

KLASIFIKASI CITRA EUROSAT MENGGUNAKAN ALGORITMA KNN, DECISION TREE DAN RANDOM FOREST

Yusril Iza Fajarendra, Yulis Rizal Fauzan, Shofwatul 'Uyun

Informatika, Universitas Islam Negeri Sunan Kalijaga Yogyakarta
Jl. Laksda Adisucipto, Papringan, Caturtunggal, Kec. Depok, Kabupaten Sleman,
Daerah Istimewa Yogyakarta 55281
Fajarendra1999@gmail.com

ABSTRAK

Penggunaan teknologi citra satelit telah berkembang pesat dan banyak dimanfaatkan untuk penelitian, mencakup pegunungan, sungai, danau, dan lainnya. Dalam penelitian ini, digunakan dataset *EuroSat* yang memiliki 10 label yang mendeskripsikan perubahan bentuk lahan dan pemukiman di wilayah Eropa. Perubahan tersebut disebabkan oleh kerusakan akibat aktivitas manusia dan alam seperti pertanian, irigasi, dan bencana alam. Selain itu, kompleksitas dan variabilitas dataset *EuroSat* yang beragam memerlukan pendekatan yang mampu mengenali pola-pola halus dan berbeda pada setiap kategori. Oleh karena itu, penelitian ini bertujuan untuk mengkomparasikan kinerja beberapa algoritma dalam klasifikasi citra satelit. Dalam hal ini, digunakan tiga algoritma yaitu *KNN*, *Decision Tree*, dan *Random Forest* untuk dibandingkan kinerjanya dalam klasifikasi citra data *EuroSat*. Metode penelitian yang digunakan meliputi import data, normalisasi data, konversi label ke one-hot encoding, ekstraksi fitur, dan normalisasi fitur menggunakan model *VGG16* yang ditraining terlebih dahulu untuk mendapatkan representasi fitur yang informatif. Fitur yang diekstraksi kemudian digunakan sebagai input untuk ketiga algoritma tersebut. Pengujian performa model dilakukan menggunakan *K-Fold Cross Validation*. Hasil yang diperoleh memperlihatkan bahwa *Random Forest* memiliki kinerja terbaik dengan *accuracy* 81,32%, *recall* 81,33%, *precision* 81,18%, dan *f1 score* 80,82%. Penelitian ini menunjukkan bahwa penggunaan algoritma *Random Forest* dapat meningkatkan akurasi dan performa dalam klasifikasi citra satelit.

Kata Kunci : Klasifikasi Citra, *K-Fold Cross Validation*, *EuroSat*, *KNN*, *Decision Tree*, *Random Forest*, *VGG16*

1. PENDAHULUAN

Pemanfaatan teknologi citra satelit telah menjadi bagian terpenting dalam proses monitoring keadaan lingkungan, pemetaan muka bumi serta perubahan yang terjadi pada alam ini. Data citra satelit sangat menyajikan deskripsi yang luas dan ketelitian tentang keadaan permukaan bumi dari ketinggian sehingga sangat bermanfaat diterapkan dalam berbagai pengaplikasian dari monitoring pertanian sampai pemetaan bencana alam [1]. Data ini sangat memberikan pandangan kepada para peneliti dan pembuat kebijakan dalam melakukan pemantauan secara real-time dan mendeteksi adanya perubahan kecil maupun besar di berbagai belahan dunia. Satu diantara dataset citra satelit yang digunakan adalah *EuroSat* yang merupakan kumpulan gambar citra satelit dari Sentinel-2 yang dioperasikan oleh European Space Agency (ESA).

Dataset ini terdiri dari 5000 citra satelit yang memiliki 10 kelas dengan resolusi spasial tinggi sehingga membuat data ini merepresentasikan berbagai jenis tutupan lahan yang ada di Eropa. Penggunaan informasi ini sangat penting termasuk sumber daya alam, mitigasi bencana, dan penelitian perubahan iklim. Pemanfaatan dataset citra memerlukan teknik klasifikasi yang baik dan efektif. Pada klasifikasi citra proses piksel dalam citra diidentifikasi dan diberikan label sesuai kelas tertentu [2].

Penggunaan *machine learning* sekarang ini juga memberikan dampak yang sangat besar dalam berbagai aspek termasuk dalam pengklasifikasian

jenis-jenis lahan. Manfaat utamanya adalah menghemat waktu dan energi dibandingkan dengan melakukan pengklasifikasian secara langsung dilapangan. Penelitian yang telah dilakukan sebelumnya berhasil mengklasifikasi citra tutupan lahan menggunakan data *EuroSat* dan menggunakan algoritma *Xception* dengan akurasi sebesar 96,6% dengan activation function *ReLU* dan 95,9% menggunakan activation function *swish* [3]. Riset ini menunjukan wujud perhatian pada perkembangan pada suatu daerah dipermukaan bumi.

Klasifikasi adalah sebuah tahapan pencarian model yang mendeskripsikan dan membedakan class data atau konseptualisasi supaya model tersebut dapat dipakai dalam memprediksi kelas dari suatu objek yang label kelasnya belum ditemukan [4]. Beberapa algoritma klasifikasi yang sering dimanfaatkan secara luas antara lain *Decision Tree*, *Naïve Bayes*, *KNN*, *CNN*, *Algoritma Genetika*, *SVM*, *Metode Rule Based*, *Rough Sets*, dan *Memory Based Reasoning*. Algoritma *random forest* adalah algoritma machine learning yang menggunakan kumpulan pohon keputusan. Algoritma ini bekerja dengan membentuk sejumlah besar pohon keputusan, di mana setiap pohon dibangun dari pengamatan yang diambil secara acak dengan penggantian dari data pelatihan [5]. Pada penelitian selanjutnya, klasifikasi citra penyakit daun menggunakan algoritma *K-Nearest Neighbors (KNN)* menghasilkan akurasi sebesar 80%. Algoritma *KNN* adalah metode klasifikasi yang digunakan untuk mengelompokkan objek ke dalam lebih dari satu kelas yang serupa atau berbeda [6]. *KNN* juga digunakan

untuk memberikan prediksi nilai asli pada sampel yang tidak diketahui

Dalam penelitian tentang klasifikasi citra burung Lovebird menggunakan Decision Tree J48 dengan empat jenis evaluasi, untuk mengurangi tahapan ekstraksi fitur dan mempercepat komputasi, hasilnya menunjukkan akurasi sebesar 100%, recall 100%, precision 100%, dan f1 score 100%. Hasil ini menunjukkan bahwa ekstraksi tingkat pertama di Decision Tree J48 mampu meningkatkan kinerja dan menghasilkan akurasi yang sangat tinggi [7].

Permasalahan utama dalam klasifikasi citra satelit adalah kompleksitas dan variabilitas data yang mana pada tutupan lahan yang beraneka ragam dan kompleks. Pendekatan yang dapat mengenali pola-pola yang halus dan berbeda disetiap kategori inilah yang memungkinkan karakteristik spektral yang tumpang tindih sehingga membuat tugas klasifikasi menjadi berat. Penerapan model klasifikasi yang membutuhkan sumber daya komputasi yang tinggi seringkali tidak praktis diterapkan dalam keadaan sumber daya terbatas.

Selain itu, tutupan lahan juga merupakan akar variabel dalam mengukur bermacam-macam fenomena yang terjadi di permukaan bumi contoh perubahan lingkungan, akibat bencana alam dan kerusakan oleh manusia, maupun mengevaluasi pengelolaan lahan dan tata wilayah [8]. Perkembangan bentuk tanah dan pemukiman disuatu daerah akan memperoleh perubahan yang mudah dilihat secara visual. Perubahan tersebut terjadi karena disebabkan oleh factor manusia dan alam seperti pertanian, perkebunan, irigasi, dan pembangunan yang sangat membantu dalam penilaian dan perkembangan suatu wilayah, usaha dalam konservasi lahan dan perencanaan pembangunan infrastruktur didaerah tersebut [9].

Berdasarkan penjelasan yang diuraikan diatas, penelitian ini berfokus pada penggunaan algoritma KNN, Decision Tree dan Random Forest dalam menganalisis klasifikasi citra dataset EuroSat yang terdiri dari 500 citra gambar dengan 10 label. Melalui kombinasi metode dan algoritma yang digunakan, diharapkan dapat mencapai klasifikasi yang efektif dan akurat untuk dataset citra tutupan lahan EuroSat.

2. TINJAUAN PUSTAKA

Pada bagian tinjauan pustaka ini menjelaskan tentang beberapa teori, penelitian sebelumnya dan literatur yang berkaitan dengan klasifikasi citra dan algoritma yang digunakan yaitu KNN, Decision Tree dan Random Forest. Selain itu juga membahas mengenai normalisasi, ekstraksi dan konversi label pada data. Tinjauan Pustaka ini berfungsi agar memberikan dasar teoritis mengenai topik yang diteliti sehingga mempermudah memahami latar belakang dan landasan ilmiah dari penelitian ini. Beberapa teori dan penelitian sebelumnya yang dibahas antara lain sebagai berikut.

2.1. Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep yang digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak atau belum diketahui [10]. Berikut ini algoritma klasifikasi yang sering dimanfaatkan untuk penelitian diantaranya Decision Tree, Naïve Bayes, KNN, CNN, Algoritma Genetika, SVM, Metode Rule Based, Rough Sets, dan Memory Based Reasoning [4].

2.2. Algoritma K-Nearest Neighbor (K-NN)

K-Nearest Neighbor (KNN) adalah salah satu algoritma yang digunakan untuk klasifikasi data citra. Metode ini diaplikasikan untuk mengatasi masalah identifikasi yang diukur secara kualitatif dan kuantitatif berdasarkan data pembelajaran dengan jarak terdekat. Untuk menemukan jarak data ke tetangga (Euclidean Distance), diperlukan penentuan nilai K tetangga (neighbor). Fungsi algoritma K-NN adalah untuk mengklasifikasikan objek baru berdasarkan atribut data dari sampel training berdasarkan jarak Euclidean Distance. Secara umum rumus persamaan algoritma K-NN dengan jarak Euclidean Distance adalah sebagai berikut [11].

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2} \quad (1)$$

Keterangan :

$d(x_i, x_j)$: Jarak tetangga (Euclidean Distance)

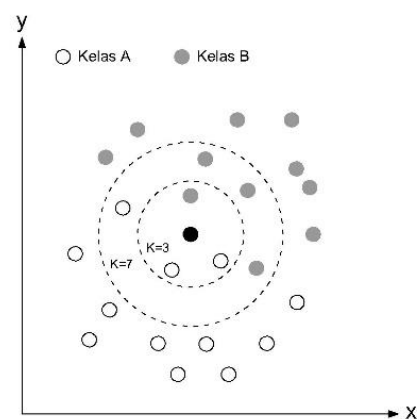
(x_i) : record ke -i

(x_j) : record ke -j

a_r : data ke-r

i, j : 1,2,3,4.....n

Ilustrasi algoritma K-NN dapat dilihat pada gambar dibawah ini.



Gambar 1. Ilustrasi algoritma KNN

2.3. Algoritma Decision Tree

Decision Tree adalah algoritma yang mempunyai bagan bentuk menyerupai pohon pada setiap node yang merupakan pengujian terhadap variabel atribut. Penggunaan algoritma ini adalah untuk menganalisis proses klasifikasi dan prediksi [12]. Berbagai penelitian telah dilakukan dengan menggunakan

algoritma Decision Tree yaitu mengatasi missing value dan pruning [13]. Persamaan algoritma Decision Tree adalah sebagai berikut.

$$Gain(S, A) = Entropi(s) - \sum_{i=0}^n \frac{|s_i|}{|s|} * Entropi s_i \quad (2)$$

Keterangan :

S : himpunan kasus

A : Atribut

n : jumlah partisi atribut A

$|s_i|$: jumlah kasus partisi ke- i

$|s|$: jumlah kasus dalam S

Sedangkan persamaan untuk mencari nilai Entropy adalah sebagai berikut.

$$Entropy(s) = \sum_{i=1}^k -p_i * \log_2 p_i \quad (3)$$

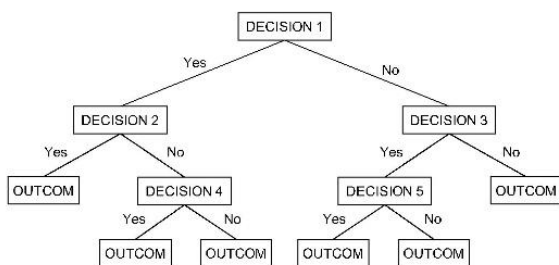
Keterangan :

S : himpunan kasus

k : jumlah partisi S

p_i : probabilitas yang diperoleh dari jumlah (ya/tidak) dibagi total kasus 3

Pemodelan bagan arsitektur algoritma Decision Tree dapat dilihat pada gambar dibawah ini.



Gambar 2. Bagan model algoritma *Decision Tree*

Semua gambar harus diberi penjelasan dalam bentuk paragraf dibawahnya, contoh: dari grafik pada gambar 1, dapat dilihat.....

- Format Gambar (in line with text) diletakkan di tengah halaman (*center alignment*) baik itu pada kolom 1 ataupun pada kolom 2 dengan judul ditengah.
- Seluruh gambar harus diberi judul dibawah gambar, dengan penomoran secara berurutan mulai angka 1 contoh Gambar 1. dan seterusnya bukan Gambar 2.1.
- Jika gambar besar boleh diubah menjadi 1 kolom diletakkan di tengah, untuk selanjutnya dikembalikan menjadi 2 kolom lagi.

2.4. Algoritma Random Forest

Random Forest adalah pemodelan dengan metode melakukan pemisahan biner rekursif agar menggapai node akhir pada bagan bentuk pohon berdasarkan pada pohon-pohon klasifikasi dan regresi [11]. Keuntungan dari penggunaan algoritma random

forest adalah dapat menghasilkan error yang relative kecil, proses dalam klasifikasi performanya baik, mampu menyelesaikan data training dalam jumlah yang banyak secara efektif, dan algoritma ini sangat baik dalam mengatasi missing data. Random Forest mampu membentuk banyak pohon independent dengan subset yang dipilih secara acak melalui bootstrap dari sampel pelatihan dan inputan pada setiap node [14]. Persamaan algoritma random forest dalam dilihat pada gambar dibawah ini.

Persamaan 1 Gini Impurity adalah pengukuran ketidakmurnian suatu simpul dalam pohon keputusan . Rumus :

$$Gini(t) = 1 - \sum_{i=1}^C p_i^2 \quad (4)$$

Keterangan :

t : Simpul pada pohon

C : Jumlah kelas

P_i : Proporsi dikelas i pada simpul t

Setiap pohon dalam Random Forest dilatih menggunakan sampel bootstrap dari dataset asli. Jika kita memiliki dataset D dengan N sampel, maka akan dipilih N sampel dengan penggantian (dengan probabilitas yang sama untuk setiap sampel) untuk membentuk dataset bootstrap D_i :

$$D_i = \{(x_{i1}, y_{i1}), (x_{i2}, y_{i2}), \dots, (x_{iN}, y_{iN})\} \quad (5)$$

Keterangan :

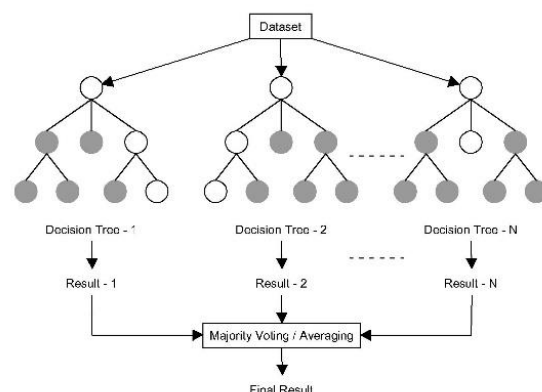
D_i : Set bootstrap ke- i , subset data yang diperoleh dengan penggantian dari dataset asli

x_{ij} : Fitur atau atribut dari sampel ke- j dalam set bootstrap ke - i

y_{ij} : Label atau nilai target dari sampel ke- j dalam set bootstrap ke - i

N : Total sampel dalam set bootstrap D_i

Ilustrasi pemodelan arsitektur algoritma random forest adalah sebagai berikut :



Gambar 3. Struktur algoritma *Random Forest*

2.5. K-Fold Cross Validation

K-Fold Cross Validation merupakan model evaluasi performa algoritma dengan Teknik validasi silang yaitu membagi data membentuk K bagian set

data yang memiliki ukuran sama. Pemanfaatan metode *K-Fold Cross Validation* berfungsi sebagai peniadaan bias sehingga evaluasi dilakukan secara optimal [15]. Data training dan testing dilakukan sejumlah K kali sehingga semua data akan menjadi data training dan data testing [16]. Tahapan evaluasi *K-Fold Cross Validation* dapat dilihat pada gambar dibawah ini:

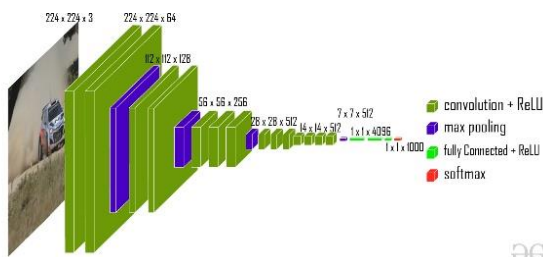
Tabel 1. Metode *K-Fold Cross Validation*

1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5

Data Pelatihan
Data Pengujian

2.6. VGG16-Net

VGG16-net adalah arsitektur deep learning yang dominan dipakai pada klasifikasi citra. Arsitektur VGG16 terdiri atas 16 layer yaitu kombinasi convolutional layer, max pooling layer dan fully connected layer. Arsitektur VGG16 ini memang sudah terbukti sangat baik dalam melakukan klasifikasi citra pada berbagai dataset citra misalnya dataset ImageNET. Berikut adalah ilustrasi VGG16 MODEL ARSITEKTUR VGG16 U [17].



Gambar 4. Model VGG16

2.7. Augmentasi Data

Augmentasi data merupakan cara yang banyak digunakan untuk mengurangi over-fitting dan noise. Pada augmentasi data diperoleh data baru dengan menggunakan transformasi data asli sehingga memungkinkan dalam meningkatkan generalisasi data [18]. Ada beberapa tahapan augmentasi data yaitu rotasi, pergeseran lebar dan tinggi, geseran, zoom dan flip horizontal. Contoh gambar yang telah melalui tahap augmentasi.



Gambar 5. Augmentasi Data

2.8. Data Citra

Data yang digunakan pada penelitian ini yaitu citra satelit EuroSat yang diperoleh dari link <https://www.kaggle.com/datasets/apollo2506/eurosat-dataset/data>. Dataset EuroSat ini adalah tentang gambar citra satelit yang digunakan untuk tugas pengenalan citra pada klasifikasi citra yang berada di wilayah benua Eropa terutama dalam penggunaan lahan pada citra geospasial. Dataset ini terdiri atas gambar yang didapat dari Sentinel-2. EuroSat memiliki 10 label diantaranya adalah tanaman tahunan, hutan, vegetasi herbal, jalan raya, industry, padang rumput, pangkas permanen, permukiman, sungai, dan danau laut dengan total gambar citra sebanyak 5000 data. Setiap gambar berukuran 64 x 64 piksel dengan Groud Sampling Distance 10 m yang ditunjukkan pada gambar dibawah ini



Gambar 6. Dataset EuroSat

2.9. Normalisasi Data

Proses yang lebih umum yang mencakup normalisasi fitur dengan contoh melakukan penanganan pada missing value atau transformasi log, penghapusan data yang duplikasi dan lainnya [19].

2.10. Konversi Label

Konversi Label merupakan proses untuk mengubah bentuk label kategori yaitu nama label atau nama kelas menjadi format numerik yang digunakan oleh algoritma machine learning [20].

2.11. Ekstraksi dan Normalisasi Fitur

Ekstraksi fitur merupakan tahapan dalam memperoleh informasi yang relevan dari data untuk dipakai dalam machine learning. Tujuannya adalah untuk mengubah data menjadi format yang lebih representatif serta informatif sehingga bisa meningkatkan kinerja model algoritma salah satunya adalah pengolahan citra. Ekstraksi fitur yang digunakan dalam penelitian ini adalah model pralatih VGG16 untuk ekstraksi fitur gambar [21]. Normalisasi fitur merupakan proses skala nilai-nilai fitur ke rentang tertentu antara 0 dan 1 serta membantu dalam membentuk fitur yang memiliki skala sebanding sehingga model machine learning tidak memberikan bobot yang berlebihan pada fitur tertentu yang mempunyai rentang nilai lebih besar. [21].

3. METODE PENELITIAN

Tahapan metode penelitian pada klasifikasi citra dataset EuroSat ini adalah import data, normalisasi data, konversi label, augmentasi data, ekstraksi fitur, normalisasi fitur, klasifikasi dan evaluasi. Berikut ini adalah flowchart metode penelitian yang digunakan pada riset ini.



Gambar 7. Metode penelitian

3.1. Normalisasi Data

Pada penelitian ini, tahapan normalisasi dataset EuroSat akan diubah dari rentang $[0, 255]$ menjadi $[0,1]$ sehingga piksel akan dibagi dengan 255.

3.2. Konversi Label

Pada penelitian ini teknik konversi label datanya adalah One-Hot encoding yaitu label kategori diubah menjadi format biner sehingga dalam setiap kategori yang diubah menjadi vector biner hanya ada 1 elemen yang bernilai 1 dan semua elemen lainnya bernilai 0.

3.3. Augmentasi Data

Tahapan augmentasi data yang dilakukan ada 6 teknik yaitu rotasi (*rotation*), pergeseran tinggi dan lebar (*width shift and height shift*), *shear*, *zoom*, dan *horizontal flip*.

3.4. Ekstraksi Fitur

Ekstraksi fitur yang digunakan dalam penelitian ini adalah model prelatih VGG16 untuk ekstraksi fitur gambar. Ketika menggunakan model VGG16 ini biasanya menghapus lapisan fully connected terakhir.

3.5. Normalisasi Fitur

Tahapan Normalisasi fitur ini menggunakan model VGG16 menggunakan teknik Min-Max Scaling yang berfungsi untuk memastikan bahwa

nilai-nilai fitur sebanding dan tidak ada mendominasi dalam skala yang berbeda.

3.6. Klasifikasi

Algoritma yang digunakan dalam klasifikasi ini adalah KNN, Decision Tree dan Random Forest dengan tujuan untuk melihat model terbaik dalam melakukan klasifikasi terhadap dataset EuroSat.

3.7. Evaluasi

Untuk melatih model ketiga algoritma yang digunakan dalam klasifikasi ini adalah dengan menggunakan K-Fold Cross Validation dengan tujuan agar semua data bisa menjadi data training dan testing serta untuk meniadakan bias pada sehingga evaluasi dilakukan secara optimal.

4. HASIL DAN PEMBAHASAN

Pada bagian ini, akan menjelaskan hasil penelitian yang dilakukan secara bertahap yang sudah tercantum dalam metode penelitian. Hasil-hasil yang diperoleh dari penelitian ini akan berfungsi untuk memahami implikasi dan relevansinya terhadap tujuan penelitian. Berikut ini penjelasan hasil dan pembahasan penelitian klasifikasi citra.

4.1. Normalisasi Data

Setelah dilakukan import data, maka tahapan selanjutnya adalah normalisasi data dengan tujuan mengurangi perbedaan skala antar fitur dan mencegah overfitting dalam mempercepat pembelajaran model. Pada konteks penelitian ini normalisasi data membagi nilai piksel oleh 255 menjadi 0-1 sehingga nilai-nilai piksel tersebut berada dalam rentang antara 0 dan 1.

```

array([[ 2.19382914e-03,  9.36742559e-03],
       [ 5.14578401e-03, -2.81299920e-03],
       [-6.11705296e-03, -7.54852460e-03],
       [-8.94661217e-03, -5.36689823e-04],
       [ 6.12040932e-03, -1.67826762e-03],
       [-3.16880249e-03,  4.69270886e-03],
       [-1.06126993e-03, -8.84525347e-03],
       [ 1.88432158e-03,  2.15350720e-03],
       [-4.73557484e-03, -4.97640662e-03],
       [ 9.96695937e-04, -7.04047067e-03],

```

Gambar 8. Normalisasi data

4.2. Konversi Label

Hasil dari konversi label pada penelitian ini adalah dalam bentuk one-hot encoding yang merupakan proses mengubah label kategori menjadi representasi biner sesuai yang digunakan dalam pembelajaran mesin. Data yang digunakan ada 10 kelas sehingga hanya satu elemen yang diaktifkan bernilai 1 yang memiliki kategori yang sesuai, sedangkan elemen lainnya adalah 0 sebanyak 10 biner. Berikut beberapa hasil dari konversi label one-hot encoding

```

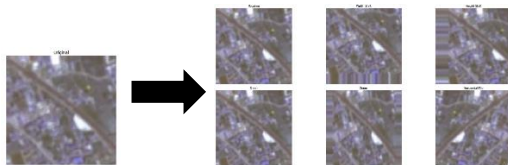
[[0] [0] [1] [0]
 [1] [1] [0] [0]
 [0] [0] [0] [1]
 [0] [1] [1] [0]
 [1] [1] [1] [1]
 [0] [1] [0] [1]
 [1] [1] [1] [1]
 [1] [1] [1] [1]
 [0] [0] [1] [0]
 [1] [0] [0] [0]
 [1] [0] [0] [0]]

```

Gambar 9. Konversi label

4.3. Augmentasi Data

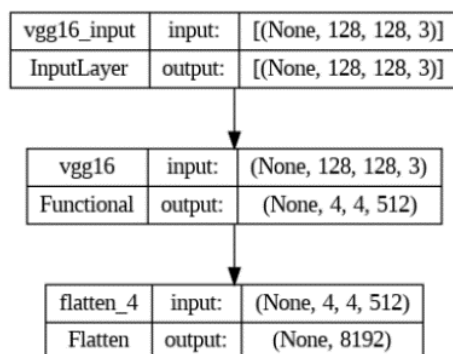
Hasil dari augmentasi dataset citra EuroSat yaitu menerapkan beberapa teknik augmentasi rotasi (*rotation*), pergeseran tinggi dan lebar (*width shift and height shift*), *shear*, *zoom*, dan *horizontal flip* dengan tujuan untuk menunjukkan variasi dari citra sampel yang kemudian akan dievaluasi serta meningkatkan variasi data dan performa model algoritma pada klasifikasi citra KNN, *Decision Tree*, dan *Random Forest*. Hasil nya dapat dilihat pada gambar dibawah ini.



Gambar 10. Hasil augmentasi data

4.4. Model VGG16

Model VGG16 pada dataset citra EuroSat memiliki arsitektur lapisan konvolusi yang cukup dalam (hingga 13 lapisan) akan tetapi pada penelitian ini hanya menampilkan 3 lapisan saja yang di ikuti oleh lapisan pooling maksimum dalam mengekstraksi fitur gambar citra Eurosat sebanyak 5000 data gambar. Pada setiap beberapa lapisan konvolusi, dilakukan max pooling untuk mengurangi dimensi spasial dari fitur yang diperoleh dalam mencapai invariansi translasi. Hasil dari pemodelan VGG16 dapat dilihat pada gambar dibawah ini.



Gambar 11. Arsitektur hasil model VGG16

4.5. Ekstraksi dan Normalisasi Fitur

Pada penelitian ini hasil ekstraksi fitur dari model VGG16 mengacu pada proses penggunaan bagian lapisan konvolusi arsitektur VGG16. Model VGG16 telah dilatih pada dataset citra EuroSat yang berarti bobot lapisan konvolusi telah disesuaikan sehingga model bisa mendapatkan pola visual umum yang

mewakili berbagai kategori objek. Pada saat dilakukannya ekstraksi fitur, biasanya membuang bagian lapisan Fully Connected (FC) yang terdapat pada bagian akhir model sehingga hanya menggunakan lapisan dan pooling sebelumnya untuk memperoleh fitur-fitur dari gambar. Setelah itu dilakukan normalisasi fitur yang menggunakan teknik normalisasi yaitu skala min-max nilai fitur yang diubah menjadi rentang 0-1 untuk memastikan bahwa fitur yang diekstraksi pada gambar citra memiliki skala atau rentang nilai yang konsisten. Hasilnya dapat dilihat pada gambar dibawah ini.

```

Vektor fitur sebelum normalisasi:
[[[0.20679563 0. 0. ... 0. 0.44415343 0. ]
 [0.16776259 0. 0. ... 0. 0.51168716 0. ]
 [0.12585187 0. 0. ... 0. 0.8339688 0. ]
 [0.16290474 0. 0. ... 0.18559931 0.3265776 0. ]
 [0.06761132 0. 0. ... 0. 0.40646932 0. ]]]

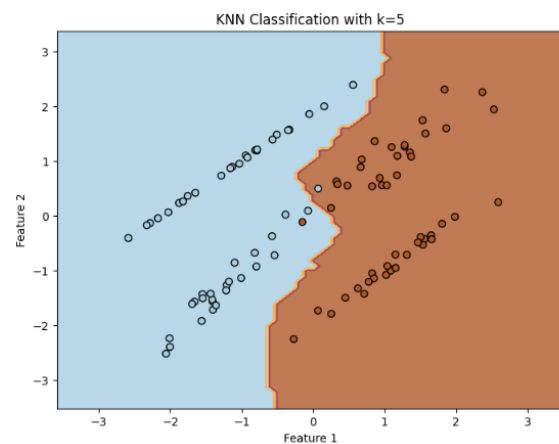
Vektor fitur setelah normalisasi:
[[[0.20679563 0. 0. ... 0. 0.34664103 0. ]
 [0.12167508 0. 0. ... 0. 0.3788491 0. ]
 [0.02501524 0. 0. ... 0. 0.46774212 0. ]
 [0.0658194 0. 0. ... 0. 0.45790952 0. ]
 [0.12188892 0. 0. ... 0. 0.23429267 0. ]]]

```

Gambar 12. Normalisasi fitur

4.6. Klasifikasi KNN

Wilayah yang berwarna memberikan arti prediksi model KNN dengan nilai K=5 pada area tersebut dan setiap warna mewakili kelas yang diprediksi oleh model KNN berdasarkan titik terdekat. Titik data latih ditampilkan diatas wilayah keputusan sesuai warna dengan kelas sebenarnya. Garis batas warna yang berbeda memperlihatkan perbatasan dimana model KNN beralih dari satu kelas ke kelas dua. Daerah inilah yang dihasilkan berdasarkan tetangga terdekat dari setiap titik digrid. Hasil visualisasi algoritma KNN dapat dilihat dibawah ini.

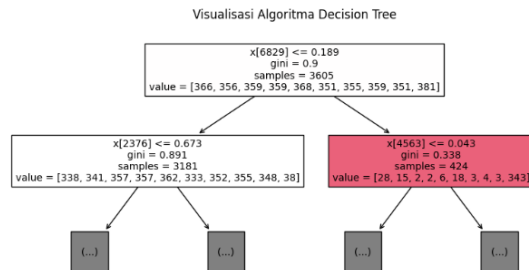


Gambar 13. Hasil visualisasi algoritma KNN

4.7. Klasifikasi Decision Tree

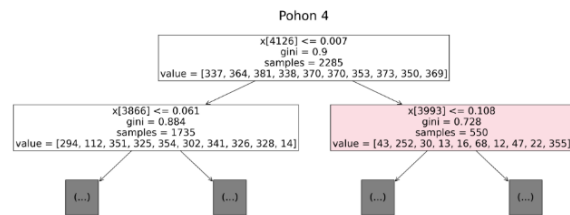
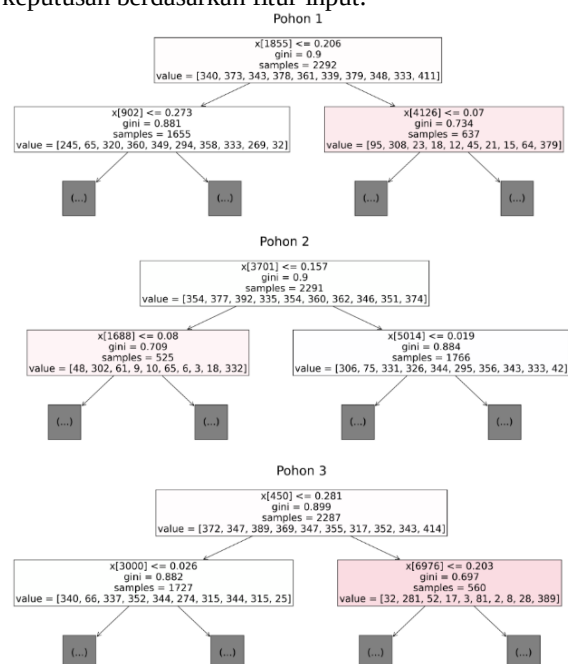
Hasil Visualisasi Decision Tree menunjukkan setiap node pada pohon mewakili keputusan berdasarkan pada nilai dan fitur tertentu dan node Root (atas) adalah titik dasar dimana keputusan pertama dibentuk. Garis yang menunjukkan bagaimana keputusan dapat dijelaskan berdasarkan alur dari atas ke bawah dan dari node ke node selanjutnya. Node yang berisi warna menandakan kelas mayoritas atau

nilai prediksi. Pada node berisi informasi tentang fitur yang dipakai untuk keputusan, jumlah sampel, nilai split dan distribusi kelas. Visualisasi dibawah ini dapat menunjukkan keputusan pertama yang dibentuk oleh model dan bagaimana proses data yang dipecah menjadi bercabang-cabang berdasarkan nilai fitur. Hasil visualisasi algoritma decision tree dapat dilihat pada gambar dibawah ini.

Gambar 14. Hasil visualisasi *Decision Tree*

4.8. Klasifikasi Random Forest

Hasil klasifikasi algoritma Random Forest merupakan hasil dari pengembangan algoritma decision tree sehingga pohon keputusan yang terbentuk banyak, bercabang-cabang dan random seperti hutan. Pada penelitian ini menampilkan 4 pohon keputusan dari model random forest. Untuk setiap node dalam pohon mewakili keputusan berdasarkan nilai dari fitur tertentu. Hampir mirip dengan algoritma decision tree dan untuk batas kedalaman hanya 2 level saja dari setiap pohon yang ditampilkan agar lebih sederhana sehingga menunjukkan variasi dalam pohon yang membentuk ensemble, yang merupakan inti dari kekuatan Random Forest dalam mengatasi overfitting dan setiap pohon individual dalam Random Forest membentuk keputusan berdasarkan fitur input.

Gambar 15. Hasil visualisasi *Random Forest*

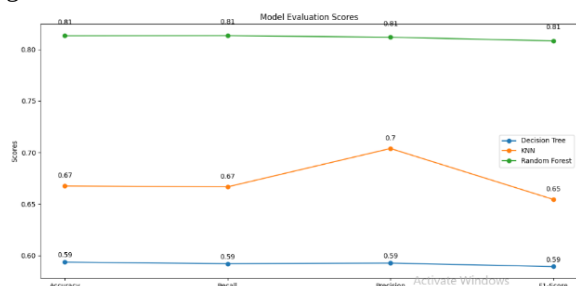
4.9. Evaluasi Model K-Fold Cross Validation

Pengujian model klasifikasi dengan menggunakan metode K-Fold Cross Validation yaitu nilai K=10 diperoleh hasil accuracy 66,74%, recall 66,66%, precision 70,38%, F1 score 65% untuk algoritma KNN. Algoritma decision tree memperoleh hasil accuracy 59,35%, recall 59,20%, precision 59,25%, dan f1 score 58,92%. Sedangkan algoritma Random Forest accuracy 81,32%, recall 81,33%, precision 81,18%, dan f1 score 80,82%. Berikut ini tabel hasil evaluasi K-Fold Cross Validation.

Tabel 2. Hasil performa *K-Fold Cross Validation*

Evaluasi Kinerja Model	KNN	Decision Tree	Random Forest
Accuracy	66,74%	59,35%	81,32%
Recall	66,66%	59,20%	81,33%
Precision	70,38%	59,25%	81,18%
F1 Score	65,44%	58,92%	80,82%

Perbandingan performa model dari algoritma KNN, Decision Tree, dan Random Forest dapat dilihat pada grafik dibawah ini.

Gambar 16. Grafik *K-Fold Cross Validation*

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan pada klasifikasi citra Klasifikasi Citra Data *EuroSat* sebanyak 5000 gambar citra satelit menggunakan perbandingan tiga algoritma KNN, decision tree, dan random forest evaluasi performa K-fold cross validation nilai k=10 dengan 10 label kelas, dapat disimpulkan bahwa algoritma *Random Forest* mendapatkan performa terbaik dibandingkan algoritma KNN dan Decision Tree dengan nilai accuracy 81,32%, recall 81,33%, precision 81,18%, dan f1 score 80,82%.

Penelitian selanjutnya disarankan untuk menggunakan algoritma klasifikasi lainnya sehingga dapat membandingkan model yang cocok diterapkan pada dataset *EuroSat*.

DAFTAR PUSTAKA

- [1] N. Inarossy and S. Y. J. Prasetyo, "Klasifikasi Wilayah Risiko Bencana Kekeringan Berbasis Citra Satelit Landsat 8 Oli Dengan Kombinasi Metode Moran's I dan Getis Ord G*(Studi Kasus: Kabupaten Boyolali dan Klaten)," *Indones. J. Comput. Model.*, vol. 2, no. 2, pp. 37–54, 2019.
- [2] I. I. A. Widyan, "Implementasi Convolutional Neural Network Menggunakan Arsitektur Visual Geometry Group-19 (Vgg-19) untuk Klasifikasi Citra Buah Salak (Studi Kasus: Klasifikasi Citra Buah Salak Asal DI Yogyakarta)." Universitas Islam Indonesia, 2024.
- [3] M. Ichwan and I. S. Hadi, "Kinerja Model EfficientNetV2M dalam Klasifikasi Citra Tutupan dan Penggunaan Lahan," *MIND (Multimedia Artif. Intell. Netw. Database) J.*, vol. 8, no. 2, pp. 203–216, 2023.
- [4] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naïve Bayes," 2018.
- [5] D. Marlina, "Klasifikasi Tutupan Lahan pada Citra Sentinel-2 Kabupaten Kuningan dengan NDVI dan Algoritme Random Forest," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.)*, vol. 7, no. 1, p. 41, 2022, doi: 10.30998/string.v7i1.12948.
- [6] E. H. Rachmawanto and H. P. Hadi, "Optimasi Ekstraksi Fitur Pada Knn Dalam Klasifikasi Penyakit Daun Jagung," *Dinamik*, vol. 26, no. 2, pp. 58–67, 2021, doi: 10.35315/dinamik.v26i2.8673.
- [7] A. Y. Rahman, "Klasifikasi Citra Burung Lovebird Menggunakan Decision Tree dengan Empat Jenis Evaluasi," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, pp. 688–696, 2021, doi: 10.29207/resti.v5i4.3210.
- [8] Y. Florensia, S. Saparudin, and S. Samsuryadi, "Kombinasi 3D Convolutional Neural Network Dalam Klasifikasi Citra Hiperspektral Pada Tutupan Lahan," *J. Media Inform. Budidarma*, vol. 5, no. 4, p. 1611, 2021, doi: 10.30865/mib.v5i4.3183.
- [9] M. Ichwan and S. Hadi, "MIND (Multimedia Artificial Intelligent Networking Database Kinerja Model EfficientNetV2M dalam Klasifikasi Citra Tutupan dan Penggunaan Lahan," *J. MIND J. | ISSN*, vol. 8, no. 2, pp. 203–216, 2023, [Online]. Available: <https://doi.org/10.26760/mindjournal.v8i2.203-216>
- [10] E. T. L. Kusrini and E. Taufiq, "Algoritma data mining," *Yogyakarta Andi Offset*, 2009.
- [11] A. Purnamawati, W. Nugroho, D. Putri, and W. F. Hidayat, "Deteksi Penyakit Daun pada Tanaman Padi Menggunakan Algoritma Decision Tree, Random Forest, Naïve Bayes, SVM dan KNN," *InfoTekJar J. Nas. Inform. dan Teknol. Jar.*, vol. 5, no. 1, pp. 212–215, 2020, [Online]. Available: <https://doi.org/10.30743/infotekjar.v5i1.2934>
- [12] P. K. S. U. A. Mas and F. A. Santoso, "Implementasi Decision Tree Untuk Menentukan Status Pinjaman".
- [13] Sutoyo, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik," *J. Pilar Nusa Mandiri*, vol. 14, no. 217–224, 2018.
- [14] F. Y. Pamuji and V. P. Ramadhan, "Komparasi Algoritma Random Forest dan Decision Tree untuk Memprediksi Keberhasilan Immunotherapy," *J. Teknol. dan Manaj. Inform.*, vol. 7, no. 1, pp. 46–50, 2021, doi: 10.26905/jtmi.v7i1.5982.
- [15] D. Normawati and D. P. Ismi, "K-Fold Cross Validation for Selection of Cardiovascular Disease Diagnosis Features by Applying Rule-Based Datamining," *Signal Image Process. Lett.*, vol. 1, no. 2, pp. 23–35, Jul. 2019, doi: 10.31763/simple.v1i2.3.
- [16] F. Tempola, M. Muhammad, and A. Khairan, "Perbandingan Klasifikasi antara KNN dan Naïve Bayes Pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation," vol. 5, no. 5, pp. 577–584, 2018, doi: 10.25126/jtiik20185983.
- [17] A. Saputro, S. Mu'min, Moch. Lutfi, and H. Putri, "Deep Transfer Learning Dengan Model Arsitektur Vgg16 Untuk Klasifikasi Jenis Varietas Tanaman Lengkeng Berdasarkan Citra Daun," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 609–614, 2022, doi: 10.36040/jati.v6i2.5456.
- [18] G. Lukas Hansel and H. Bunyamin, "Penggunaan Augmentasi Data pada Klasifikasi Jenis Kanker Payudara dengan Model Resnet-34," *J. Strateg.*, vol. 3, no. 1, pp. 187–193, 2021.
- [19] Z. Setiawan *et al.*, *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia, 2023.
- [20] I. D. Id, *Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python*, vol. 1. Unri Press, 2021.
- [21] A. Dwi Putro and H. Tantyoko, "Hybrid Algoritma Vgg16-Net Dengan Support Vector Machine Untuk Klasifikasi Jenis Buah dan sayuran," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 5, no. 2, pp. 56–65, 2023, doi: 10.35746/jtim.v5i2.335.