

KLASIFIKASI TUTUPAN LAHAN BERDASARKAN RANDOM FOREST ALGORITHM MENGGUNAKAN CLOUD COMPUTING PLATFORM

(Land Cover Classification Based on Random Forest Algorithm Using Cloud Computing Platform)

Hady Suryono¹, Arif Handoyo Marsuhandi², dan Setia Pramana³

¹Badan Pusat Statistik

*e-mail: hadys@go.id

Abstrak

Statistik pertanian merupakan salah satu data yang vital di dunia dan memiliki kontribusi besar terhadap pencapaian tujuan program Sustainable Development Goals (SDGs). Dalam SDGs, perhatian terhadap ketahanan pangan difokuskan pada indikator kunci kedua yaitu nol kelaparan (SDG 2). Ketersediaan data tutupan lahan yang akurat diperlukan sebagai data dasar untuk luasan baku sawah yang akan digunakan untuk mengukur tingkat ketahanan pangan. Pemetaan tanaman membutuhkan pemrosesan dan pengelolaan data citra satelit dengan volume yang sangat besar dan tidak terstruktur yang mengarah pada permasalahan Geo Big Data dan menuntut teknologi dan sumber daya baru yang mampu menangani citra satelit dalam jumlah besar. Secara khusus, munculnya sumber daya cloud computing, seperti Google Earth Engine telah mengatasi masalah Geo Big Data ini. Kami menggunakan algoritma Random Forest (RF) pada platform Google Earth Engine (GEE) di Kota Jakarta Utara pada tahun 2019 untuk mengklasifikasikan tutupan lahan. Hasil penelitian menunjukkan bahwa overall accuracy (OA) menggunakan algoritma RF adalah 93,37%, statistik kappa (κ) adalah 0,86 dan F1 score (F1) adalah 86,24%.

Kata kunci: sustainable development goals, klasifikasi, tutupan lahan, google earth engine, random forest

Abstract

The agricultural sector is one of the vital sectors in the world and has a major contribution to the achievement of the goals of the Sustainable Development Goals (SDGs) program. In the SDGs, attention to food security is focused on the second key indicator, namely zero hunger (SDG 2). The availability of accurate land cover data is needed as basic data for the raw area of rice fields that will be used to measure the level of food security. Plant mapping requires the processing and management of very large volumes of unstructured satellite image data which leads to Geo Big Data problems and demands new technology and resources capable of handling large amounts of satellite imagery. In particular, the emergence of cloud computing resources, such as Google Earth Engine, has addressed this Geo Big Data problem. We used the Random Forest (RF) algorithm on the Google Earth Engine (GEE) platform in North Jakarta City in 2019 to classify land cover. The results showed that the overall accuracy (OA) using the RF algorithm was 93.71%, the kappa statistic (κ) was 0.86 and the F1 score (F1) was 86,24%.

Keywords: sustainable development goals, classification, land cover, google earth engine, random forest

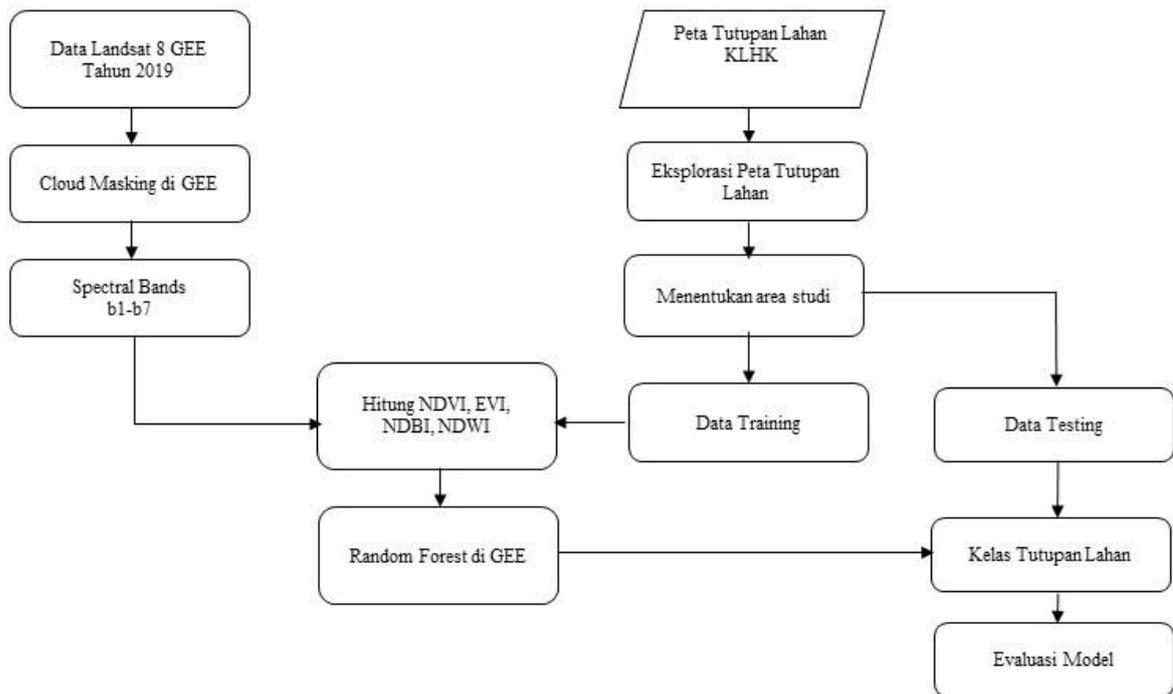
PENDAHULUAN

Statistik pertanian merupakan salah satu data penting di dunia dan memiliki kontribusi yang sangat signifikan terhadap pencapaian program Sustainable Development Goals (SDGs). SDGs yang diadopsi oleh negara-negara di dunia memiliki implikasi yang signifikan dan bersifat global terhadap perencanaan pembangunan pada tahun 2015-2030 (Omilola & Robele, 2017). Pada indikator kunci SDGs, perhatian terhadap ketahanan pangan difokuskan pada tujuan kedua yaitu zero hunger (SDG 2), mewujudkan ketahanan pangan, perbaikan gizi, dan mendorong budidaya pertanian berkelanjutan (FAO, 2016). Pertanian merupakan bagian penting dari ekosistem darat dan memainkan peran penting dalam mata pencaharian manusia (Ramankutty, *et al.*, 2018). Berdasarkan data BPS, produksi padi di Indonesia Januari-September 2018 sebesar 49,65 juta ton Gabah Kering Giling (GKG) atau setara dengan 34,42 juta ton beras, sedangkan konsumsi beras di Indonesia hingga Desember 2018 diperkirakan mencapai 29,57 juta ton.

Melihat kebutuhan beras yang sangat besar tersebut maka tersedianya data pertanian sebagai bahan pangan pokok yang tepat waktu dan akurat memiliki nilai strategis dan mempunyai pengaruh yang besar dalam bidang ekonomi, lingkungan dan sosial politik. Kebutuhan pangan terus meningkat seiring dengan pertumbuhan populasi. Sementara itu, lahan pertanian belum dapat memenuhi seluruh kebutuhan pangan penduduk Indonesia, sehingga sangat memungkinkan perluasan lahan dan intensifikasi pertanian. Pemantauan tutupan lahan dilakukan untuk menjawab tujuan SDGs kedua. Klasifikasi tutupan lahan merupakan langkah awal dalam pemantauan kebutuhan lahan pertanian dan ketersediaan data Luas Baku sawah. Klasifikasi tutupan lahan dapat menghasilkan 5 kelas tutupan lahan, antara lain: kelas perairan, kelas pertanian, kelas permukiman, kelas hutan dan kelas lahan terbuka.

Pada bidang pertanian, data penginderaan jauh seperti Landsat 8 memberikan kemampuan untuk memperoleh informasi pertanian yang akurat dan tepat waktu (Thenkabail, *et al.*, 2019), pemetaan tutupan lahan (Congalton & Green, 2019), pemodelan pertumbuhan padi dengan fitur spasio-temporal (Parsa, *et al.*, 2017), klasifikasi jenis tanaman (Azar, *et al.*, 2016), estimasi luas panen (You, *et al.*, 2017), pemetaan padi (Zhang, *et al.*, 2018).

Pemodelan data penginderaan jauh untuk melakukan klasifikasi tutupan lahan memerlukan pengolahan dan pengelolaan data citra satelit dengan volume yang sangat besar dan tidak terstruktur yang mengarah pada permasalahan "Big Data". Hal ini dikenal sebagai masalah "Geo Big Data" dan membutuhkan teknologi dan sumber daya baru yang mampu menangani citra satelit dalam jumlah besar (Shelestov, *et al.*, 2017). Secara khusus, munculnya sumber daya cloud computing platform, seperti Google Earth Engine (GEE), dapat memecahkan masalah Geo Big Data (Mutanga & Kumar, 2019). Penggunaan cloud platform yang tersedia secara terbuka, banyaknya pilihan multiple sensor (Landsat, sentinel) dan machine learning algorithms menjadikan GEE sangat populer beberapa tahun terakhir untuk melakukan proses data satellite imagery. Peningkatan volume dan variasi data penginderaan jauh mengakibatkan masalah Big Data membutuhkan tantangan baru dalam menangani kumpulan data sehingga diperlukan pendekatan baru untuk mengekstrak informasi yang relevan dan akurat dari data penginderaan jauh (Kussul, *et al.*, 2016). Algoritma Random Forest (RF) dapat digunakan sebagai pendekatan baru untuk memetakan tutupan lahan melalui Google Earth Engine (Teluguntla & Thenkabail, 2018).



Gambar 1. Kerangka Konseptual Untuk Klasifikasi Tutupan Lahan

Makalah ini bertujuan untuk mengeksplorasi akurasi penggunaan GEE cloud computing platform saat mengklasifikasikan citra satelit Landsat 8 untuk tutupan lahan. Analisis spasial dilakukan menggunakan peta tutupan untuk mengidentifikasi luas pertanian serta memberikan informasi terkait jenis tutupan lahan mana yang dapat menjadi perluasan areal pertanian. Pendekatan yang dilakukan adalah dengan menerapkan algoritma RF sehingga berpotensi meningkatkan kinerja prediktif secara keseluruhan. Oleh karena itu, kami mengusulkan algoritma RF menggunakan Google Earth Engine *cloud computing platform* untuk mengklasifikasikan tutupan lahan dari citra satelit Landsat 8.

METODE

Penelitian ini menggunakan metode Random Forest (RF) yang tersedia pada algoritma data processing Google Earth Engine untuk memberikan solusi permasalahan Big data remote sensing pada klasifikasi tutupan lahan. Kami melakukan pemrosesan dan komputasi data citra satelit menggunakan GEE yang memungkinkan

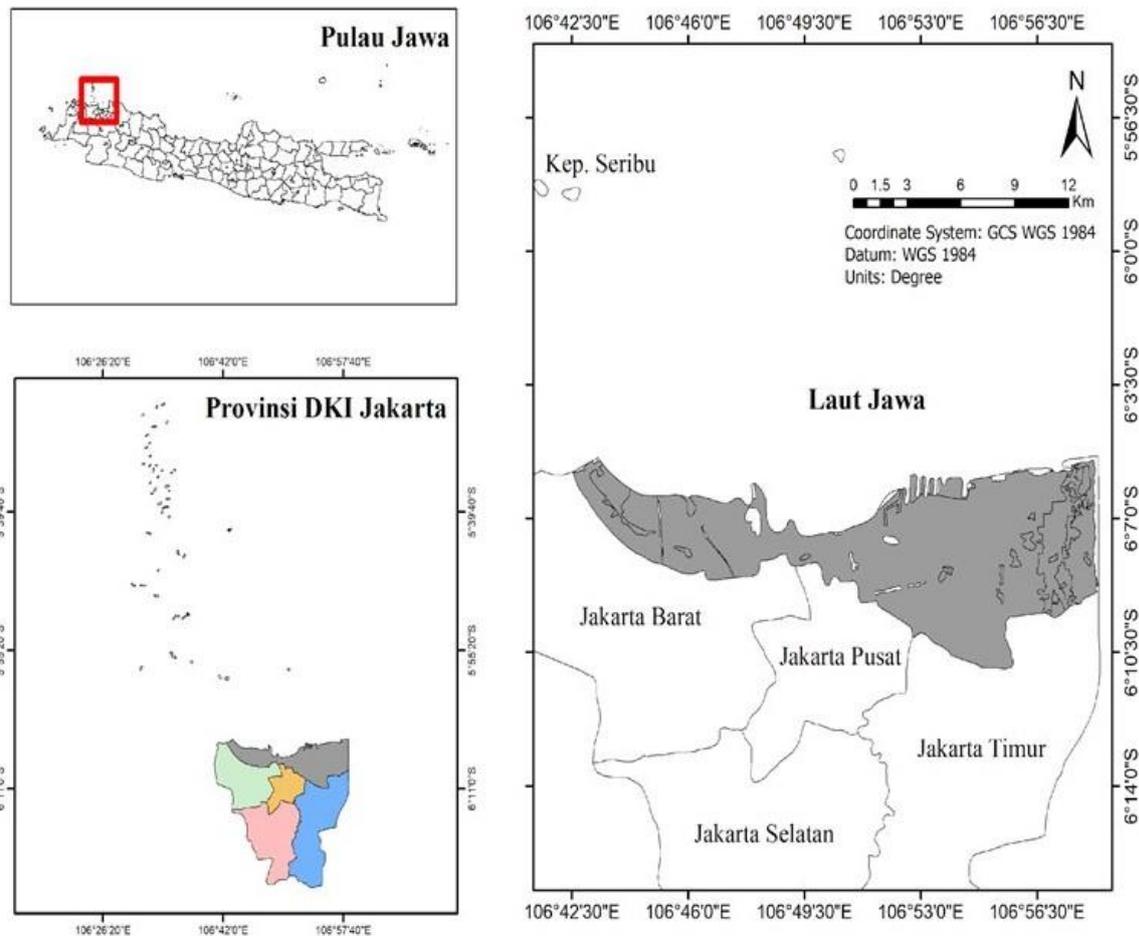
pengolahan secara cloud dan paralel di server Google dengan lancar dan cepat (Teluguntla & Thenkabail, 2018). Gambar 1. menunjukkan kerangka konseptual untuk klasifikasi tutupan lahan.

Studi Area

Jakarta Utara adalah sebuah kota administrasi di bagian utara Daerah Khusus Ibukota Jakarta. Kota ini berada di path/row 122/064 pada citra Landsat 8 *Operational Land Imager* (OLI). Secara geografis, Kota Jakarta Utara terletak pada 1060 20' 00" Bujur Timur 060 10' 00" Lintang selatan. Kota Jakarta Utara berbatasan dengan Laut Jawa di sebelah utara, Kota Jakarta Timur dan Kabupaten Bekasi, Jawa Barat di sebelah timur, Kota Jakarta Barat, Kota Jakarta Pusat dan Kota Jakarta Timur di sebelah selatan, serta Kota Jakarta Barat dan Kabupaten Tangerang, Banten di sebelah barat (lihat Gambar 2).

Data

Data yang digunakan dalam penelitian ini adalah data peta penutupan (tutupan) lahan, Citra satelit Landsat 8 OLI Januari-Desember 2019 dan SHP batas provinsi serta kabupaten dari Badan Pusat



Gambar 2. Letak Geografis dan Studi Area Kota Jakarta Utara

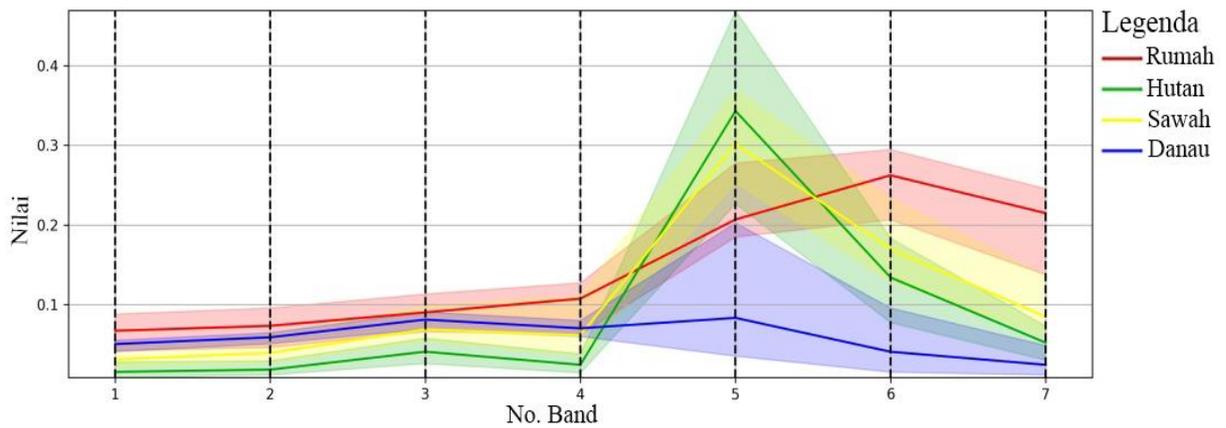
Statistik (BPS) tahun 2019. Peta tutupan lahan dari Kementerian Lingkungan Hidup dan Kehutanan Indonesia (KLHK) bekerjasama dengan Lembaga Antariksa dan Penerbangan Nasional (LAPAN), Badan Informasi Geospasial (BIG) dan

Balai Pemantapan Kawasan Hutan (BPKH) menghasilkan 5 kelas tutupan lahan yang diberi label sebagai lahan pertanian, perairan, hutan/rumput/semak, pemukiman dan area kosong atau lahan terbuka.

Tabel 1. Jenis dan kegunaan band Landsat 8 (yang digunakan pada penelitian)

No band	Nama band	Panjang gelombang (μm)	Aplikasi Band
1	Coastal/Aerosol	0,43 - 0,45	Studi pesisir dan aerosol
2	Biru	0,45 - 0,51	Pemetaan batimetri, membedakan tanah dari vegetasi
3	Hijau	0,53 - 0,59	Menekankan puncak vegetasi, yang berguna untuk menilai kekuatan tanaman
4	Merah	0,63 - 0,67	Membedakan lereng vegetasi
5	NIR	0,85 - 0,88	Menekankan konten biomassa dan garis pantai
6	SWIR 1	1,57 - 1,65	Diskriminasi kadar air tanah dan tumbuh-tumbuhan, menembus awan tipis
7	SWIR 2	2,11 - 2,29	Peningkatan kadar air tanah dan vegetasi dan penetrasi awan tipis

Sumber: Landsat 8 (L8) Data User Handbook, USGS 2018 (<https://www.usgs.gov/media/files/landsat-8-data-users-handbook>)



Gambar 3. Grafik Reflektansi band landsat 8 di Kota Jakarta Utara

Data arsip Landsat 8 pada GEE

Landsat 8 adalah salah satu satelit yang diluncurkan untuk melakukan pemantauan bumi. Satelit ini terdiri dari 9 sensor *Operational Land Image* dan 2 sensor *Thermal Infrared Sensor* (TIRS). OLI memiliki sembilan kelas band yang beroperasi pada kisaran panjang gelombang 0,433-2,300 μm dan memberikan gambar dengan resolusi maksimum 15 m (USGS 2018). Setiap spektral band dalam citra satelit akan menghasilkan nilai reflektansi yang berbeda antar lokasi. Misalkan pada lahan laut, nilai reflektansi pada semua kanal band berada pada nilai yang rendah. GEE merupakan inovasi teknologi geospasial yang menyediakan akses online untuk data Landsat 8.

Setiap band dalam citra satelit akan menghasilkan nilai reflektansi yang berbeda antar lokasi. Contohnya nilai reflektansi pada lahan laut, nilai semua kanal band berada pada nilai yang rendah (lihat Gambar 3).

Indeks Vegetasi

Indeks Vegetasi atau Vegetation Indices (VI) adalah pengukuran optis tingkat kehijauan vegetasi dari informasi reflektansi gelombang elektromagnetik dari kanopi menggunakan sensor pasif (Xue & Su, 2017). Beberapa indeks vegetasi yang diturunkan dari citra satelit untuk mendeteksi tutupan lahan diantaranya adalah Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Normalized Difference Built-Up

Index (NDBI), dan Normalized Difference Water Index (NDWI). ρ_{NIR} , ρ_{RED} , ρ_{BLUE} dan ρ_{SWIR} mewakili reflektansi pada gelombang untuk pita masing-masing. Data NDVI, EVI, NDBI dan NDWI dapat didefinisikan:

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}$$

$$EVI = 2,5 \frac{\rho_{NIR} - \rho_{RED}}{(1 + \rho_{NIR} + 6\rho_{RED} - 7,5\rho_{BLUE})}$$

$$NDBI = \frac{\rho_{SWIR} - \rho_{NIR}}{\rho_{SWIR} + \rho_{NIR}}$$

$$NDWI = \frac{\rho_{NIR} - \rho_{SWIR}}{\rho_{NIR} + \rho_{SWIR}}$$

dimana:

ρ_{NIR} = nilai band *near infrared* (NIR)

ρ_{RED} = nilai band merah

ρ_{BLUE} = nilai band biru

ρ_{SWIR} = nilai band *shortwave infrared* (SWIR)

Cloud Computing Platform

Pengolahan yang digunakan dalam penelitian ini menggunakan cloud computing platform yaitu Google Earth Engine yang meliputi pengolahan citra satelit Landsat 8 OLI dengan periode waktu tahun 2019. Landsat 8 mempunyai kemampuan untuk mengidentifikasi kelas tutupan lahan yang memiliki resolusi spasial 30x30 meter.

Google Earth Engine (GEE) adalah portal web yang menyediakan citra satelit deret waktu global (lebih dari 40 tahun), *cloud computing*, dan algoritma untuk

memproses data (Mutanga & Kumar, 2019). Data yang tersedia berasal dari beberapa satelit, seperti: Moderate Resolution Imaging Spectrometer (MODIS); National Oceanographic and Atmospheric Administration Advanced resolusi sangat tinggi radiometer (NOAA AVHRR); Landsat 7 dan 8; Sentinel 1, 2, dan 3; Satelit Pengamat Darat Lanjutan (ALOS). Data citra satelit yang digunakan telah terkoreksi secara geometrik dan *radiometric surface reflectance* dari United States Geological Survey (USGS) dan kemudian dengan GEE dilakukan *cloud masking* agar data terbebas dari awan.

Random Forest Algorithms

Random forest (Breiman, 2001) merupakan pengembangan CART dengan menerapkan bagging dan random feature selection pada algoritma Decision Tree Machine Learning, yaitu memilih secara acak beberapa fitur dalam setiap iterasinya.. Pohon yang dihasilkan sebanyak iterasinya sehingga menyerupai hutan. Keputusan klasifikasi diambil dari suara terbanyak (*majority vote*) diantara semua pohon (Sammut & Webb, 2017). Alasan penelitian ini memanfaatkan Random Forest (RF) dalam penginderaan jauh adalah untuk menangani data berdimensi tinggi (*high data dimensionality*), multikolinearitas dan juga karena tidak sensitif terhadap overfitting (Belgiu & Drăguț, 2016). Pembagian data untuk training dan testing adalah 70% data training dan 30% data testing.

Evaluasi Kinerja Model

Kinerja prediksi dievaluasi menggunakan fungsi akurasi, presisi, dan sensitivitas. *Confusion matrix* merupakan ringkasan dari hasil prediksi dan ukuran kinerja untuk masalah klasifikasi (lihat Tabel 2). Jumlah prediksi benar dan salah untuk setiap kelas dirangkum di mana nilai

(AP, AN) mewakili data tes positif dan negatif, dan nilai (PP, PN) mewakili hasil prediksi untuk kelas positif dan negatif (Visa, *et al.*, 2011).

Persamaan fungsi kriteria pengukuran kinerja model adalah sebagai berikut:

$$\text{akurasi} = \frac{BP + BN}{BP + BN + SP + SN},$$

$$\text{Presisi} = \frac{BP}{BP + SP},$$

$$\text{Sensitifitas} = \frac{BP}{BP + SN}.$$

Statistik Kappa (Viera & Garrett, 2005) digunakan untuk memberikan ukuran kuantitatif dari besarnya kesepakatan antara pengamat (rater) dan dapat disebut dengan *Inter-Rater reliability* (IRR). Koefisien Kappa dapat melihat tingkat kesepakatan (agreement) antar pengamat dalam menilai setiap indikator pada instrumen. IRR akan memberikan gambaran berupa skor tentang sejauh mana tingkat kesepakatan yang diberikan pengamat. Persamaan statistik kappa adalah sebagai berikut:

$$\kappa = \frac{\sum_{i=1}^v p_{ii} - \sum_{i=1}^v p_{i+} p_{+i}}{1 - \sum_{i=1}^v p_{i+} p_{+i}},$$

dimana:

κ = Statistik Kappa dengan kategori: buruk ($\kappa < 0,4$), cukup ($0,4 < \kappa < 0,6$), baik ($0,6 < \kappa < 0,75$), sangat baik ($\kappa > 0,75$)

$\sum_{i=1}^v p_{ii}$ = Proporsi kesepakatan teramati

$\sum_{i=1}^v p_{i+} p_{+i}$ = Proporsi kesepakatan teramati.

F1 Score adalah rata-rata harmonik dari presisi dan sensitifitas. Persamaan F1 score adalah sebagai berikut:

$$F_1 = 2 \frac{\text{presisi} \times \text{sensitifitas}}{(\text{presisi} + \text{sensitifitas})},$$

Tabel 2. *Confusion Matrix*

	Aktual Positif (AP)	Aktual Negatif (AN)
Prediksi Positif (PP)	Benar Positif (BP)	Salah Positif (SP)
Prediksi Negatif (PN)	Salah Negatif (SN)	Benar Negatif (BN)

HASIL DAN PEMBAHASAN

Setiap *band* pada citra satelit akan menghasilkan nilai reflektansi yang berbeda antar lokasi pada setiap koordinat. Peta grafik reflektansi *band* dibangun sebagai informasi awal untuk mendukung klasifikasi tutupan lahan. Koordinat AOI (*Area of Interest*) secara acak ditentukan untuk objek perairan (Kelas 0), pertanian (Kelas 1), permukiman (Kelas 2), hutan (Kelas 3), dan lahan terbuka (Kelas 4) di Kota Jakarta Utara. Grafik Karakteristik Spektral memperlihatkan ciri khas yang dimiliki dari masing-masing pita dalam menangkap gelombang pantul dari masing-masing objek tutupan lahan. Kurva spektral pada objek vegetasi pertanian dan non-pertanian menunjukkan bahwa reflektansi objek pertanian lebih tinggi daripada reflektansi objek non-pertanian pada pita 4 sampai 7. Kurva spektral dengan objek air menunjukkan kemiringan yang hampir sama di setiap pita (lihat Gambar 3).

Tabel 3 menampilkan hasil dalam hal akurasi, presisi, dan recall. Akurasi atau *Overall Accuracy* (OA) menggunakan RF untuk model klasifikasi tutupan lahan sangat baik yaitu 93,37%. Terlihat bahwa

ketepatan model dalam memprediksi kelas lahan terbuka hanya sebesar 69,90% dan sensitivitasnya juga hanya 71,35%. Hal ini menunjukkan bahwa model ini tidak dapat secara efektif memprediksi Kelas lahan terbuka.

Sedangkan presisi dan sensitivitas untuk kelas lain seperti Kelas pertanian mendapatkan nilai akurasi yang sangat baik yaitu 90,80% dan 89,27%. Nilai tersebut menunjukkan bahwa model dapat dengan sangat baik mengklasifikasikan Kelas pertanian dalam kasus klasifikasi tutupan lahan di Kota Jakarta Utara. Tabel 4 menunjukkan bahwa Kelas 1, Kelas 2 dan Kelas 3 dapat dipetakan dengan *Producer's Accuracy* yang sangat baik. Hal ini menunjukkan bahwa model dapat mengklasifikasikan tutupan lahan secara efektif dengan kebenaran klasifikasi sebesar 90,80%, 96,77% dan 93,92%. Statistik kappa (κ) untuk model sebesar 0,86, yang artinya memiliki kategori yang sangat baik karena $\kappa > 0,75$ dan F1 Score sebesar 86,24%.

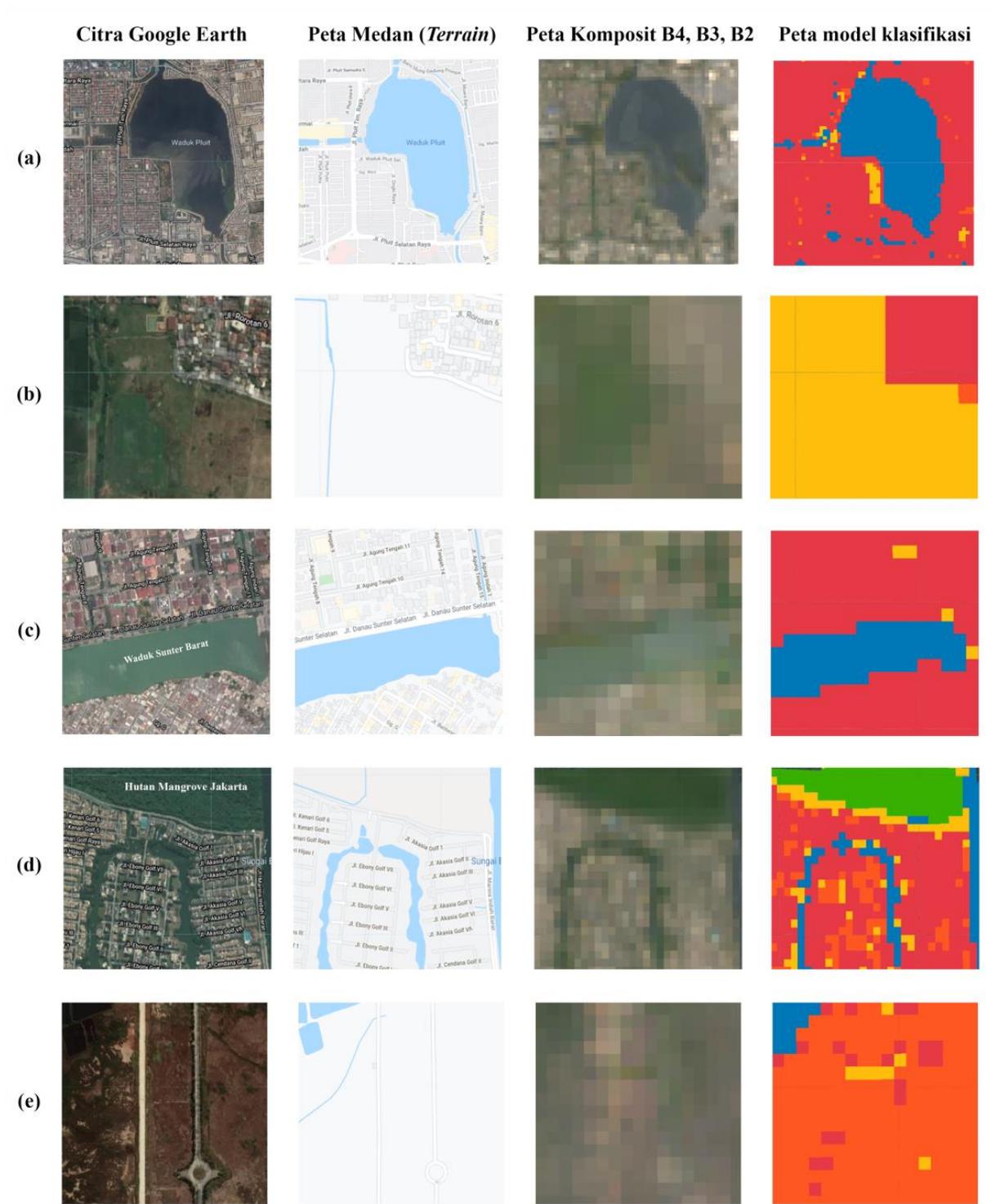
Berdasarkan interpretasi visual dan perbandingan dengan citra satelit resolusi

Tabel 3. Performa prediksi hasil klasifikasi tutupan lahan dengan Random Forest Algorithm.

Algoritma	Kelas	Akurasi (%)	Presisi (%)	Sensitivitas (%)
Random forest algorithm	Perairan (0)		77,59%	84,67%
	Pertanian (1)		90,80%	89,27%
	Permukiman (2)	93,37%	96,77%	96,46%
	Hutan (3)		93,92%	92,05%
	Lahan terbuka (4)		69,90%	71,35%

Tabel 4. *Confusion table* dari prediksi hasil klasifikasi tutupan lahan dengan Random Forest Algorithm.

	Kelas Aktual					<i>Producer's Accuracy</i> (PA)	
	0	1	2	3	4		
Kelas Prediksi	0	232	25	41	0	1	77,59%
	1	38	790	29	12	1	90,80%
	2	1	55	3268	0	53	96,77%
	3	3	6	0	139	0	93,92%
	4	0	9	50	0	137	69,90%
<i>User's Accuracy</i> (UA)	84,67%	89,27%	96,46%	92,05%	71,35%		
Akurasi dan F1 Score Statistik Kappa						93,37% dan 86,24% 0,86	



Gambar 4. Peta Citra pada Google Earth, Peta Medan (*Terrain*), Peta dari Komposit Band 4, 3 dan 2 serta Peta hasil Model Klasifikasi menggunakan algoritma RF pada GEE platform di Kota Jakarta Utara dari (a) Kelas perairan, (b) Kelas pertanian, (c) Kelas permukiman, (d) Kelas Hutan, dan (e) Kelas lahan terbuka.

tinggi, metodologi pemrosesan Geo Big Data yang diusulkan mampu menggambarkan kelas secara baik dan memberikan gambaran visual yang akurat dari hampir semua kelas. Peta tutupan lahan

Kota Jakarta Utara diilustrasikan pada Gambar 4.

Kami membuat perbandingan antara output peta di GEE platform dengan hasil klasifikasi model untuk melihat akurasi

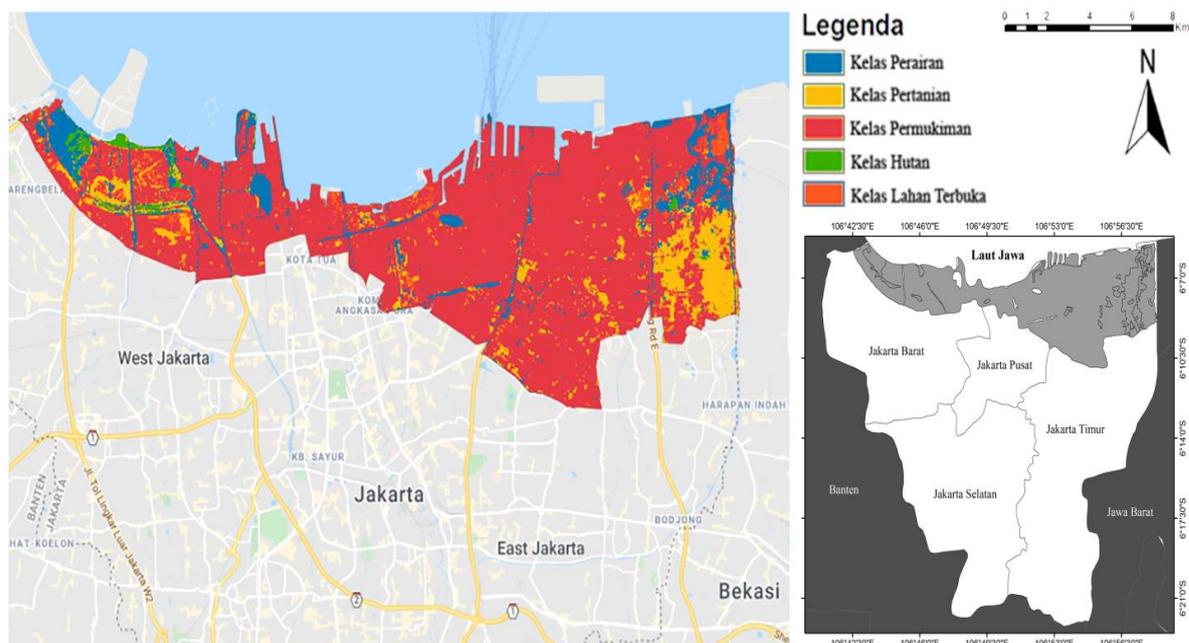
secara visual pada 5 kelas tutupan lahan. Setiap Kelas dipilih satu area dari area studi yang dapat memperlihatkan kesesuaian tampilan citra dengan hasil peta model klasifikasi. Misalnya, area pada Gambar 4(a) adalah area perairan yang merupakan Kelas perairan. Pada kelas tersebut kami memilih area disekitar Waduk Pluit. Visualisasi warna digunakan sebagai keterwakilan Kelas pada peta hasil model klasifikasi (lihat Gambar 5). Peta model klasifikasi untuk Kelas perairan pada Gambar 4(a) telah mendeteksi reflektansi dari permukaan air dan juga reflektansi dari bangunan di sekitar waduk tersebut. Walau ada beberapa noise di area tertentu di Kelas perairan namun secara akurasi sebesar 77,59%. Model yang dibangun juga dapat mengklasifikasikan Kelas Hutan dimana area Hutan Mangrove yang berada di Utara area studi dapat diidentifikasi dengan baik yang ditunjukkan oleh Gambar 4(b). Akurasi untuk Kelas Hutan sebesar 69,90%.

Gambar 4(c) menunjukkan area permukiman yang dipilih di sekitar Waduk Sunter Barat. Peta medan (*terrain*) pada Kelas permukiman memperlihatkan bahwa daerah perairan dan daratan dapat diidentifikasi, namun tidak terlihat jelas pada visualisasi *Natural Map* (peta komposit Band 4, 3 dan 2) Kelas

permukiman. Peta model klasifikasi untuk Kelas permukiman dapat memperlihatkan secara visual batas klasifikasi antara permukiman (ditampilkan dengan legenda warna merah) dan perairan yang merupakan Waduk Sunter Barat (ditampilkan dengan legenda warna biru).

Kelas pertanian diperlihatkan pada Gambar 4(b). Visualisasi yang dihasilkan untuk Kelas pertanian sangat baik. Hal ini menunjukkan model tutupan lahan dengan algoritma Machine Learning Random Forest dapat mengklasifikasikan Kelas pertanian dengan baik dan benar sebesar 90,80%. Klasifikasi yang baik untuk Kelas pertanian pada tutupan lahan sangat penting karena Kelas pertanian sangat berguna sebagai dasar untuk perhitungan luas baku sawah untuk menghasilkan satu data pertanian sehingga dapat mewujudkan ketahanan pangan, perbaikan gizi, dan mendorong budidaya pertanian berkelanjutan dan dapat menjawab tujuan ke-2 SDGs yaitu Zero Hunger.

Gambar 5. Menampilkan gambaran visual secara keseluruhan peta model klasifikasi tutupan lahan Kota Jakarta Utara. Secara umum, klasifikasi yang dihasilkan untuk mengidentifikasi bagian perairan dapat dengan baik dilakukan untuk beberapa danau, waduk, tambak serta aliran



Gambar 5. Peta tutupan lahan menggunakan GEE platform dari Landsat 8 di Kota Jakarta Utara

sungai di daerah Kota Jakarta Utara. Kelas permukiman juga dapat dengan baik diklasifikasikan dengan akurasi terbesar dari semua kelas yaitu 96,77%. Pada bagian utara, model dapat mengklasifikasikan Hutan bakau yang memang ada di Kota Jakarta Utara. Sedangkan untuk Kelas lahan terbuka, akurasi yang didapatkan tidak terlalu baik yang mungkin terjadi karena data training yang dipilih tidak mewakili atau mungkin datanya bisa bersifat tidak seimbang. *Data imbalanced* yang terjadi terlihat dari jumlah data pada setiap kelas tidak sama (Lihat Tabel 4).

KESIMPULAN

Secara ringkas dapat disimpulkan bahwa Random Forest Algorithm merupakan teknik yang sangat baik untuk mengatasi masalah klasifikasi data pada kasus tutupan lahan di Kota Jakarta Utara. Berdasarkan hasil penelitian, algoritma klasifikasi Random Forest memiliki kinerja dalam hal persentase akurasi untuk klasifikasi tutupan lahan dengan akurasi, statistik kappa dan F1 score (93,37%, 0,86 dan 86,24%). Model tersebut juga berhasil mengklasifikasikan Kelas 1 yaitu Kelas pertanian, Kelas 2 yaitu Kelas permukiman dan Kelas 3 yaitu Kelas Hutan memiliki *Producer's Accuracy* diatas 90%. Hasil penelitian menunjukkan bahwa metode yang kami usulkan mencapai kinerja terbaik dalam hal klasifikasi tutupan lahan untuk mendukung data ketahanan pangan dalam berkontribusi pada pencapaian tujuan SDGs. Oleh karena itu, pengembangan penelitian dengan menggunakan pendekatan Algoritma *Machine Learning* lain dan jenis citra yang berbeda, seperti Sentinel 2 merupakan langkah penting berikutnya dalam mengklasifikasikan *Geo Big Data* dengan lebih tepat. Metode lain, seperti SVM dan NN, dapat digunakan untuk meningkatkan kinerja model.

DAFTAR PUSTAKA

- Azar, R., Villa, P., Stroppiana, D., Crema, A., Boschetti, M., & Brivio, P. A. (2016). Assessing In-Season Crop Classification Performance Using Satellite Data: A Test Case In Northern Italy. *Eur. J. Remote Sens.* 49, 361-380.
- Belgiu, M., & Drăguț, L. (2016). Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS Journal of Photogrammetry and Remote Sensing* 114, 24–31.
- Breiman, L. (2001). Random forests. *Machine learning*, vol. 45, no. 1, 5–32.
- Congalton, R. G., & Green, K. (2019). Assessing the accuracy of remotely sensed data: principles and practices. CRC Press.
- FAO. (2016). Food and Agriculture Key To Achieving The 2030 Agenda For Sustainable Development. FAO.
- Kussul, N., Lemoine, G., Gallego, F. J., Skakun, S. V., Lavreniuk, M., & Shelestov, A. Y. (2016). Parcel-Based Crop Classification in Ukraine Using Landsat-8 Data and Sentinel-1A Data. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, 2500-2508.
- Mutanga, O., & Kumar, L. (2019). Google Earth Engine Applications. *Remote Sens.*, 11, 591.
- Omilola, B., & Robele, S. (2017). *The Central Position of Agriculture within the 2030 Agenda for Sustainable Development*. Washington, DC: International Food Policy Research Institute.
- Parsa, I. M., Dirgahayu, D., Manalu, J., & Carolita, I. (2017). Uji Model Fase Pertumbuhan Padi Berbasis Citra Modis Multiwaktu Di Pulau Lombok. *Jurnal Penginderaan Jauh*, 51-64.
- Ramankutty, N., Mehrabi, Z., Waha, K., Jarvis, L., Kremen, C., Herrero, M., le ba bangwe. (2018). Trends in global agricultural land use: implications for environmental health and food security. *Annu. Rev. Plant Biol.* 69, 789–815.
- Sammut, C., & Webb, G. I. (2017). *Encyclopedia of Machine Learning and Data Mining*. New York, AS: Springer.

- Shelestov, A., Lavreniuk, M., Kussul, N., Novikov, A., & Skakun, S. (2017). Exploring google earth engine platform for big data processing: Classification of multi-temporal satellite imagery for crop mapping. *Front. Earth Sci.*, vol. 5, 17.
- Teluguntla, P., & Thenkabail, P. (2018). A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. *ISPRS J. Photogramm. Remote Sens.*, 325-340.
- Thenkabail, P., Lyon, J., & Huete, A. (2019). *Advanced Applications in Remote Sensing of Agricultural Crops and Natural Vegetation*. Boca Raton: CRC Press.
- Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The Kappa Statistic. *Family Medicine*, 37, 360-63.
- Visa, S., Ramsay, B., Ralescu, A., & Van Der Knaap, E. (2011). Confusion matrix-based feature selection. *MAICS*, 710, 120–127.
- Xue, J., & Su, B. (2017). Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *Journal of Sensors*, 1–17.
- You, J., Li, X., Low, M., Lobell, D., & Ermon, S. (2017). Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data. *31th AAAI Conf. Artificial Intelligence*, (dits. 4559–4565).
- Zhang, C., Zhang, H., & Zhang, L. (2018). An Automated Paddy Rice Extent Extraction with Time Stacks of Sentinel Data: A Case Study In Jiangnan Plain, Hubei, China. *7th Int. Conference Agro-Geoinformatics 8*, (dits. 1-6).

