

Teknik *Decision Tree* dalam Pengklasifikasian Penggunaan Lahan Dengan Menggunakan Citra Sentinel-2A MSI

Sri Azizah Nazhifah¹, Andriani Putri²

¹Program Studi Informatika, Universitas Syiah Kuala

²Program Studi Teknologi Informasi, Universitas Teuku Umar

¹Jl. Syech Abdurrauf, Banda Aceh

²Jl. Alue Peunyareng, Aceh Barat

¹sriazizah07@unsyiah.ac.id

²andrianiputri@utu.ac.id

Abstract - Land use classification is an interesting thing to study and research. There are several classification methods that are widely used in some researches, one of which is using machine learning methods, such as SVM, Naive, and decision trees for both aerial photography and satellite imagery. This study uses MSI Sentinel-2A satellite imagery to classify the land use of Kota Langsa as the study area. As for the method used, the first step is downloading the Sentinel-2A image with a band that has a resolution of 10 m and 20 m. This image has an extremely good resolution compared to other optical remote sensing images. For the second step, because the image of the earth's surface is covered by clouds, it is necessary to do a cloud masking process to reduce pixel classification errors. Next, the process of collecting training datasets obtained from each class and represented by pixel values. Then, the training dataset is used to perform a supervised classification using the decision tree method. This method will categorize each pixel into eight classes. As a validation stage, it is necessary to obtain the accuracy of the classification results by using a confusion matrix and calculating the overall accuracy (OA). From the accuracy, it can be concluded that the decision tree can give good results in land use classification of Kota Langsa with 94% of OA value.

Keywords - Decision Tree, Multispectral Instrument, Sentinel-2A, OA, and Supervised Classification.

Abstrak - Klasifikasi penggunaan lahan merupakan suatu hal yang menarik untuk dipelajari dan diteliti. Ada beberapa metode klasifikasi yang banyak digunakan dalam penelitian, salah satunya yaitu menggunakan metode *machine learning*, seperti SVM, Naive, dan *decision tree* baik untuk foto udara dan citra satelit. Penelitian ini menggunakan citra satelit Sentinel-2A MSI dalam melakukan klasifikasi lahan di Kota Langsa sebagai lokasi kajian. Adapun metode yang digunakan, langkah pertama dengan mengunduh citra Sentinel-2A dengan band yang memiliki resolusi 10 m dan 20 m. Citra ini memiliki resolusi yang cukup baik dibandingkan dengan citra *optical remote sensing* lainnya. Langkah kedua, karena citra permukaan bumi yang tertutup awan maka perlu dilakukan proses *cloud masking* untuk mengurangi kesalahan klasifikasi piksel. Selanjutnya, proses pengumpulan *training dataset* yang diperoleh dari setiap kelasnya dan diwakilkan oleh nilai-nilai piksel. Kemudian, *training dataset* tersebut digunakan untuk melakukan klasifikasi terbimbing dengan menggunakan metode pohon keputusan. Metode ini akan mengategorikan setiap pikselnya ke dalam delapan kelas. Terakhir, sebagai tahap validasi, maka perlu dilakukan uji akurasi pada hasil klasifikasi lahan dengan menggunakan *confusion matrix* dan menghitung *overall accuracy* (OA). Dari hasil akurasi tersebut dapat disimpulkan bahwa pohon keputusan dapat memberikan hasil yang baik dalam klasifikasi penggunaan lahan di Kota Langsa dengan nilai OA sebesar 94%.

Kata Kunci - Pohon Keputusan, Multispectral Instrumen, Sentinel-2A, OA, Klasifikasi Terbimbing.

I. PENDAHULUAN

Kota Langsa merupakan salah satu kotamadya dari lima kotamadya yang ada di Provinsi Aceh. Kota Langsa memiliki luas wilayah 7.731,83 Ha dengan luas terbesar berada di Desa Cinta Raja Kecamatan Langsa Timur. Jika dilihat dari peta RTRW Kota Langsa maka dapat dilihat penggunaan lahan tersebut terbagi kepada beberapa bagian. Penggunaan lahan tersebut terdiri dari penggunaan lahan dermaga sungai, gedung/bangunan, padang rumput, pendidikan dasar, perkebunan/kebun, permukiman dan tempat kegiatan, rawa, sawah, sawah tadah hujan, semak belukar, sungai, tambak, tanah kosong/gundul, serta tegalan/ladang. Dari total luas wilayah Kota Langsa,

guna lahan tambak seluas 2.296,26 Ha mendominasi penggunaan lahan di kota tersebut. Hal ini pun tak lepas dari adanya pengembangan wisata hutan mangrove yang terus didorong oleh pemerintah.

Penggunaan lahan sendiri merupakan hal yang ditimbulkan dari campur tangan manusia, baik secara permanen maupun secara sementara terhadap sumberdaya alam dan sumber daya buatan secara keseluruhan dengan tujuan untuk mencukupi kebutuhan-kebutuhannya baik secara material maupun spiritual ataupun dua-duanya [1]. Penggunaan lahan timbul sebagai akibat adanya kebutuhan dari aktivitas hidup manusia. Aktivitas manusia ini berupa tempat tinggal, mata pencaharian, transportasi dan lain-lain. Contohnya untuk daerah

perkotaan biasanya digunakan permukiman, perkantoran, dan industri. Berbeda dengan daerah pedesaan yang biasanya diperuntukan sebagai lahan pertanian, perkebunan, dan peternakan.

Menentukan kuantitas dan kualitas penggunaan lahan sangat diperlukan guna mengukur pemanfaatan dari luasan lahan tersebut bagi masyarakat sekitar. Hal ini dapat dilakukan dengan melakukan klasifikasi penggunaan lahan. Klasifikasi adalah proses penetapan objek-objek, kenampakan atau satuan-satuan menjadi kumpulan-kumpulan, di dalam suatu sistem pengelompokan yang dibedakan berdasarkan sifat-sifat khusus, atau berdasarkan kandungan isinya [2].

Machine Learning adalah salah satu algoritma yang sering digunakan dalam pengklasifikasian data. *Machine Learning* sendiri akan bekerja secara otomatis dan digunakan untuk memprediksi nilai output dari input yang diberikan. Dua tugas utama dari machine learning adalah klasifikasi dan regresi. Maka ada beberapa metode machine learning yang biasa digunakan untuk klasifikasi penggunaan lahan menggunakan citra satelit ataupun foto udara, antara lainnya adalah menggunakan *decision tree*, *Maximum Likelihood Enhance Neighbor classifier*, *Multi Resolution Segmentation (MRS)* dan metode *Support Vector Machine (SVM)* [3].

Metode *decision tree* mengklasifikasikan suatu sampel data yang belum diketahui kelasnya kedalam kelas – kelas yang ada. Penggunaan *decision tree* dapat mengurangi *overfitting* pada sebuah data untuk mencapai akurasi yang maksimal. Metode *Random forest* merupakan salah satu metode dalam *decision tree*. *Random forest* adalah kombinasi dari masing – masing *tree* yang baik kemudian dikombinasikan ke dalam satu model. *Random Forest* bergantung pada sebuah nilai vector random dengan distribusikan ke semua pohon yang masing - masing *decision tree*nya memiliki kedalaman yang maksimal [4].

Adapun citra satelit yang banyak digunakan dalam pengamatan operasional seperti mendeteksi tutupan lahan dan medeteksi perubahan lahan adalah Sentinel. Ada beberapa jenis citra Sentinel, yaitu Sentinel-1 hingga Sentinel-6. Namun pada studi ini akan digunakan Sentinel-2A MSI yang banyak diaplikasikan pada lahan atau air laut. Klasifikasi penggunaan lahan di Kota Langsa dilakukan dengan menggunakan teknik *decision tree* dari *machine learning* atau sering dikenal dengan pohon keputusan. Sebagai tahap evaluasi dari hasil klasifikasi penggunaan lahan, maka akan dihitung *overall accuracy* dengan menggunakan *confusion matrix* dari delapan kelas yang diberikan.

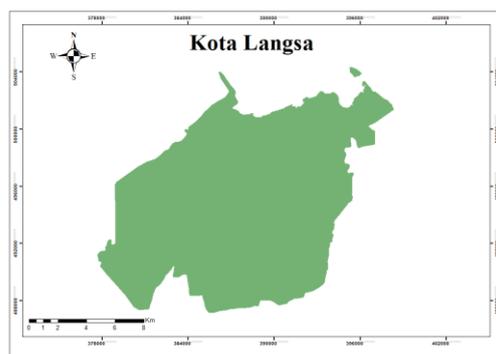
II. METODE PENELITIAN

Data dan Lokasi Penelitian

Data yang digunakan pada penelitian ini adalah citra Sentinel-2A MSI (Multi Spectral Instrumen), karena memiliki resolusi yang tinggi untuk klasifikasi sebuah area. Satelit ini memiliki 13 *spectral band* yang dibagi menjadi 3 jenis spasial resolusinya:

- a. 10 m : B2 , B3, B4 dan B8
- b. 20 m : B5, B6, B7, B8a, B11 dan B12
- c. 60 m : B1, B9 dan B10

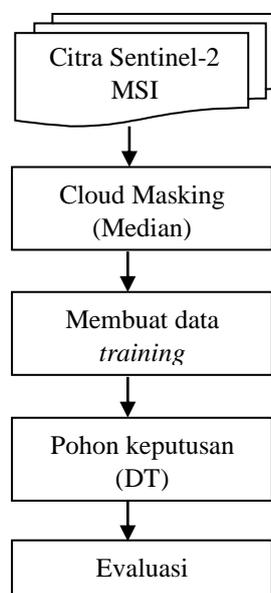
Pada penelitian ini kami menggunakan *band* dengan resolusi spasial 10 m dan 20 m, karean area penelitiannya relative kecil dengan luasan 262,41km², lihat Gambar 1.



Gambar 1. Lokasi Penelitian

Prosedur Penelitian

Penelitian ini menggunakan teknik klasifikasi terbimbing (*supervised classification*) yaitu *decision tree* (DT) dalam mengklasifikasi tutupan lahannya. Adapun tahapan penelitian lebih detail dalam dilihat pada Gambar 2 berikut ini.

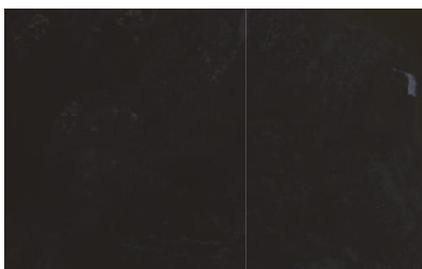


Gambar 2. Tahapan Penelitian

Optical remote sensing image adalah citra yang tidak terlepas dari tutupan awan (*cloud cover*) termasuk citra Sentinel-2 yang digunakan dalam penelitian ini. Maka perlu dilakukan *cloud masking* atau membuang pixel yang mengandung awan karena akan terjadi kesalahan klasifikasi. Setelah pixel yang mengandung awan dihapus maka akan terlihat seperti gambar berikut Gambar 3. Area di dalam *rectangle* adalah contoh pixel yang mengandung awan. Kemudian, area tersebut harus diisi atau ditambah dengan pixel baru, pada penelitian ini kami menggunakan fungsi median filter untuk mengisi area tersebut. Nilai-nilainya diambil dari citra yang pikselnya tidak mengandung awan. Maka hasilnya akan seperti Gambar 4.



Gambar 3. Citra hasil *cloud masking*



Gambar 4. Citra hasil menggunakan median filter

Tahapan berikutnya, membuat *training dataset* untuk 8 kelas, adapun pembagian kelasnya terdapat pada tabel 1.

Tabel 1. Kelas Tutupan Lahan

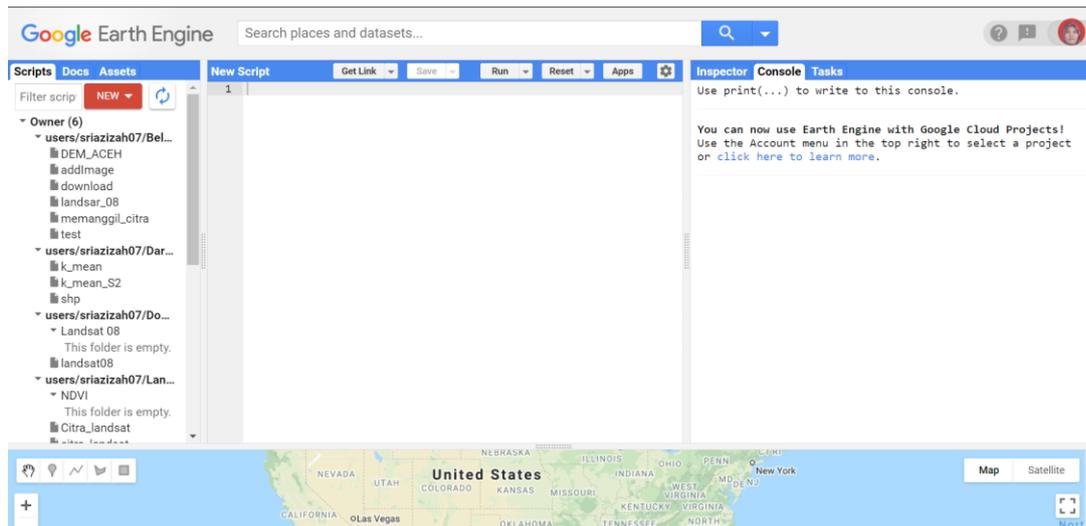
| Code Kelas | Kelas | Keterangan |
|------------|--------------------|---|
| 1 | Bakau | Vegetasi yang tumbuh pada daerah payau |
| 2 | Badan air | Daerah yang memiliki perairan |
| 3 | Tambak | Tempat pemeliharaan ikan dan memiliki pola tertentu |
| 4 | Lahan terbuka | Seluruh lahan yang dapat digunakan pemukiman atau pembangunan |
| 5 | Pemukiman/bangunan | Kawasan permukiman baik perkotaan, pedesaan, pelabuhan, industri dll yang memperlihatkan pola |

| | | |
|---|---------------|---|
| | | alur yang rapat |
| 6 | Semak belukar | Vegetasi alami heterogen |
| 7 | Sawah | Tanah yang digarap dan diairi untuk tempat menanam padi |
| 8 | Sawit | Pohon yang menyerupai kelapa |

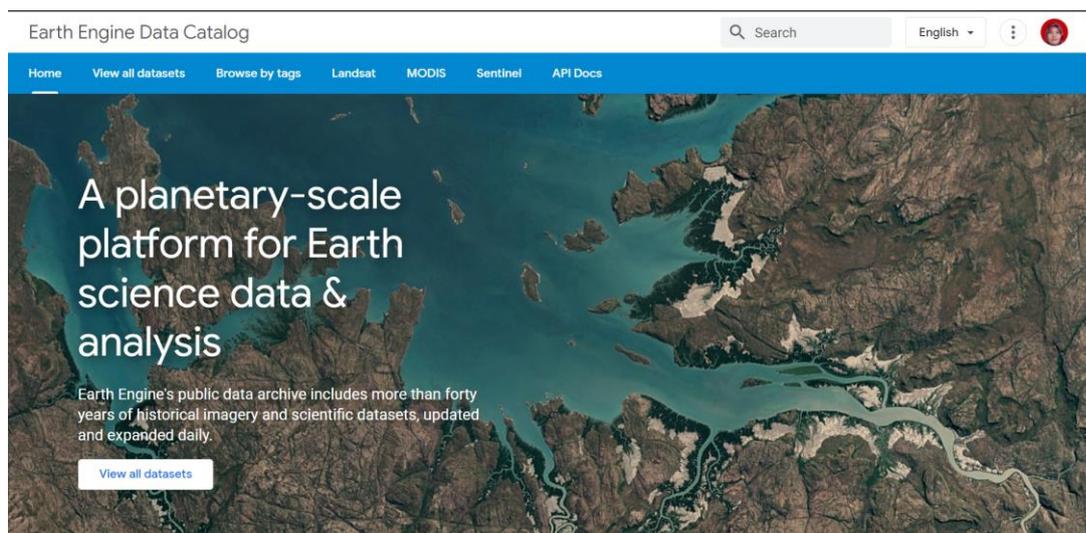
Training dataset yang sudah dibuat berdasarkan kelas-kelas di atas kemudian di gunakan sebagai inputan mengklasifikasi piksel-piksel pada citra Sentinel-2 dengan menggunakan *classifier* DT atau disebut juga dengan pohon keputusan. DT dapat melakukan klasifikasi lahan terhadap citra multispektral dan kemudian dapat menghasilkan hasil akurasi yang baik [5]. Metode ini juga digunakan pada citra Sentinel-2A yang memanfaatkan nilai indeks vegetasi sebagai pemisah antar kelasnya [6]. Selain itu, pohon keputusan juga dapat melakukan klasifikasi berdasarkan karakteristik dari bentuk dan warna [7].

Pohon keputusan adalah metode yang mengubah data menjadi sebuah *decision tree*, agar data dapat lebih mudah diinterpretasikan [8], dalam penelitian ini data yang dimaksud adalah piksel-piksel yang sudah dikumpulkan di dalam *training dataset*. Adapun prosedur dari pada pohon keputusan ini [9] adalah 1) *training dataset*, dimana masing-masing kelas memiliki nilai piksel yang berbeda-beda dengan labelnya masing-masing 2) sejumlah training dataset membagi kelas-kelas ke dalam subsetnya masing-masing 3) Membuat kriteria yang dapat melakukan penghentian dalam pengembangan pohonnya..

Proses klasifikasi pohon keputusan dilakukan dengan membangun algoritma yang diterapkan dalam bahasa pemrograman javascript. Bahasa ini ditulis di dalam Google Earth Engine (GEE). GEE merupakan *platform* yang dikembangkan oleh Google untuk pengolahan data *remote sensing*. Data yang diperoleh dengan sistem *remote sensing* telah berkembang sangat pesat, maka dari itu pihak Google mengembangkan sebuah *cloud computing platform* yang dikenal dengan GEE. *Platform* ini mempermudah penggunaannya untuk dalam proses pengolahan datanya, *user* tidak perlu mengunduh citra satelit secara terpisah, adapun citra satelit yang dapat digunakan pada *platform* ini adalah MODIS, Landsat, Sentinel Series. *User* dapat menuliskan program untuk pemanggilan citra yang diinginkan [10]. Selanjutnya, pada saat ingin menggunakan sebuah *classifier*, *user* cukup menggunakan *classifier* DT. Lihat gambar Gambar 6 tampilan workspace GEE dan Gambar 7 data *catalog* GEE.



Gambar 6. Workspace GEE



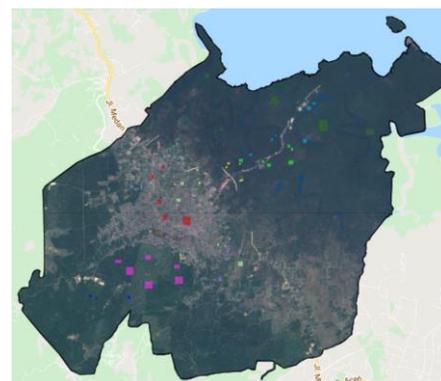
Gambar 7. Data Catalog GEE

Selanjutnya dilakukan uji akurasi terhadap hasil klasifikasinya. Proses uji akurasi menggunakan sample yang diperoleh dari citra GEE, dimana citra ini memiliki resolusi yang sangat baik atau tinggi. Sehingga peneliti dapat membedakan mana pemukiman, lahan terbuka, tambak, badan air dan kelas lainnya dari sampel yang dikumpulkan.

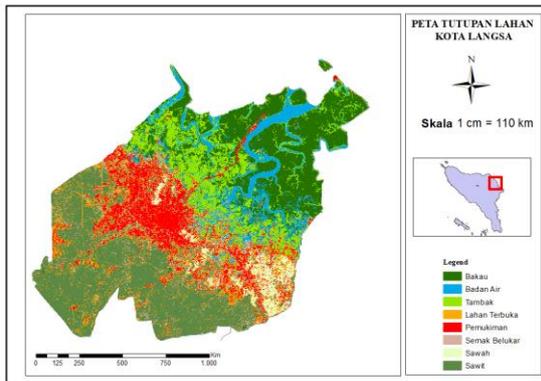
III. HASIL DAN PEMBAHASAN

Kota Langsa merupakan pemekaran dari beberapa kabupaten, sehingga luas areanya tidak begitu besar. Dari segi topografi kota ini juga memiliki daerah perbukitan dan perkebunan, seperti perkebunan sawit. Ada beberapa daerah yang memiliki lahan sawah yang berdampingan dengan area pemukiman.

Hasil klasifikasi yang diperoleh dengan menggunakan citra Sentinel-2 MSI dapat dilihat pada Gambar 8. Gambar tersebut menunjukkan data training yang dibuat dengan mendigitasi citra satelit terlebih dahulu untuk setiap classnya.



Gambar 8. Data training



Gambar 9. Tutupan lahan Kota Langsa

Gambar diatas didapatkan dari hasil metode *supervised classification* yaitu dengan menggunakan *classifier cart*, potongan program yang digunakan dapat dilihat dalam Gambar 10.

```
var classifier = ee.Classifier.smileCart().train({
  features: training,
  classProperty: 'lc',
  inputProperties:
  bands
});
```

Gambar 10. Javascript code classifier decision tree

Kemudian menguji hasil klasifikasi dengan mengumpulkan sampel pada setiap kelas yang dimasukkan ke dalam *confusion* matrik. Berikut adalah tabel 2 *confusion* matrik untuk hasil klasifikasi dan data sample.

Tabel 2. Confusion matrik

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Total |
|---|-----|-----|-----|-----|----|----|-----|-----|-------|
| 1 | 103 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 103 |
| 2 | 1 | 121 | 4 | 0 | 0 | 0 | 0 | 0 | 126 |
| 3 | 0 | 10 | 100 | 0 | 0 | 0 | 0 | 0 | 110 |
| 4 | 0 | 0 | 0 | 96 | 14 | 0 | 3 | 0 | 113 |
| 5 | 0 | 0 | 0 | 3 | 80 | 0 | 1 | 0 | 84 |
| 6 | 0 | 0 | 0 | 3 | 0 | 87 | 0 | 2 | 92 |
| 7 | 0 | 0 | 0 | 2 | 1 | 0 | 104 | 0 | 107 |
| 8 | 1 | 0 | 0 | 2 | 0 | 2 | 0 | 98 | 103 |
| | 105 | 131 | 104 | 106 | 95 | 89 | 108 | 100 | 838 |

Piksel Benar
 Piksel Salah

$$Overall Accuracy (OA) = \frac{Total\ Piksel\ Benar}{Total\ Keseluruhan\ piksel}$$

$$OA = 789 / 838 = 0,94 \times 100 = 94 \%$$

Dari formula diatas maka didapatkan OA senilai 94%, maka hasil klasifikasi menggunakan supervised classification dengan pohon keputusan adalah benar dan model ini dapat digunakan untuk proses penelitian

lebih lanjut dan dapat dijadikan referensi dalam pengelolaan tata ruang Kota Langsa.

IV. KESIMPULAN

Berdasarkan penelitian, implementasi dan pengujian, maka dapat diambil kesimpulan bahwa supervised classification dengan menggunakan metode CART atau yang lebih dikenal dengan *decision tree* dapat melakukan klasifikasi tutupan lahan terhadap citra satelit Sentinel-2 MSI yang memiliki resolusi 10 m dan 20 m. Model yang dihasilkan menunjukkan performa yang baik dengan nilai akurasi 94%.

DAFTAR PUSTAKA

- [1] M. D. Kusumaningrat, S. Subiyanto, and B. D. Yuwono, "Analisis Perubahan Penggunaan Dan Pemanfaatan Lahan Terhadap Rencana Tata Ruang Wilayah Tahun 2009 Dan 2017 (Studi Kasus: Kabupaten Boyolali)," *Jurnal Geodesi Undip*, vol. 06, no. 4, 2014.
- [2] R.J. Manalu, A. Sutanto, and B. Trisakti "Perbandingan Metode Klasifikasi Penutup Lahan Berbasis Pikel Dan Berbasis Obyek Menggunakan Data Pisar-L2," *Jurnal Penginderaan Jauh*, vol. 13, no. 1, 2016.
- [3] B. van Leeuwen, Z. Tobak, and F. Kovács "Machine Learning Techniques for Land Use/Land Cover Classification of Medium Resolution Optical Satellite Imagery Focusing on Temporary Inundated Areas," *Journal of Environmental Geography*, vol. 13, no. 1-2, pp. 43 – 52, 2020.
- [4] H. Zhang, D. Nettleton, and Z. Zhu, "Classification and Regression by Random Forest," *ResearchGate*, 2019.
- [5] L. Timur, "Perbandingan Klasifikasi Svm Dan Decision Tree Untuk Pemetaan Mangrove Berbasis Objek Menggunakan Citra Satelit Sentinel-2B Di Gili Sulat, Lombok Timur," *J. Nat. Resour.*, vol. 9, no. 3, pp. 746–757, 2019, doi: 10.29244/JPSL.9.3.746-757.
- [6] R. Awaliyan and Y. B. Sulistyoadi, "Klasifikasi Penutupan Lahan Pada Citra Satelit Sentinel-2a Dengan Metode Tree Algorithm," *ULIN J. Hutan Trop.*, vol. 2, no. 2, pp. 98–104, 2018, doi: 10.32522/u-jht.v2i2.1363.
- [3] T. N. Lina *et al.*, "Penerapan Metode Decision Tree Untuk Penentuan," vol. 09, no. 01, pp.

- 10–19, 2017.
- [7] J. Rogan, J. Miller, D. Stow, J. Franklin, L. Levien, and C. Fischer, “Land-cover change monitoring with classification trees using Landsat TM and ancillary data,” *Photogramm. Eng. Remote Sensing*, vol. 69, no. 7, pp. 793–804, 2003, doi: 10.14358/PERS.69.7.793.
- [8] M. Pal, “Factors influencing the accuracy of remote sensing classifications : a comparative study FACTORS INFLUENCING THE ACCURACY OF REMOTE SENSING CLASSIFICATIONS : Thesis submitted to the University of Nottingham for,” no. January 2002, 2014.
- [9] B. Charbuty and A. Abdulazeez, “Classification Based on Decision Tree Algorithm for Machine Learning,” *J. Appl. Sci. Technol. Trends*, vol. 2, no. 01, pp. 20–28, 2021, doi: 10.38094/jastt20165.
- [10] M. Amani *et al.*, “Google Earth Engine Cloud Computing Platform for Remote Sensing Big Data Applications: A Comprehensive Review,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, no. September, pp. 5326–5350, 2020, doi: 10.1109/JSTARS.2020.3021052.