjadi data latih dan data validasi.

Proses klasifikasi tutupan lahan dilakukan menggunakan dua jenis algoritma, yaitu Random Forest (RF) dan Support Vector Machine (SVM) dengan kernel linear. Pemilihan RF didasarkan pada kemampuannya untuk menangani data dengan dimensi tinggi dan fitur yang saling berhubungan, serta keunggulannya dalam menghasilkan hasil klasifikasi yang robust meskipun terdapat noise dalam data. RF juga memungkinkan analisis pengaruh jumlah pohon keputusan terhadap akurasi klasifikasi, sehingga memberikan fleksibilitas dalam penyesuaian parameter model. Di sisi lain, SVM dipilih karena keandalannya dalam memisahkan data dengan margin maksimum, khususnya pada dataset yang tidak terlalu besar, serta kemampuannya memberikan hasil klasifikasi yang presisi dengan pendekatan kernel.

Hasil klasifikasi dari kedua algoritma berupa peta tutupan lahan kawasan IKN, yang kemudian dievaluasi menggunakan metrik seperti akurasi, presisi, recall, dan F1-score. Evaluasi ini bertujuan untuk menentukan algoritma yang paling efektif dalam menganalisis tutupan lahan di kawasan tersebut.

3.2. Pengumpulan Data Citra Satelit Sentinel-2

Penelitian ini memanfaatkan data citra satelit Sentinel-2 yang mencakup wilayah Ibu Kota Nusantara (IKN) dengan periode pengambilan data dari 1 Januari 2024 hingga 31 Juli 2024. Data tersebut diakses melalui platform Google Earth Engine. Dalam proses pengumpulan data, langkah awal yang dilakukan adalah menentukan Region of Interest (ROI), yaitu area yang menjadi fokus utama penelitian untuk proses klasifikasi. Pada penelitian ini, ROI yang ditetapkan meliputi Kawasan Ibu Kota Negara (IKN) dengan luas mencapai 56.181 hektar.

Setelah menentukan Region of Interest (ROI), langkah berikutnya adalah mengumpulkan citra satelit berdasarkan periode waktu yang telah ditentukan untuk menghasilkan komposit citra satelit. Komposit atau mosaik citra satelit merupakan proses penggabungan beberapa citra menjadi satu citra tunggal yang memiliki kualitas optimal. Di platform Google Earth Engine, pembuatan komposit dilakukan dengan menetapkan rentang waktu pengambilan citra melalui parameter tanggal awal dan akhir. Data citra satelit Sentinel-2 diakses melalui koleksi "Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-2A," yang tersedia di Google Earth Engine dengan nama koleksi "COPERNICUS/S2_SR_HARMONIZED."

Tahap berikutnya adalah melakukan penyaringan data Sentinel-2 berdasarkan parameter Region of Interest (ROI) dan periode waktu yang telah ditetapkan. Penyaringan ini dilakukan menggunakan fungsi `filterBounds` dan `filterDate` pada platform Google Earth Engine. Proses tersebut menghasilkan 171 citra satelit Sentinel-2 yang mencakup wilayah Ibu Kota Nusantara (IKN). Dari total 171 citra tersebut, sebuah citra komposit kemudian disusun menggunakan metode median filter. Metode ini menggabungkan data dari koleksi "COPERNICUS/S2_SR_HARMONIZED" yang telah difilter menjadi citra tunggal dengan kualitas yang optimal.

3.3. Preprocessing

Sebelum proses klasifikasi dilakukan, data citra satelit perlu menjalani tahap praproses lanjutan untuk menghilangkan gangguan akibat keberadaan awan. Tahapan ini menggunakan teknik Cloud Masking yang mengintegrasikan penyaringan dan pemanfaatan QA Band. Citra komposit yang dihasilkan sebelumnya masih mengandung banyak area tertutup awan, yang dapat mengganggu akurasi proses klasifikasi.

Langkah awal dalam praproses ini adalah menyaring Image Collection berdasarkan parameter `CLOUD_PIXEL_PERCENTAGE` dengan nilai kurang dari 50%. Penyaringan ini bertujuan untuk memilih citra yang memiliki tutupan awan di bawah 50%, sehingga menghasilkan data yang lebih jernih. Namun, pendekatan ini dapat menyebabkan munculnya area kosong atau gap pada citra akibat terbatasnya ketersediaan data di wilayah tertentu. Untuk mengatasi kekosongan tersebut, diterapkan metode Cloud Masking dengan memanfaatkan Quality Assessment Band (QA Band) dari Sentinel-2, yaitu Band QA60, guna meningkatkan kualitas citra akhir.

Tahapan cloud masking diawali dengan memasukkan citra komposit Sentinel-2 untuk kawasan penelitian. Langkah berikutnya adalah memilih band QA60 dari citra tersebut, yang berisi informasi deteksi awan. Selanjutnya, bitmask didefinisikan untuk mendeteksi awan tebal (Bit 10) dan awan cirrus (Bit 11). Setelah itu, operasi bitwise AND dilakukan untuk setiap piksel pada band QA60 guna menentukan apakah piksel tersebut mengandung awan atau tidak. Jika piksel bebas dari awan, data akan digunakan untuk analisis. Namun, jika piksel tertutup awan, piksel tersebut akan di-masked untuk memastikan hanya piksel berkualitas tinggi yang dipertahankan dalam proses selanjutnya.

3.4. Ekstraksi Fitur

Ekstraksi fitur dilakukan untuk memperoleh informasi penting dari citra satelit yang mendukung proses klasifikasi. Fitur-fitur yang diekstraksi dari citra Sentinel-2 meliputi berbagai indeks spektral, seperti NDVI (Normalized Difference Vegetation Index), NDBI (Normalized Difference Built-up Index), dan NDWI (Normalized Difference Water Index).

NDVI dihitung dengan menggunakan band inframerah dekat (B8) dan band merah (B4), yang bermanfaat untuk memantau kondisi kesehatan vegetasi; nilai NDVI yang lebih tinggi mengindikasikan vegetasi yang lebih sehat. NDBI, yang dihitung dengan memanfaatkan band inframerah dekat (B8) dan band inframerah gelombang pendek (B11), berfungsi untuk mendeteksi area terbangun atau perkotaan, dengan nilai yang lebih tinggi menandakan keberadaan bangunan. Sedangkan NDWI dihitung menggunakan band hijau (B3) dan band inframerah dekat (B8), di mana nilai positifnya menunjukkan adanya keberadaan air di wilayah tersebut.

Ketiga indeks ini dihitung melalui platform berbasis awan Google Earth Engine (GEE) menggunakan rumus yang sesuai untuk masing-masing indeks pada rumus 1, 2, dan 3. Setiap indeks kemudian ditambahkan ke dalam citra penginderaan jauh secara bertahap. Rumus untuk menghitung masing-masing indeks adalah sebagai berikut:

$$NDVI = \frac{\rho NIR - \rho RED}{\rho NIR + \rho RED} \quad (1)$$

$$NDBI = \frac{\rho NIR - \rho SWIR}{\rho NIR + \rho SWIR} \quad (2)$$

$$NDWI = \frac{\rho GREEN - \rho NIR}{\rho GREEN + \rho NIR} \quad (3)$$

Di mana ρNIR , ρRED , $\rho SWIR$, dan $\rho GREEN$ masing-masing berasal dari band spektral pada Sentinel-2, dengan ρNIR diperoleh dari Band 8, ρRED dari Band 4, $\rho SWIR$ dari Band 11, dan $\rho GREEN$ dari Band 3

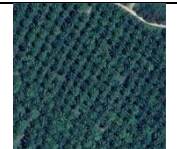
3.5. Pengumpulan Data Sampel

Data sampel memiliki peran yang sangat penting dalam menentukan tingkat akurasi hasil klasifikasi tutupan lahan. Sampel yang tidak memadai atau kurang representatif dapat menyebabkan rendahnya akurasi hasil klasifikasi. Oleh karena itu, pemilihan sampel dilakukan dengan mempertimbangkan kondisi aktual di lapangan serta mengacu pada data tutupan lahan yang sudah tersedia.

Penentuan kelas dalam penelitian ini didasarkan pada pedoman yang tercantum dalam dokumen SNI 7645-1-2014 tentang Klasifikasi Penutup Lahan untuk Skala Kecil dan Menengah [15]. Klasifikasi tersebut kemudian disesuaikan dengan kondisi tutupan lahan di lokasi penelitian, sehingga menghasilkan pembagian kelas sebagaimana ditampilkan pada Tabel 1.

Tabel 1. Pembagian Kelas pada Data Sampel

No	Nama Kelas	Contoh Data
1	Natural Vegetation (Vegetasi Alami)	

2	Cultivation Vegetation (Vegetasi Budidaya)	
3	Open Land (Lahan Terbuka)	
4	Built-up Land (Lahan Terbangun)	
5	Water Body (Badan Air)	

3.6. Pembagian Data Latih dan Data Validasi

Sebelum proses klasifikasi dilakukan, data sampel akan dibagi menjadi dua subset, yaitu data pelatihan dan data validasi. Data pelatihan digunakan untuk membangun atau melatih model klasifikasi, sedangkan data validasi berfungsi untuk mengevaluasi performa model. Pada penelitian ini, proporsi data dibagi sebesar 70% untuk pelatihan dan 30% untuk validasi, sebagaimana diusulkan dalam [14].

Langkah pertama adalah mengimpor data sampel dalam bentuk Feature Collection. Setelah data berhasil diimpor, pembagian dataset dilakukan menggunakan metode stratified random sampling. Teknik ini memastikan bahwa setiap kelas tutupan lahan (landcover class) terwakili secara proporsional dalam subset pelatihan dan validasi. Dengan demikian, distribusi kelas pada subset tetap sesuai dengan distribusi awal data sampel, sehingga model yang dihasilkan lebih akurat dalam menggeneralisasi data [16]. Metode stratified random sampling sangat penting dalam analisis data yang memiliki kelas tidak seimbang, karena teknik ini memastikan representasi yang seimbang untuk setiap kelas dalam dataset. Dengan demikian, model yang dilatih dapat mempelajari pola secara lebih merata dan menghindari bias terhadap kelas yang lebih dominan.

3.7. Skenario Pengujian

Rangkaian skenario pengujian disusun sebagai pedoman untuk mengevaluasi kinerja algoritma dalam mengklasifikasikan tutupan lahan menggunakan citra satelit di wilayah Kawasan IKN. Pengujian ini dirancang dengan tujuan untuk menentukan metode dan konfigurasi parameter yang paling optimal pada masing-masing algoritma, yaitu Random Forest dan SVM. Detail skenario pengujian dijabarkan secara lengkap pada Tabel 2 berikut.

Tabel 2. Skenario Pengujian

Nama Skenario	Parameter
RF_1tree	numberOfTrees = 1, minLeafPopulation = 5
RF_10trees	numberOfTrees = 10, minLeafPopulation = 5
RF_50trees	numberOfTrees = 50, minLeafPopulation = 10
RF_100trees	numberOfTrees = 100, minLeafPopulation = 10
RF_200trees	numberOfTrees = 200, minLeafPopulation = 10
RF_500trees	numberOfTrees = 500, minLeafPopulation = 10
SVM_LINEAR	kernel = LINEAR, cost = 1,0
SVM_LINEAR_cost10	kernel = LINEAR, cost = 10
SVM_LINEAR_cost100	kernel = LINEAR, cost = 100
SVM_POLY	kernel = POLY, cost = 1,0, gamma = 0,1, degree = 3
SVM_POLY_degree5	kernel = POLY, cost = 1,0, gamma = 0,1, degree = 5
SVM_POLY_gamma1	kernel = POLY, cost = 1,0, gamma = 1,0, degree = 3

3.9. Evaluasi dan Perbandingan Performa

Evaluasi dilakukan untuk menilai dan membandingkan kinerja masing-masing model atau skenario yang telah diterapkan. Proses evaluasi ini menghasilkan perbandingan berbentuk confusion matrix serta sejumlah metrik seperti overall accuracy, precision, recall, dan F1-score.

Confusion matrix merupakan tabel yang digunakan untuk menganalisis performa model klasifikasi [17]. Tabel ini menggambarkan hubungan antara prediksi model dengan label aktual data, baik untuk prediksi yang benar maupun salah. Contoh struktur confusion matrix disajikan pada Tabel 3 di bawah.

Tabel 3. Confusion Matrix

		Kelas Sebenarnya	
		1	0
Kelas Prediksi	1	True Positive	False Negative
	0	False Positive	True Negative

True Positive (TP) merupakan jumlah data dari kelas 1 yang berhasil diklasifikasikan dengan benar sebagai kelas 1 oleh model. True Negative (TN) menggambarkan jumlah data dari kelas 0 yang juga diklasifikasikan dengan benar sebagai kelas 0. Sebaliknya, False Positive (FP) adalah jumlah data dari kelas 0 yang secara keliru diklasifikasikan oleh model sebagai kelas 1, sedangkan False Negative (FN) mengacu pada jumlah data dari kelas 1 yang salah diklasifikasikan sebagai kelas 0. Berdasarkan confusion matrix, dapat dihitung nilai overall accuracy (OA), precision, recall, dan F1-score menggunakan formula yang tertera pada Rumus 4, 5, 6, dan 7 berikut.

$$\text{Overall Accuracy (OA)} = \frac{TP + TN}{\text{Total}} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

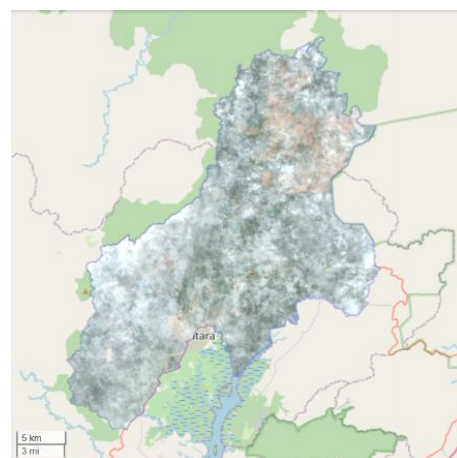
Overall Accuracy (OA) menunjukkan persentase total data yang diklasifikasikan dengan benar, baik positif maupun negatif, dan memberikan gambaran umum kinerja model. Precision mengukur ketepatan model dalam memprediksi data positif, yaitu seberapa banyak prediksi positif yang benar dibandingkan dengan keseluruhan prediksi positif. Recall mencerminkan sensitivitas model, yaitu seberapa baik model mengenali seluruh data positif yang seharusnya terklasifikasi sebagai positif. F1-Score adalah metrik yang menggabungkan precision dan recall dalam satu nilai rata-rata harmonis, yang berguna untuk mengevaluasi model dengan mempertimbangkan keseimbangan antara ketepatan dan sensitivitas.

4. HASIL DAN PEMBAHASAN

Hasil analisis dan pembahasan terkait perbandingan algoritma klasifikasi Random Forest dan SVM akan dimanfaatkan untuk menentukan algoritma yang lebih unggul dalam mengklasifikasikan tutupan lahan berdasarkan citra satelit.

4.1. Hasil Tahap Pengumpulan Citra dan Preprocessing

Pada tahap pengumpulan citra satelit Sentinel-2 di kawasan IKN, menghasilkan komposit citra yang menunjukkan citra yang masih sebagian besar tertutupi oleh awan seperti yang ditunjukkan pada Gambar 2. Kemudian, pada tahap preprocessing, dengan menggunakan teknik cloud masking dengan memanfaatkan QA Band satelit Sentinel-2 dan filter cloud_percentage di bawah 50% menghasilkan citra yang bebas awan seperti yang ditunjukkan pada Gambar 3.



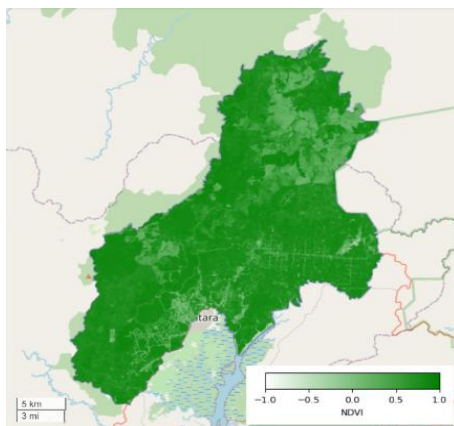
Gambar 2. Citra tertutup awan



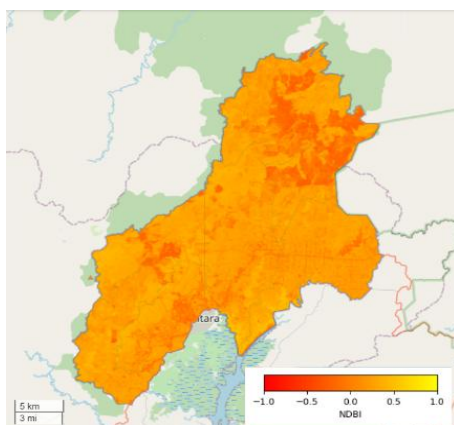
Gambar 3. Citra bebas awan

4.2. Hasil Ekstraksi Fitur

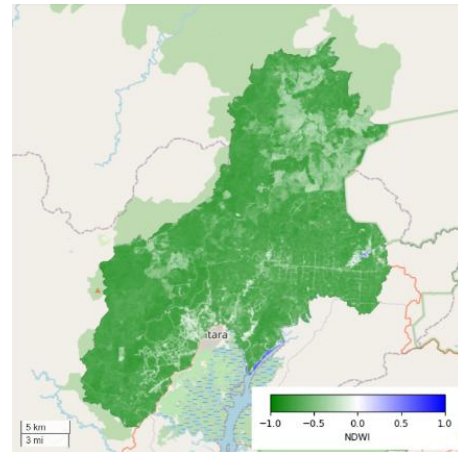
Setelah melalui tahap preprocessing, citra satelit kemudian diproses untuk dilakukan ekstraksi fitur untuk mengekstrak informasi dari citra satelit yang dapat mendukung proses klasifikasi. Fitur yang diambil dari citra satelit Sentinel-2 meliputi indeks spektral, seperti NDVI (Normalized Difference Vegetation Index), NDBI (Normalized Difference Built-up Index), dan NDWI (Normalized Difference Water Index). Hasil dari ekstraksi fitur dengan indeks spektral NDVI, NDBI, dan NDWI pada kawasan IKN, masing-masing ditunjukkan pada Gambar 4, 5, dan 6 berikut.



Gambar 4. Hasil NDVI



Gambar 5. Hasil NDBI



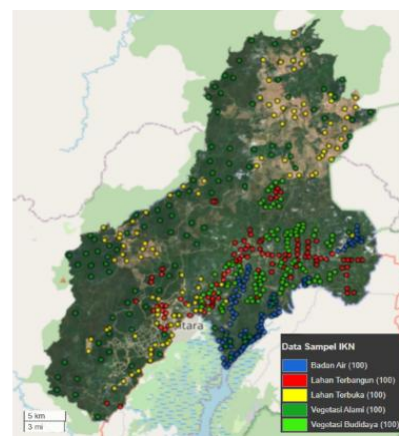
Gambar 6. Hasil NDWI

Pada Gambar 4, warna hijau muda menggambarkan daerah seperti padang rumput atau vegetasi dengan kepadatan rendah, sementara warna hijau tua menunjukkan hutan lebat atau kawasan pertanian dengan vegetasi yang subur. Area hijau muda pada peta menggambarkan lahan terbuka atau lahan terbangun.

Pada Gambar 5, area berwarna kuning hingga oranye menunjukkan lahan non-terbangun, sedangkan warna merah menunjukkan lahan terbangun. Peta NDBI (Normalized Difference Built-up Index) mempermudah visualisasi distribusi permukaan yang dibangun dan membantu dalam pemantauan urbanisasi serta perubahan tutupan lahan. Wilayah pemukiman di kawasan IKN sering kali terlihat samar dengan warna merah pada peta.

Pada Gambar 6, area berwarna biru pada peta NDWI umumnya mencakup badan air seperti sungai, danau, atau waduk, dengan warna biru yang lebih intens menandakan konsentrasi air yang lebih tinggi. Sebaliknya, warna hijau hingga putih menunjukkan tanah kering, vegetasi, atau infrastruktur tanpa kandungan air. Pada peta ini, hanya terdeteksi sedikit badan air di bagian bawah yang merupakan muara sungai dan di bagian kanan peta terdapat waduk.

4.3. Peta Sebaran Data Sampel



Gambar 7. Peta sebaran sampel data

Pada peta distribusi Gambar 7, terlihat bahwa titik sampel tersebar merata di seluruh kawasan IKN. Kelas area terbangun (built-up land), yang ditandai dengan warna merah muda, cenderung terkonsentrasi pada satu wilayah karena pemukiman biasanya berada dalam jarak yang dekat satu sama lain. Sementara itu, kelas badan air (water body) terfokus di sepanjang aliran sungai dan di area muara sungai.

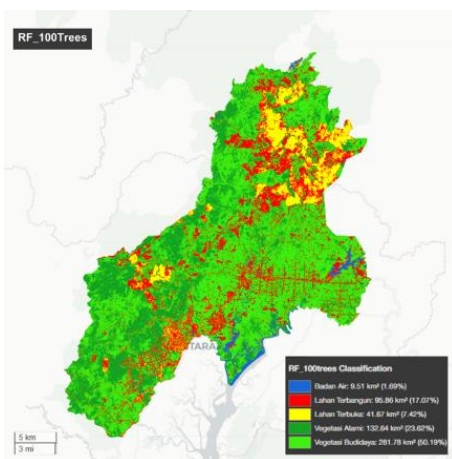
4.4. Hasil Skenario Pengujian

Terdapat 12 skenario pengujian yang dilakukan, setiap skenario dirancang untuk menguji model dalam berbagai kondisi, sehingga dapat ditentukan parameter yang memberikan kinerja terbaik. Untuk parameter-parameter yang digunakan pada masing-masing skenario tertera pada Tabel 4 di bawah ini. Hasil yang tertera pada tabel adalah dalam bentuk persentase.

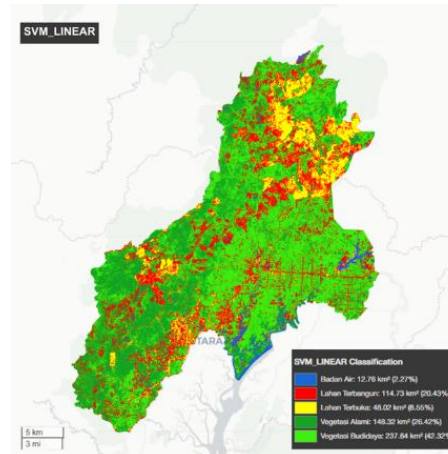
Tabel 4. Hasil Skenario Pengujian

Skenario	Akurasi	Presisi	Recall	F1-Score
RF_1tree	76,67	78,44	76,67	76,35
RF_10trees	84	84,78	84	84
RF_50trees	85,33	85,59	85,33	85,37
RF_100trees	88	88,22	88	88,01
RF_200trees	86,67	86,83	86,67	86,7
RF_500trees	87,33	87,51	87,33	87,36
SVM_LINEAR	88	88,12	88	87,99
SVM_LINEAR_cost10	86,67	86,7	86,67	86,66
SVM_LINEAR_cost100	87,33	87,61	87,33	87,31
SVM_POLY	79,33	79,47	79,33	79,23
SVM_POLY_degree5	70,67	72,23	70,67	69,64
SVM_POLY_gamma1	78,67	78,94	78,67	78,54

Dari 12 skenario pengujian dapat dilihat untuk skenario dengan hasil terbaik dari masing-masing metode, yakni skenario 4 RF_100Trees dan skenario 7 SVM_LINEAR yang menghasilkan peta hasil klasifikasi seperti yang ditampilkan pada Gambar 9 dan 10 di bawah ini.

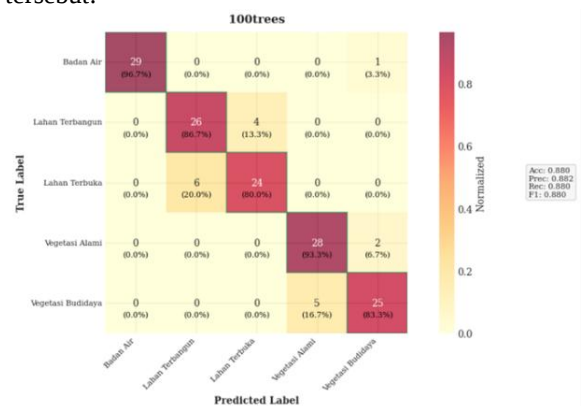


Gambar 8. Peta hasil skenario RF_100Tres



Gambar 9. Peta hasil skenario SVM_LINEAR

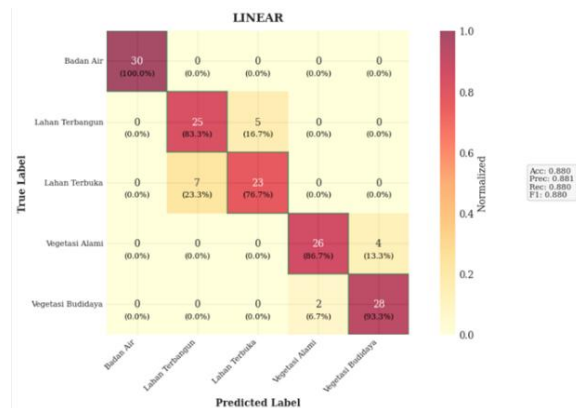
Skenario 4, dengan algoritma Random Forest 100 trees yang ditunjukkan Gambar 9, menunjukkan peningkatan lebih lanjut dalam detail dan presisi, khususnya di wilayah lahan terbuka dan terbangun, terutama di area pembangunan IKN. Meskipun akurasi meningkat, masih ada tumpang tindih kecil antara vegetasi alami dan budidaya. Skenario 7 dengan SVM Linear yang ditunjukkan pada Gambar 10 menunjukkan kinerja yang baik dalam membedakan vegetasi, namun kurang efektif dalam mengklasifikasikan lahan terbangun dan terbuka, dengan distribusi yang kurang detail di area-area tersebut.



Gambar 10. Confusion matrix skenario RF_100Trees

Berdasarkan confusion matrix skenario 4 seperti yang ditunjukkan pada Gambar 10, Random Forest dengan 100 pohon keputusan (RF_100trees) menunjukkan performa lebih baik dibandingkan RF_50trees, dengan akurasi meningkat dari 85,3% menjadi 88%. Precision, recall, dan F1-score masing-masing mencapai 88,2%, 88%, dan 88%, dengan peningkatan akurasi terutama pada kelas Lahan Terbuka dan Vegetasi Alami. Akurasi tertinggi dicapai pada kelas Badan Air (96,7%), diikuti Vegetasi Alami (93,3%), Lahan Terbangun (86,7%), Vegetasi Budidaya (83,3%), dan Lahan Terbuka (80%). Meskipun masih terdapat kesalahan klasifikasi, terutama antara Lahan Terbangun dan Lahan Terbuka serta Vegetasi Budidaya dan Vegetasi Alami, model

RF_100trees lebih mampu mengenali pola yang lebih kompleks dalam data, sehingga hasil klasifikasi menjadi lebih akurat.



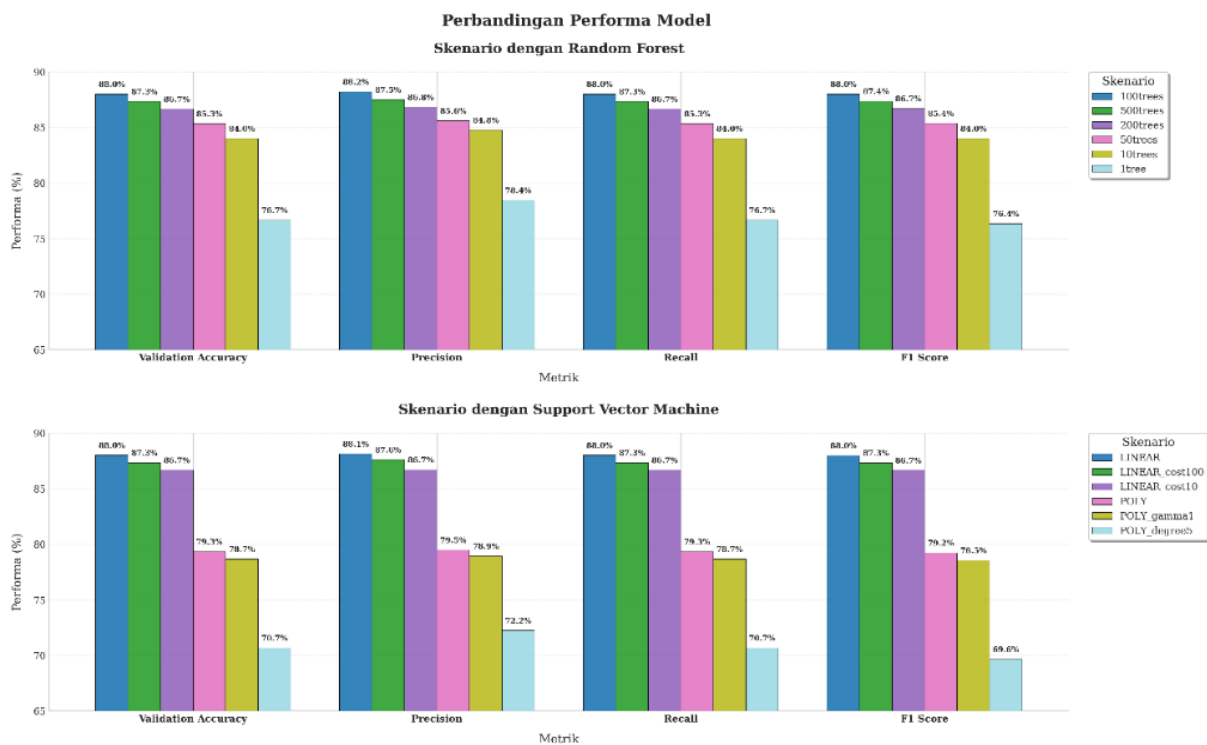
Gambar 11. Confusion matrix skenario SVM_LINEAR

Sedangkan pada confusion matrix skenario 7 seperti yang ditunjukkan pada Gambar 12, model

SVM dengan kernel linear dan cost 1,0 (SVM_Linear) mencapai akurasi keseluruhan 88%, dengan precision 88,1%, recall 88%, dan F1-score 88%. Model ini menunjukkan perbedaan performa dibandingkan Random Forest, terutama dengan akurasi sempurna pada kelas Badan Air (100%) dan Vegetasi Budidaya yang lebih tinggi (93,3%). Namun, akurasi kelas Lahan Terbangun (83,3%) dan Lahan Terbuka (76,7%) lebih rendah dibanding beberapa skenario Random Forest. Hasil ini menunjukkan bahwa SVM Linear lebih unggul dalam mengklasifikasikan Badan Air dan Vegetasi Budidaya, tetapi kurang optimal untuk Lahan Terbuka dan Lahan Terbangun.

4.5. Perbandingan Performa Antar Skenario

Dengan menggunakan metrik akurasi validasi, presisi, recall, dan F1-score, masing-masing skenario yang telah dijalankan akan dievaluasi. Hasil analisis seperti yang ditunjukkan pada Gambar 12 di bawah ini.



Gambar 12. Diagram perbandingan performa model

Berdasarkan Gambar 12, Diagram batang menunjukkan bahwa peningkatan jumlah pohon dalam Random Forest meningkatkan performa model, tetapi tidak selalu linear dan mencapai titik jenuh. Peningkatan signifikan terjadi dari 1 ke 10 pohon dan dari 10 ke 50 pohon, dengan performa terbaik pada 100 pohon (accuracy 88%, precision 88.2%, recall 88%, F1-score 88%). Penambahan lebih dari 100 pohon hanya memberikan peningkatan kecil atau bahkan sedikit penurunan pada beberapa metrik seperti recall. Sementara itu, SVM dengan kernel linear

menunjukkan performa terbaik dengan akurasi 88%, sementara peningkatan cost tidak memberikan manfaat signifikan. Sebaliknya, SVM dengan kernel polynomial mengalami penurunan performa, terutama pada degree 5 (70,67%) dan gamma 1 (78,67%), menunjukkan bahwa kompleksitas yang berlebihan menghambat generalisasi. Dengan demikian, SVM kernel linear adalah pilihan terbaik untuk dataset ini, menawarkan keseimbangan optimal antara akurasi dan efisiensi.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil evaluasi performa model, Random Forest dengan 100 pohon (RF_100trees) dan SVM dengan kernel linear (SVM_LINEAR) memiliki performa terbaik, keduanya mencapai akurasi validasi sebesar 88%. Random Forest menunjukkan peningkatan performa yang signifikan dengan bertambahnya jumlah pohon, namun setelah 100 pohon, peningkatan menjadi tidak signifikan, bahkan sedikit menurun pada 200 dan 500 pohon dalam beberapa metrik seperti recall dan F1-score. Sementara itu, SVM linear memiliki performa yang stabil tanpa peningkatan berarti meskipun parameter cost ditingkatkan. SVM dengan kernel polynomial mengalami penurunan akurasi yang cukup besar, menunjukkan sensitivitas tinggi terhadap pemilihan parameter. Secara keseluruhan, RF_100trees dan SVM_LINEAR merupakan model terbaik untuk klasifikasi tutupan lahan di Kawasan IKN, dengan Random Forest unggul dalam kestabilannya pada jumlah pohon yang cukup besar, sementara SVM lebih sensitif terhadap pemilihan parameter kernel.

DAFTAR PUSTAKA

- [1] M. W. Naikoo, M. Rihan, dan M. Ishtiaque, "Analyses of land use land cover (LULC) change and built-up expansion in the suburb of a metropolitan city: Spatio-temporal analysis of Delhi NCR using ...," *Journal of Urban Management*. Elsevier, 2020. [Daring]. Tersedia pada: <https://www.sciencedirect.com/science/article/pii/S2226585620300637>
- [2] K. Jia, X. Wei, X. Gu, Y. Yao, X. Xie, dan B. Li, "Land cover classification using Landsat 8 operational land imager data in Beijing, China," *Geocarto international*, 2014, doi: 10.1080/10106049.2014.894586.
- [3] D. Marlina, "Klasifikasi Tutupan Lahan pada Citra Sentinel-2 Kabupaten Kuningan dengan NDVI dan Algoritme Random Forest," *STRING (Satuan Tulisan Riset dan Inovasi ...)*, 2022, [Daring]. Tersedia pada: <https://journal.lppmunindra.ac.id/index.php/STRING/article/view/12948>
- [4] A. Shelestov, M. Lavreniuk, N. Kussul, dan ..., "Exploring Google Earth Engine platform for big data processing: Classification of multi-temporal satellite imagery for crop mapping," *frontiers in Earth ...*, frontiersin.org, 2017. doi: 10.3389/feart.2017.00017.
- [5] C. A. Ramezan, T. A. Warner, A. E. Maxwell, dan B. S. Price, "Effects of training set size on supervised machine-learning land-cover classification of large-area high-resolution remotely sensed data," *Remote Sensing*, 2021, [Daring]. Tersedia pada: <https://www.mdpi.com/2072-4292/13/3/368>
- [6] H. Prayogi, H. Setiadi, S. Supriatna, dan D. Dewayany, "Land cover change analysis in Majalengka Regency using the pan-sharpening method and random forest machine learning algorithm," *Jurnal Pendidikan Geografi: Kajian ...*, 2023.
- [7] T. C. Novianti, "Klasifikasi Tutupan Lahan Menggunakan Google Earth Engine," *JURNAL SWARNABHUMI: Jurnal ...*, 2021, [Daring]. Tersedia pada: <https://jurnal.univpgri-palembang.ac.id/index.php/swarna/article/view/5105>
- [8] D. J. Mutaqin, M. B. Muslim, dan ..., "Analisis konsep forest city dalam rencana pembangunan Ibu Kota Negara," *Bappenas ...*, 2021.
- [9] D. Phiri, "Sentinel-2 data for land cover/use mapping: A review," *Remote Sensing*, vol. 12, no. 14, 2020. doi: 10.3390/rs12142291.
- [10] X. Gao, H. Chi, J. Huang, Y. Han, Y. Li, dan F. Ling, "Comparison of Cloud-Mask Algorithms and Machine-Learning Methods Using Sentinel-2 Imagery for Mapping Paddy Rice in Jiangnan Plain," *Remote Sensing*. mdpi.com, 2024. [Daring]. Tersedia pada: <https://www.mdpi.com/2072-4292/16/7/1305>
- [11] S. Abburu dan S. B. Golla, "Satellite image classification methods and techniques: A review," *International journal of computer applications*. Citeseer, 2015. [Daring]. Tersedia pada: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=69613390ca76bf103791ef251e1568deb5fe91dd>
- [12] T. Adugna, W. Xu, dan J. Fan, "Comparison of random forest and support vector machine classifiers for regional land cover mapping using coarse resolution FY-3C images," *Remote Sensing*. mdpi.com, 2022. [Daring]. Tersedia pada: <https://www.mdpi.com/2072-4292/14/3/574>