

## Penerapan Metode K-Nearest Neighbor Untuk Klasifikasi Tutupan Lahan

Febriana Santi Wahyuni<sup>1)</sup>, Hani Zulfia Zah'ro<sup>2)</sup>, Agung Panji Sasmito<sup>3)</sup>, Mukhammad Zainul Musyafa<sup>4)</sup>

<sup>1),2),3 )4)</sup> Teknik Informatika, Institut Teknologi Nasional Malang  
Jl. Sigura-gura 2 Malang  
Email : vebryana@lecturer.itn.ac.id

**Abstrak.** Tutupan lahan adalah jenis hamparan objek yang menutupi permukaan bumi, misalnya tumbuhan tanaman keras. Perkembangan data penginderaan jauh dari foto udara hingga fotografi digital tidak lepas dari sejarah perkembangan komputer. Pengolahan data penginderaan jauh yang memerlukan interpretasi visual dengan teknik manual beralih ke teknik analisis digital, oleh karena itu penggunaannya semakin berkembang dengan beragam teknik untuk meningkatkan akurasi hasil. Keuntungan analisis citra komputer adalah mengurangi unsur subjektif dalam penilaian manusia. Salah satu kendala dalam mengidentifikasi pola spektral data penginderaan jauh adalah resolusi gambarnya, dimana semakin tinggi resolusi maka volume data semakin besar dan algoritma analisis semakin kompleks. Oleh karena itu, sejumlah algoritma mulai dikembangkan untuk mampu menyimpan data dalam jumlah besar, menganalisis dan memvalidasinya. Pendekatan berbasis pengetahuan kemudian mulai beralih ke pendekatan berbasis data, yang merupakan titik temu antara statistik dan ilmu komputer, khususnya machine learning. Algoritma K-Nearest Neighbor (K-NN) adalah sebuah metode klasifikasi terhadap sekumpulan data maupun dokumen berdasarkan pembelajaran data yang sudah terklasifikasi sebelumnya. Algoritma ini termasuk dalam supervised learning, dimana hasil query instance yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam K-NN. Tujuan dari penelitian ini adalah mengklasifikasi tutupan lahan dengan menggunakan metode K Nearest Neighbor. Setelah melalui proses pengujian diperoleh tingkat akurasi dari system sebesar 67% dan error rate sebesar 33%.

**Katakunci :** K-Nearest Neighbor, Tutupan Lahan, Sistem Cerdas, Klasifikasi.

### 1. Pendahuluan

Tutupan lahan adalah jenis hamparan objek yang menutupi permukaan bumi, misalnya tumbuhan tanaman keras. Perkembangan data penginderaan jauh dari foto udara hingga fotografi digital tidak lepas dari sejarah perkembangan komputer. Pengolahan data penginderaan jauh yang memerlukan interpretasi visual dengan teknik manual beralih ke teknik analisis digital, oleh karena itu penggunaannya semakin berkembang dengan beragam teknik untuk meningkatkan akurasi hasil. Keuntungan analisis citra komputer adalah mengurangi unsur subjektif dalam penilaian manusia. Salah satu kendala dalam mengidentifikasi pola spektral data penginderaan jauh adalah resolusi gambarnya, dimana semakin tinggi resolusi maka volume data semakin besar dan algoritma analisis semakin kompleks. Oleh karena itu, sejumlah algoritma mulai dikembangkan untuk mampu menyimpan data dalam jumlah besar, menganalisis dan memvalidasinya. Pendekatan berbasis pengetahuan kemudian mulai beralih ke pendekatan berbasis data, yang merupakan titik temu antara statistik dan ilmu komputer, khususnya machine learning.

Meningkatnya penggunaan machine learning dalam penginderaan jauh karena karakteristik data gambar piksel yang tinggi, berbagai noise dan ketidakpastian dalam data serta nonlinier dan redundansi spasial [1]. Pada tahun 2016, Arif dan Wahyuni melakukan penelitian dengan judul Penggunaan Metode Machine Learning Untuk Pengenalan Pola Tutupan Lahan Pada Citra Satelit. Pada penelitian ini dilakukan klasifikasi untuk mengenali pola tutupan lahan dengan menggunakan metode Support Vector Machine dan Artificial Neural Network. Adapun hasil dari penelitian tersebut adalah klasifikasi tutupan lahan citra dengan menggunakan metode Support Vector Machine lebih dapat diunggulkan dengan akurasi (88,02%) dibandingkan dengan metode Artificial Neural Network

(81,78%), tetapi hanya dapat menangani dataset dalam jumlah yang kecil [2]. Menurut Rina (2019), Algoritma *K Nearest Neighbour* mampu melakukan klasifikasi lahan sawah, dari hasil pengujian dengan nilai k yang berbeda-beda dan data training yang sama hasil akurasi tertinggi diperoleh pada k=1 sebesar 95% dan terendah pada nilai k sebesar 86,66%[3]. Penelitian yang dilakukan oleh Miranda dan Aryuni (2021) menyimpulkan bahwa klasifikasi lahan menggunakan *convolutional neural network* pada citra satelit daerah Semarang dari dokumen RSNI-1 Badan Standar Nasional Indonesia untuk klasifikasi kelas penutupan lahan. Penelitian ini menggunakan tiga nilai fitur dari hasil proses ekstraksi fitur yaitu fitur spektral, indeks spektral dan fitur spasial. Klasifikasi tutupan lahan dilakukan menggunakan metode *Convolutional Neural Network* untuk mengklasifikasikan objek ke kelas tutupan lahan hutan kering primer, hutan kering sekunder, hutan tanaman, padang rumput, pemukiman, badan air atau tanah terbuka. Metode *Convolutional Neural Network* yang digunakan mampu mengklasifikasi dengan baik yaitu overall accuracy 98.4% untuk data pelatihan dan data pengujian [4]. Berdasarkan penelitian-penelitian tersebut, penelitian ini bertujuan untuk melakukan klasifikasi pada tutupan lahan dengan menggunakan metode *K Nearest Neighbor*.

### 1.1. Metode K-Nearest Neighbour

Algoritma *K-Nearest Neighbor* (K-NN) adalah sebuah metode klasifikasi terhadap sekumpulan data maupun dokumen berdasarkan pembelajaran data yang sudah terklasifikasikan sebelumnya. Algoritma ini termasuk dalam *supervised learning*, dimana hasil *query instance* yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam K-NN

Algoritma dari metode *K-Nearest Neighbor* adalah sebagai berikut :

- Menentukan parameter k (jumlah tetangga paling dekat).
- Menghitung jarak objek terhadap data training yang diberikan.
- Mengurutkan hasil dari langkah poin 2 secara ascending
- Mengumpulkan data sejumlah K yang telah ditentukan sebelumnya.
- Pilih kategori mayoritas atau kategori yang paling sering muncul sebagai hasil dari klasifikasi.

Rumus jarak yang digunakan pada penelitian ini adalah jarak Euclidian. Adapun persamaan dari perhitungan jarak Euclidian ditunjukkan pada Persamaan 1

$$distance = \sqrt{\sum_{i=1}^n (X_{training}^i - X_{testing})^2} \dots\dots\dots (1)$$

Dimana :

$X_{training}^i$  = Data training ke-i

$X_{testing}$  = Data testing

i = Record (baris) ke-i dari tabel

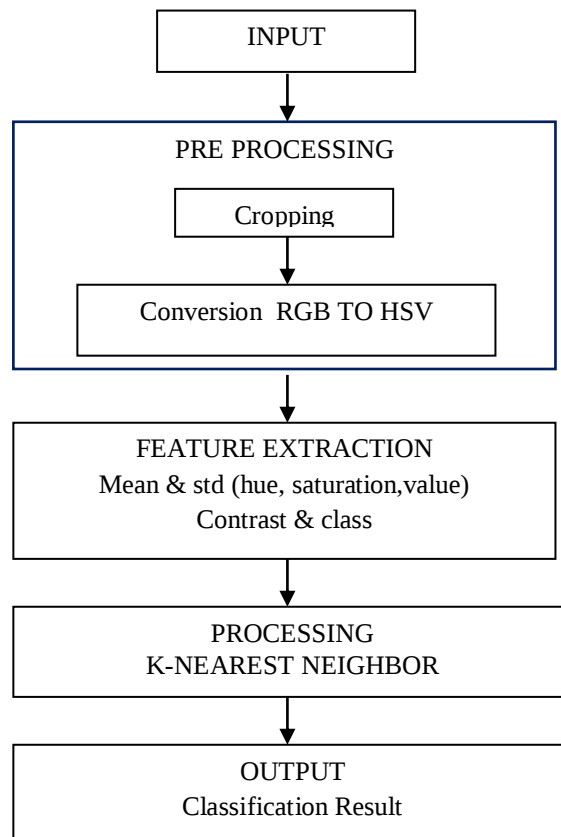
n = Jumlah data training

### 1.2. Metode Evaluasi Klasifikasi

Pada penelitian ini dilakukan evaluasi dengan menggunakan metode *confusion matrix*. *Confusion matrix* adalah alat yang berguna untuk menganalisis seberapa baik *classifier* mengenali *tuple* dari kelas yang berbeda [5]. Nilai dari *True-Positive* dan *True-Negative* memberikan informasi ketika *classifier* melakukan klasifikasi data bernilai benar, sedangkan *false-positive* dan *false-negative* memberikan informasi ketika *classifier* salah dalam melakukan klasifikasi data [6].

## 2. Pembahasan

Adapun langkah-langkah dalam penelitian ini digambarkan pada diagram blok yang ditunjukkan pada Gambar 1.



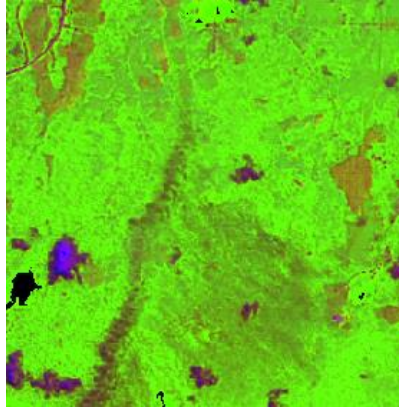
Gambar 1. Diagram Blok Sistem

### 1. Input

Dataset yang digunakan sebagai input dalam penelitian ini adalah sekumpulan data citra satelit kawasan tutupan lahan yang diperoleh dari website <https://stanfordmlgroup.github.io/projects/forestnet/> [7]. Dataset berjumlah 187 buah citra dengan format .png dengan resolusi 512x512 dari dua kelompok citra vegetasi yakni perkebunan dan pertanian.



Gambar 2. Citra satelit format RGB



Gambar 3. Citra hasil konversi RGB to HSV

## 2. Pre Processing

Citra-citra tersebut kemudian dilakukan proses *cropping* menjadi 320 x 320, hal ini dilakukan karena ukuran dataset awal terlalu besar. Selanjutnya dilakukan konversi citra dari RGB (*red green blue*) ke HSV (*hue saturation value*) dengan tujuan untuk mempermudah proses perolehan ekstraksi fitur. Adapun contoh citra dari dataset yang berformat RGB ditunjukkan pada Gambar 2., sedangkan citra hasil konversi yang berformat HSV ditunjukkan pada Gambar 3.

## 3. Feature Extraction

*Feature Extraction* adalah suatu tahapan mengekstrak ciri atau informasi yang dimiliki oleh suatu obyek citra. Ciri tersebut dapat digunakan untuk membedakan antara objek satu dengan lainnya. Pada penelitian ini fitur yang diekstrak meliputi fitur rata-rata nilai hue, rata-rata nilai saturation, rata-rata nilai value, standart deviation hue, standart deviation saturation, standart deviation value, dan contrast. Untuk selanjutnya nilai-nilai tersebut akan diproses dengan menggunakan metode K Nearest Neighbor.

## 4. Pemrosesan dengan metode K Nearest Neighbor

### a. Menentukan variable independent

Variable independent meliputi rata-rata nilai hue, rata-rata nilai saturation, rata-rata nilai value, *standart deviation hue*, *standart deviation saturation*, *standart deviation value*, dan *contrast*.

|   | hue_mean  | saturation_mean | value_mean | hue_std   | saturation_std | value_std | contrast  | class |
|---|-----------|-----------------|------------|-----------|----------------|-----------|-----------|-------|
| 0 | 70.281143 | 96.384453       | 38.834756  | 12.905452 | 19.198329      | 8.389318  | 9.825940  | 1     |
| 1 | 66.521250 | 148.187178      | 28.515625  | 7.907390  | 32.414633      | 18.078571 | 24.105398 | 1     |
| 2 | 73.642021 | 247.173408      | 13.385518  | 11.090223 | 25.707028      | 7.708227  | 9.155163  | 1     |
| 3 | 66.070557 | 146.063057      | 30.619961  | 7.371980  | 32.885180      | 16.649294 | 26.166614 | 1     |
| 4 | 63.599170 | 231.964619      | 17.660381  | 16.479968 | 58.923133      | 12.510671 | 25.403566 | 1     |

Gambar 4. Variabel independent

### b. Menentukan variable dependent

Variable dependent nya adalah class atau kategori dari klasifikasi, yaitu nilai 1= tutupan lahan dan 2=non tutupan lahan

```
0    1
1    1
2    1
3    1
4    1
Name: class, dtype: int64
```

Gambar 5. Variabel dependent



### 3. Kesimpulan

Berdasarkan hasil pengujian dan evaluasi *classifier* dapat disimpulkan sebagai berikut :

1. Algoritma K *Nearest Neighbor* dapat digunakan sebagai metode untuk melakukan klasifikasi tutupan lahan.
2. Tingkat akurasi system dalam melakukan klasifikasi adalah sebesar 67%, dapat dinyatakan bahwa *classifier* ini cukup akurat untuk melakukan klasifikasi tutupan lahan.

### Daftar Pustaka

- [1]. Camps-Valls,G, "Machine Learning in Remote Sensing Data Processing. Machine Learning for Signal Processing", MLSP 2009
- [2]. Arif, Nursida, Febriana S.W, "Penggunaan Metode *Machine Learning* Untuk Pengenalan Pola Tutupan Lahan Pada Citra Satelit", in Proc. Semnasteknomedia 2016, pp1.2-1, 6-7 Februari 2016.
- [3]. Rina, Khadiza., "Klasifikasi Lahan Sawah Menggunakan Algoritma K-Nearest Network"2019 (Online : <https://www.scribd.com/document/549538945/131402065>)
- [4]. E. Miranda and M. Aryuni, "Klasifikasi Tutupan Lahan Menggunakan Convolutional," Jurnal Sistem Informasi, 2021.
- [5]. (<https://www.scribd.com/document/549538945/131402065>)
- [6]. Han,Jiawei., Micheline Kamber and Jian Pei (2011)., Data Mining Concept And Techniques 3rded. ISBN 978-0123814791 Morgan Kauffman.
- [7]. Fibrianda, M. F., dan Bhawiyuga, A. (2018). Analisis perbandingan akurasi deteksi serangan pada jaringan komputer dengan metode naïve bayes dan support vector machine (svm). Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 2(9)