

# Penerapan Metode *K-Nearest Neighbor (K-NN)* untuk Klasifikasi Tutupan Lahan Kabupaten Empat Lawang

Kanaya Dea Thalita Akhmad<sup>1,a,\*</sup>, Farah Faizah<sup>2,b,\*</sup>, Yunaena Maratul Kirom<sup>3,c,\*</sup>, Yosia Adwily Nainggolan<sup>4,d,\*</sup>, Elsyah Sapyrah<sup>5,e,\*</sup>

<sup>1</sup>121450001, <sup>2</sup>121450002, <sup>3</sup>121450080, <sup>4</sup>121450063, <sup>5</sup>121450096

<sup>a</sup>[kanaya.121450001@student.itera.ac.id](mailto:kanaya.121450001@student.itera.ac.id), <sup>b</sup>[farah.121450002@student.itera.ac.id](mailto:farah.121450002@student.itera.ac.id),  
<sup>c</sup>[yunaena.121450080@student.itera.ac.id](mailto:yunaena.121450080@student.itera.ac.id), <sup>d</sup>[yosia.121450063@student.itera.ac.id](mailto:yosia.121450063@student.itera.ac.id),  
<sup>e</sup>[elsyah.121450096@student.itera.ac.id](mailto:elsyah.121450096@student.itera.ac.id)

\*Program studi Sains Data, Fakultas Sains, Institut Teknologi Sumatera

---

## Abstrak

Klasifikasi tutupan lahan merupakan komponen penting dalam berbagai bidang, seperti perencanaan tata ruang, pengelolaan sumber daya alam, monitoring lingkungan, dan mitigasi bencana. Salah satu sumber data yang dapat digunakan untuk mengidentifikasi tutupan lahan adalah citra satelit multispektral. Namun, untuk mengolah data citra satelit menjadi informasi yang berguna diperlukan metode klasifikasi yang tepat. Penelitian ini bertujuan untuk menerapkan metode *K-Nearest Neighbor (KNN)* dalam mengklasifikasikan tutupan lahan khususnya di Kabupaten Empat Lawang yang terletak di Provinsi Sumatera Selatan menggunakan data citra satelit landsat 8. Proses penelitian ini meliputi pengumpulan data, preprocessing, kompilasi data latih, modelling, evaluasi, dan hasil. Hasil penelitian menunjukkan bahwa metode *K-Nearest Neighbor (K-NN)* dapat menghasilkan klasifikasi tutupan lahan dengan akurasi yang sangat baik, dengan nilai *K* optimal pada *K*=5 yang menghasilkan akurasi keseluruhan sebesar 0.09788092835519677 atau sekitar 97.88% dibandingkan set validasi sebesar 0.961. Kelas tutupan lahan yang teridentifikasi meliputi lahan terbuka, sungai, jalan, vegetasi, dan pemukiman. Penerapan metode *K-Nearest Neighbor (K-NN)* terbukti efektif untuk pemetaan dan pemantauan tutupan lahan di masa yang akan mendatang.

**Kata kunci :** Citra Satelit, Klasifikasi, *K-Nearest Neighbor*, Tutupan Lahan.

## I. Pendahuluan

Tutupan lahan merupakan informasi penting dalam berbagai bidang seperti perencanaan tata ruang, pengelolaan sumber daya alam, pengelolaan lingkungan, dan mitigasi bencana. Tutupan lahan paling umum biasanya adalah kategori urban (perkotaan) dan keperluan pertanian [1]. Informasi mengenai jenis dan distribusi tutupan lahan diperlukan untuk memahami dinamika lingkungan dan membantu dalam pengambilan keputusan yang berbasis data. Salah satu cara untuk mendapatkan informasi tersebut adalah melalui klasifikasi tutupan lahan.

Citra satelit multispektral merupakan salah satu sumber data yang sangat berguna dalam pemetaan tutupan lahan. Citra ini mampu merekam informasi di berbagai spektrum elektromagnetik sehingga memungkinkan identifikasi berbagai jenis tutupan lahan berdasarkan karakteristik spektral yang unik dari setiap jenisnya. Beberapa jenis citra multispektral yang umum digunakan diantaranya adalah Landsat, QuickBird dan

IKONOS [2]. Namun, untuk mengolah data citra satelit menjadi informasi yang berguna diperlukan metode klasifikasi yang tepat.

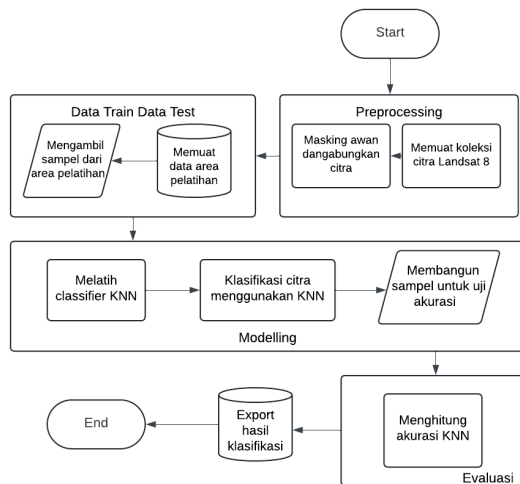
Metode *K-Nearest Neighbor (K-NN)* adalah sebuah metode klasifikasi yang populer terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasi sebelumnya. *K-NN* dilakukan dengan mencari kelompok *k* objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data testing [3].

Dalam konteks klasifikasi tutupan lahan, data geospasial seperti peta penggunaan lahan, data elevasi, dan informasi geografi lainnya dapat digunakan untuk meningkatkan akurasi model *K-NN*. Kombinasi antara citra satelit multispektral dan data geospasial memungkinkan model untuk lebih memahami konteks dan variasi lokal sehingga dapat menghasilkan klasifikasi yang lebih tepat. Dengan menggunakan citra satelit resolusi tinggi dan teknik ekstraksi fitur yang canggih, *K-NN* dapat menghasilkan peta tutupan lahan dengan akurasi yang tinggi.

Di Indonesia, khususnya di Kabupaten Empat Lawang yang terletak di Provinsi Sumatera Selatan, kebutuhan akan informasi tutupan lahan yang akurat sangat penting mengingat daerah ini memiliki keanekaragaman ekosistem dan tantangan lingkungan yang kompleks. Penerapan metode K-NN untuk klasifikasi tutupan lahan dapat membantu dalam pemantauan perubahan lahan, pengelolaan sumber daya alam, serta upaya konservasi dan restorasi lingkungan. Selain itu, informasi tutupan lahan yang akurat juga mendukung perencanaan pembangunan yang berkelanjutan dan mitigasi risiko bencana alam seperti banjir dan tanah longsor.

## II. Metode Penelitian

### 2.1. Diagram Alir



**Gambar 1.** Diagram Alir Penelitian

### 2.2 Pengumpulan Data

Penelitian menggunakan citra satelit Landsat 8 yang diperoleh dari Google Earth Engine (GEE).

### 2.3 Pre-processing Data

Pada preprocessing data dilakukan melalui beberapa langkah untuk memastikan bahwa data yang digunakan dalam klasifikasi tutupan lahan bersih dan siap untuk dianalisis. Berikut langkah-langkah preprocessing data yang dilakukan :

#### 2.3.1 Masking Awan pada Citra Landsat 8

Pada tahap ini digunakan fungsi “maskL8sr” dimana untuk menghilangkan awan dan bayangan awan dari citra Landsat 8 sebelum digunakan dalam analisis lebih lanjut. Proses ini penting dilakukan karena dengan adanya awan dan

bayangan awan dapat mengganggu hasil akurasi klasifikasi.

#### 2.3.2 Memuat dan Memfilter Koleksi Citra

Pada tahap ini memuat koleksi citra Landsat 8, memfilter berdasarkan wilayah tertentu (Empat Lawang), dan jangka waktu yang sudah ditentukan (April 2024). Setiap citra kemudian diproses dengan menggunakan fungsi “maskL8sr”.

#### 2.3.3 Menggabungkan Citra menjadi satu Mosaic dan Memilih Band yang Diperlukan

Pada tahap ini menggabungkan citra-citra dalam koleksi menjadi satu citra mosaic dan memotong wilayahnya sesuai dengan yang sudah ditentukan (Empat Lawang). Dan memastikan hanya band yang diperlukan yang dipilih (band spektral yang dibersihkan dari awan dan bayangan awan).

### 2.4 Kompilasi Data Latih

Kompilasi data latih ini adalah proses pengumpulan dan penyusunan data yang akan digunakan untuk melatih model machine learning. Dimana pada tahap ini untuk memastikan bahwa data yang digunakan untuk melatih model klasifikasi tutupan lahan sudah siap dan sesuai dan dalam penelitian ini, pengumpulan data latih yang dilakukan pelabelan diambil dengan cara digitasi data raster dari data kabupaten Empat Lawang dan banyak ukuran pixel yang akan diambil tergantung dari luas wilayah kabupaten yang akan dilakukan klasifikasi nantinya. Dengan melakukan tahap ini, memastikan bahwa dataset pelatihan memiliki variasi yang cukup, fitur-fitur yang relevan, dan label kelas yang konsisten untuk menghasilkan model klasifikasi yang akurat. Berikut langkah-langkah kompilasi data latih :

#### 2.4.1 Menggabungkan Koleksi Fitur untuk Pelatihan

Tahap ini dilakukan untuk menggabungkan berbagai kelas tutupan lahan menjadi satu koleksi fitur untuk membentuk dataset pelatihan. Koleksi fitur yang mewakili berbagai jenis tutupan lahan seperti lahan terbuka, sungai, jalan, vegetasi, dan pemukiman digabungkan menjadi satu. Setiap koleksi fitur berisi titik-titik referensi yang telah diklasifikasikan secara manual atau berdasarkan data lapangan.

#### 2.4.2 Mengambil Sampel dari Area Pelatihan

Tahap ini dilakukan untuk mengambil sampel dari citra mosaic berdasarkan titik-titik referensi (labels) untuk menciptakan dataset pelatihan yang digunakan oleh model. Properti “tutupanlahan” digunakan sebagai kelas dalam dataset pelatihan.

Dengan skala (resolusi) diatur ke 30 meter, yang sesuai dengan resolusi citra Landsat 8.

## 2.5 Model K-Nearest Neighbour

Metode *k-Nearest Neighbor* (K-NN) adalah salah satu metode klasifikasi klasik yang paling sederhana. Seringkali disebut juga sebagai Pembelajaran Berbasis Instance (*Instance-Based Learning*), K-NN melakukan klasifikasi suatu objek berdasarkan jarak antara objek tersebut dengan objek lainnya. Metode ini menggunakan prinsip ketetanggaan (*neighbor*) untuk memprediksi kelas baru, di mana jumlah tetangga yang digunakan untuk prediksi adalah sebanyak  $k$  tetangga.

Setelah memilih  $k$  tetangga terdekat pertama, langkah selanjutnya adalah menghitung jumlah data yang termasuk dalam masing-masing kelas dari  $k$  tetangga tersebut. Kelas dengan jumlah data terbanyak akan menjadi kelas pemenang, kemudian diberikan sebagai label kelas pada data. Pada k-NN, nilai  $k$  memiliki pengaruh signifikan terhadap performa klasifikasi. Jika nilai  $k$  terlalu kecil, hasil klasifikasi mungkin tidak akurat [4].

Metode K-Nearest Neighbor (K-NN) dapat dihitung dengan tahapan berikut :

1. Menentukan parameter K
2. Menghitung jarak antara data training dan data testing

Hitung jarak antara setiap data dalam set pelatihan dengan data uji. Perhitungan jarak yang paling umum digunakan dalam algoritma k-NN adalah jarak Euclidean, dengan rumus sebagai berikut :

$$euc = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \dots\dots\dots (1)$$

dimana :

$p_i$  = sample data / data training

$q_i$  = data uji / data testing

$i$  = variabel data

$n$  = dimensi data

3. Mengurutkan jarak yang terbentuk dari hasil perhitungan euclidean sebelumnya.
4. Tahapan selanjutnya dalam menentukan jarak terdekat sampai urutan K, dalam penelitian kali ini digunakan nilai ketetanggaan K senilai 5. Ini menggambarkan akan membandingkan

sebanyak 5 kelas dari data yang terdekat dari data yang ingin dilakukan klasifikasi. Dan untuk melihat tetangga terdekat dilihat dari jarak euclidean terendah.

5. Data inputan sebagai data latih yang sudah dilakukan label ini memiliki nilai dari masing masing hasil digitasi tersebut. dimana data raster ini akan membaca nilai digital number yang akan membentuk suatu kelas yang telah dilakukan pelabelan tersebut. Nilai DN dari piksel yang diambil dari hasil digitasi manual digunakan sebagai fitur dalam algoritma KNN, yang mana DN ini adalah nilai spektral dari masing masing landsat citra 8 yang digunakan pada penelitian kali ini.
6. Mencari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi pada tahap berikutnya[3]

## 2.6 Pengukuran Performa

Penerapan Metode K-Nearest Neighbor (K-NN) untuk klasifikasi tutupan lahan berfokus pada pengukuran performa dengan melibatkan pemrosesan data, pelatihan model dan evaluasi menggunakan matriks kesalahan dan matriks performa.

class 0	-	-	-
class 1	-	-	-
class 2	-	-	-
actual/pred	class 0	class 1	class 2

Tabel 1. Matrix Pengukuran performa

Dari matrix diatas, didapatkan untuk sumbu x adalah nilai dari prediksi dan sumbu y sebagai nilai aktual yang disajikan dalam matriks performansi hasil yang membuktikan kebenaran antara nilai hasil prediksi dan aktualnya. Langkah-langkah ini memastikan bahwa model tidak hanya belajar dari data pelatihan, tetapi juga dapat menggeneralisasi dengan baik ke data baru, yang sangat penting dalam aplikasi nyata seperti klasifikasi tutupan lahan.

## 2.7 Ekspor Hasil

Hasil yang telah didapatkan dari klasifikasi data nantinya akan diekspor dalam bentuk

visualisasi data vektor yaitu berupa shp, cpg, dbr, fix, prj, shx.

## 2.8 Pseudocode

```
MASUKKAN data input (batas wilayah, rentang
tanggal, koleksi citra, data pelatihan, data uji
akurasi)

PROSES citra:
  MUAT citra sesuai input
  MASKING awan
  GABUNGAN menjadi mosaik
  POTONG sesuai batas wilayah

BANGUN model klasifikasi:
  GABUNGAN data pelatihan
  TENTUKAN band
  AMBIL sampel pelatihan dari citra
  LATIH classifier KNN

KLASIFIKASI citra dengan model

EVALUASI akurasi:
  BANGUN sampel uji akurasi
  HITUNG matriks error dan akurasi

EXPORT hasil klasifikasi (opsional)
```

## III. Hasil dan Pembahasan

### 3.1 Preprocessing

Data citra diambil dari koleksi citra Landsat 8 Collection2 Level 2 (LANDSAT/LC08/C02/T1\_L2). Data citra yang digunakan merupakan data Kabupaten Empat Lawang yang diambil pada rentang waktu 1 April 2024 hingga 30 April 2024.

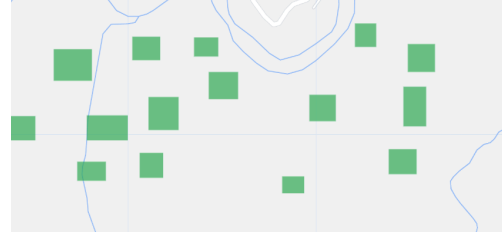
Setelah melakukan penginputan data, dilakukan tahap preprocessing. Pada **Gambar 2**, ditunjukkan hasil dari preprocessing data citra yang telah dilakukan filtering.



**Gambar 2.** Data Landsat di Kabupaten Empat Lawang

### 3.2 Data Latih dan Data Uji

Dilakukan pengambilan data train dan data test dengan melakukan digitasi pada kelas yang akan ditentukan. Dalam tahap ini, dilakukan identifikasi area atau wilayah untuk kelas lahan terbuka, sungai, jalan, vegetasi, dan pemukiman. Perbandingan dari data latih dan data uji adalah 70:30.



**Gambar 3.** Data Train dan Data Test

#### 3.2.1 Perhitungan Piksel

Luas wilayah =

$$2276.4449665297734 \text{ km}^2 \times 1,000,000 \text{ m}^2/\text{km}^2 \\ = 2,276,444,966 \text{ m}^2$$

$$\text{Luas Satu Piksel} = 30 \text{ m} \times 30 \text{ m} = 900 \text{ m}^2$$

$$\text{Jumlah Piksel} = \frac{\text{Luas Wilayah}}{\text{Luas Satu Piksel}} \\ = \frac{2,276,444,966.5297734 \text{ m}^2}{900 \text{ m}^2/\text{piksel}} \\ \approx 2,529,383 \text{ piksel}$$

#### 3.2.2 Perhitungan Sampel

Berdasarkan literatur dan praktik umum, jumlah sampel yang diambil biasanya adalah antara 0,5% hingga 2% dari total piksel.

$$\text{Jumlah sampel minimum} = 0.005 \times 2,529,383 \\ \approx 12,647 \text{ piksel}$$

$$\text{Jumlah sampel optimal} = 0.01 \times 2,529,383 \\ \approx 25,294 \text{ piksel}$$

$$\text{Jumlah sampel maksimal} = 0.02 \times 2,529,383 \\ \approx 50,588 \text{ piksel}$$

Penelitian ini menggunakan 2% dari pixel total, maka perhitungannya

- 2% dari total piksel

$$\text{Jumlah poligon per kelas} = \frac{10,118 \text{ sampel}}{100 \text{ piksel per poligon}} \\ \approx 101 \text{ poligon per kelas}$$

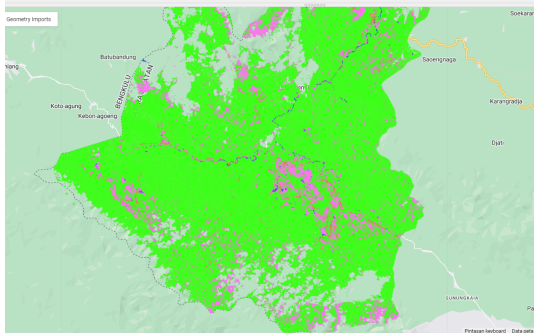
Penggunaan rata-rata disetiap polygon digitasi memiliki rata-rata bernilai 100 piksel, dimana ini menggambarkan jarak sejauh 300 meter jika

dilihat dari peta. yang mana ukuran skala untuk landsat 8 adalah 1:3000 (30 meter).

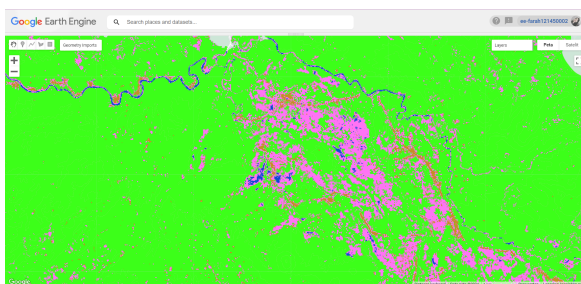
Dalam penelitian kali ini menggunakan jenis 2% dari total pixel yang ada sehingga menggunakan sebanyak 101 data digitasi per kelas. sehingga untuk jumlah total keseluruhan adalah 505 data yang sudah dibandingkan sebanyak 70% data training dan 30% data testing.

### 3.3 Modelling

Metode K-NN dalam klasifikasi tutupan lahan menggunakan algoritma yang mengklasifikasikan setiap piksel atau segmen berdasarkan kemiripannya dengan data sampel atau data latih yang telah dilabeli sebelumnya. Fitur-fitur seperti nilai spektral, tekstur dan informasi spasial digunakan untuk menghitung kemiripan dan mengklasifikasikan ke dalam kelas tutupan lahan yang paling mirip.



**Gambar 4.** Hasil Klasifikasi Tutupan Lahan di Kabupaten Empat Lawang



**Gambar 5.** Hasil Klasifikasi Tutupan Lahan di Kabupaten Empat Lawang yang diperbesar

Berdasarkan **Gambar 5**. Terlihat area berwarna pink yang cukup luas diklasifikasikan sebagai lahan terbuka. Ini mencakup lahan kosong, ladang atau area terbuka lainnya yang tidak tertutup vegetasi atau bangunan. Garis-garis biru memanjang menunjukkan klasifikasi sungai atau badan air lainnya yang mengalir di wilayah tersebut. Area berwarna hijau merupakan

klasifikasi tutupan lahan yang didominasi oleh vegetasi seperti hutan, perkebunan, atau area bervegetasi lainnya. Area coklat yang tersebar di beberapa wilayah lahan terbuka menunjukkan area yang dihuni atau area terbangun dengan bangunan dan infrastruktur perkotaan. Pada pengklasifikasian ini juga terdapat kelas lain yakni jalan yang diwakili warna ungu, tetapi sayangnya pada gambar kurang terlihat jelas karena tertutup oleh area klasifikasi lainnya seperti vegetasi dan lahan terbuka.

### 3.5 Evaluasi

Dalam melakukan evaluasi kali ini menggunakan matrix dari Validation error matrix dan KNN Error Matrix dalam mengevaluasi model.

<b>Class 0</b>	2	0	0	0	0
<b>Class 1</b>	11	30	0	0	0
<b>Class 2</b>	0	0	0	0	1
<b>Class 3</b>	22	0	0	1635	0
<b>Class 4</b>	35	0	0	2	82
<b>0,961</b>	<b>C0</b>	<b>C1</b>	<b>C4</b>	<b>C3</b>	<b>C4</b>

**Tabel 2.** Validation error matrix

Matriks ini memberikan informasi tentang jumlah sampel yang diklasifikasikan dengan benar dan salah untuk setiap kelas. Sebagai contoh, pada kelas 0, terdapat 2 sampel yang diklasifikasikan dengan benar, dan tidak ada kesalahan klasifikasi. Pada kelas 3, terdapat 1635 sampel yang diklasifikasikan dengan benar, dan 22 sampel yang salah diklasifikasikan ke kelas lain.

<b>Class 0</b>	149	0	0	6	1
<b>Class 1</b>	4	26	0	1	0
<b>Class 2</b>	0	0	0	0	0
<b>Class 3</b>	1	0	0	509	0
<b>Class 4</b>	9	0	0	0	31
<b>0,979</b>	<b>C0</b>	<b>C1</b>	<b>C2</b>	<b>C3</b>	<b>C4</b>

**Tabel 3.** KNN Error Matrix

Matriks ini memberikan informasi tentang jumlah kesalahan klasifikasi antar kelas. Sebagai contoh, pada kelas 0, terdapat 209 sampel yang salah diklasifikasikan ke kelas lain, dan 7 sampel dari kelas lain yang salah diklasifikasikan ke kelas 0. Pada kelas 3, terdapat 679 sampel yang salah diklasifikasikan ke kelas lain, dan 1 sampel dari kelas lain yang salah diklasifikasikan ke kelas 3.

Validation Accuracy	KKN Accuracy
<b>0,960989010989011</b>	<b>0.9700272479564033</b>

**Tabel 4.** Tabel Perbandingan akurasi

Algoritma K-NN menunjukkan akurasi yang lebih tinggi (0,970) dibandingkan set validasi (0,961), yang mengindikasikan bahwa model KNN memiliki performa lebih baik secara keseluruhan.

#### IV. Kesimpulan

Penerapan metode K-NN untuk klasifikasi tutupan lahan di kabupaten Empat Lawang berhasil membedakan lima kelas tutupan lahan dengan visualisasi yang cukup jelas. Peta hasil klasifikasi dapat digunakan untuk berbagai keperluan, seperti perencanaan tata ruang, monitoring lingkungan dan studi perubahan lahan. Penerapan metode K-Nearest Neighbor (K-NN) terbukti efektif untuk pemetaan dan pemantauan tutupan lahan di masa yang akan mendatang.

Algoritma KNN menunjukkan performa klasifikasi yang kuat dengan akurasi tinggi (0.9700272479564033) dan metrik yang kokoh di sebagian besar kelas, terutama untuk kelas yang lebih besar seperti kelas 3. Secara keseluruhan, model KNN yang diterapkan memberikan hasil yang sangat memuaskan dengan akurasi tinggi, menunjukkan potensi kuat untuk aplikasi serupa di wilayah lain atau dengan dataset berbeda.

#### V. REFERENSI

[1] R. P. E. Dhova Afasel, "Klasifikasi Tutupan Lahan Menggunakan Supervised Machine Learning Pada Citra Satelit Menggunakan Google Earth Engine," *e-Proceeding of*

*Engineering*, vol. 8, no. 6, pp. 3278-3287, 2022.

[2] J. G. B. S. Septiyan Firmansyah, "Perbandingan Klasifikasi SVM dan Decision Tree Untuk Pemetaan Mangrove Berbasis Objek Menggunakan Citra Satelit Sentinel-2B di Gili Sulat, Lombok Timur," *Journal of Natural Resources and Environmental Management*, vol. 9, no. 3, pp. 746-757, 2019.

[3] A. Fadhila Tangguh Admojo, "Klasifikasi Aroma Alkohol Menggunakan Metode KNN," *Indonesian Journal of Data and Science*, vol. 1, no. 2, pp. 34-38, 2020.

[4] R. Siringoringo, "KLASIFIKASI DATA TIDAK SEIMBANG MENGGUNAKAN ALGORITMA SMOTE DAN k-NEAREST NEIGHBOR," *Jurnal ISD*, vol. 3, no. 1, pp. 44-49, 2018.

[5] H. Z. Z. A. P. S. ,. M. Z. Febriana Santi Wahyuni, "Penerapan Metode K-Nearest Neighbor Untuk Klasifikasi Tutupan Lahan," *SENIATI*, pp. 91-96, 2023.