

PERBANDINGAN METODE SPLIT ATRIBUT MENGGUNAKAN *INFORMATION GAIN* DAN *GAIN RATIO* PADA ALGORITMA C4.5 UNTUK KLASIFIKASI LAHAN KRITIS DI KABUPATEN GROBOGAN

Hevi Wahyu Nursafa'ah¹, Tiani Wahyu Utami², Rochdi Wasono³

¹²³Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Muhammadiyah Semarang

Alamat e-mail : heviwn97@gmail.com

ABSTRAK

Salah satu aspek penting dalam kehidupan manusia dan makhluk hidup lainnya adalah keberadaan lahan, akan tetapi persoalan kerusakan hutan dan lahan terus terjadi dan mengalami peningkatan sehingga mengakibatkan lahan menjadi kritis. Satu faktor penting dalam menentukan kesuksesan pemetaan penggunaan lahan terletak pada pemilihan skema klasifikasi lahannya. *Decision Tree* merupakan metode klasifikasi yang telah banyak digunakan. C4.5 merupakan algoritma *Decision Tree* yang memiliki tingkat akurasi yang baik. Split atribut merupakan proses utama dalam pembentukan pohon keputusan (*Decision Tree*). Penelitian ini menggunakan metode *Information Gain* dan *Gain Ratio* dalam pemilihan split atribut dan selanjutnya membandingkan hasil kinerja dari metode *Information Gain* dan *Gain Ratio*. Hasil kinerja di evaluasi dengan melihat nilai akurasi, *recall*, *precision* dan *F-measure*. Penelitian ini menggunakan data lahan kritis di Kabupaten Grobogan. Atribut yang digunakan adalah Vegetasi Permannen, Kemiringan Lereng Tingkat Bahaya Erosi, Manajemen Lahan. Sedangkan label yang digunakan adalah kekritisn lahan dengan kategori potensial kritis, kritis, agak kritis dan tidak kritis. Hasil dari metode *information gain* didapat nilai *recall* sebesar 95.46%, nilai *precision* sebesar 97.46%, nilai *F-measure* sebesar 96.46% dan nilai akurasi sebesar 95.24%. Hasil dari metode *gain ratio* didapat nilai *recall* sebesar 70.97%, nilai *precision* sebesar 95.16%, nilai *F-measure* sebesar 81.30% dan nilai akurasi sebesar 93.04%. Sehingga metode split atribut *information gain* lebih baik digunakan dalam klasifikasi lahan kritis di Kabupaten Grobogan.

Kata Kunci: Lahan Kritis, Klasifikasi, C4.5, *Gain Ratio*, *Information Gain*

ABSTRACT

One of the important aspects of humans life and other living things is a land existence, however forest and land damage is still happening and increasing thus causing our land to become critical. One of the main factors that determine the success of mapping land utilization is the selection of the scheme land classification. *Decision Tree* is a classification method which has been widely used. C4.5 is an algorithms *Decision Tree* that have a good level of accuracy. A split attribute is the main process of the *Decision Tree*. This research used *Information Gain* and *Gain Ratio* method in the election of the split attribute and then compared the results of *Information Gain* and *Gain Ratio* method. The results were evaluated by looking at the accuracy value, the recall, precision and F-measure. This research used the data of critical land from Grobogan District. The attributes include Permanent Vegetation, Slope of the Erosion Hazard Level, Land Management. While the label used land criticality with potential critical categories, critical, approximately critical, and uncritical. The result of the recall score of information gain method is 95.16%, precision score 97.46%, F-measure score 96.46% and the accuracy score 95.24%. The result of the recall score of gain ratio is 70.97%, precision score 95.16%, F-measure score 81.30% and the accuracy score 93.04%. Therefore split attribute information gain method is better to be used in the classification of critical land in the Grobogan District.

Keywords: Critical Land, Classification, C4.5, *Gain Ratio*, *Information Gain*

PENDAHULUAN

Keberadaan lahan merupakan aspek penting dalam kehidupan manusia dan makhluk hidup lainnya, akan tetapi persoalan kerusakan hutan dan lahan terus terjadi dan mengalami peningkatan sehingga mengakibatkan lahan menjadi kritis (Ramayanti, Yuwono, & Awaluddin, 2015). Akibat pemanfaatan sumber daya alam yang berlebih dan tidak dibarengi dengan usaha konservasi tanah dan air, juga dapat menimbulkan ketidakseimbangan lingkungan yaitu terus meningkatnya luas lahan kritis (Nugroho, 2000). Dalam Keputusan Menteri Kehutanan Nomor 52/KptsII/2001 tentang pedoman penyelenggaraan pengelolaan Daerah Aliran Sungai dijelaskan bahwa lahan kritis merupakan lahan yang telah mengalami kerusakan sehingga lahan tersebut tidak dapat berfungsi secara baik sesuai dengan peruntukannya sebagai media produksi maupun sebagai media tata air (Prabandaru, Nugraha, & Sukmono, 2016). Penentuan lahan kritis dilakukan dengan pembagian 3 kawasan yaitu kawasan budidaya pertanian, kawasan hutan lindung dan kawasan lindung di luar kawasan hutan. Menentukan tingkat kekritisian lahan mengacu pada Peraturan Direktur Jenderal Bina Pengelolaan Daerah Aliran Sungai dan Perhutanan Sosial tentang petunjuk teknis penyusunan data spasial lahan kritis dengan peraturan nomor P.4/V-SET/2013, yaitu pemberian skor serta pembobotan tiap parameter pada masing-masing Kawasan (Rosyada, Prasetyo, & Hani'ah, 2015).

Laju kerusakan hutan di Indonesia tercatat telah mencapai 59,63 juta hektar. Yaitu Kawasan hutan konservasi mencapai 4,7 hektar, hutan lindung mencapai 10,5 juta hektar dan hutan alam produksi mencapai 44,4 juta hektar (Zuhud, 2009). Kementerian Kehutanan mencanangkan kegiatan Rehabilitasi Hutan dan Lahan (RHL) untuk mengatasi lahan kritis serta memulihkan, mempertahankan dan meningkatkan fungsi hutan dan lahan sehingga daya dukung, produktivitas dan peranannya sebagai penyangga kehidupan

tetap terjaga (Indrihastuti, Murtilaksono, & Tjahjono, 2016). Oleh karena itu, peta dan data lahan kritis adalah suatu hal yang sangat penting untuk penyusunan program RHL agar efektif dan program berjalan dengan optimal, serta untuk harmonisasi program antar sektor yang terkait dengan rehabilitasi hutan dan lahan (Dephut, 2014).

Kabupaten Grobogan merupakan salah satu kabupaten di Provinsi Jawa Tengah dengan luas 1976 km². Kabupaten Grobogan tercatat memiliki lahan kritis seluas 428.687 ha. Sebesar 238.170 ha adalah lahan kritis di luar kawasan hutan dan 190.517 ha merupakan lahan kritis di dalam kawasan hutan (Ariyanto, Sunarminto, & Shiddieq, 2011). Berdasarkan laporan Badan Pusat Statistik tahun 2016 Kabupaten Grobogan menempati urutan pertama di Kabupaten Grobogan yang memiliki luas lahan kritis sebesar 203.131,10 ha (BPS, 2016). Parameter yang menjadi faktor kekritisian lahan, yaitu tingkat bahaya erosi, produktifitas lahan, erosi lahan, manajemen lahan, singkapan batuan, kemiringan lereng dan curah hujan (Hendro HS, Nahdi, Budiastuti, & Purnomo, 2012). Pengklasifikasian lahan kritis akan berfungsi untuk mengelaskan berbagai macam tingkatan lahan kritis berdasarkan rentang nilai hasil berbagai parameter (Prasetya & Gunawan, 2006).

Decision Tree merupakan metode klasifikasi yang telah banyak digunakan, karena merupakan salah satu metode yang populer (Handarko, 2015). *Decision tree* banyak digunakan karena kemampuan mereka dalam menghasilkan konsep yang jelas (Bhavitha & Madhuri, 2012). Kefleksibelan membuat metode ini atraktif, khususnya karena memberikan keuntungan berupa visualisasi saran (dalam bentuk *Decision tree*) yang membuat prosedur prediksinya dapat diamati (Gorunescu, 2011). Visualisasi saran tersebut dikenal sederhana dan menarik bahkan untuk praktisi awam yang tidak memiliki latar belakang statistik (Sohn & Kim, 2012).

Metode ini juga mampu melakukan klasifikasi sekaligus menunjukkan hubungan antar atribut (Ariadi & Arieshanti, 2015). Induksi di *Decision Tree* telah dikembangkan secara mandiri di statistik dan di komunitas *machine learning* (Rahayu, Satria, & Supriyanto, 2015). Ada beberapa pengembangan algoritma yang telah dikembangkan berdasarkan *Decision Tree* seperti ID3 (*Iterative Dichotomiser 3*), C4.5, CART (*Classification and Regression Tree*), CHAID (*Chi-squared Automatic Interaction Detector*) (Gupta, Rawat, Jain, Arora, & Dhama, 2017).

Pada penelitian terkait komparasi ID3, CART dan C4.5 yang dilakukan oleh Mohankumar, Amuthakkani, dan Jeyamala pada tahun 2016 dengan menggunakan dataset kelayakan pinjaman nasabah bank diperoleh hasil bahwa C4.5 memiliki hasil *precision*, *recall*, dan akurasi lebih baik daripada ID3 dan CART. Lim dan shan shih juga telah membandingkan tingkat akurasi, kompleksitas dan waktu training dari ID3, CART dan C4.5 dan hasilnya menunjukkan bahwa C4.5 mempunyai tingkat akurasi yang bagus dan mudah untuk diinterpretasikan (Lim, Loh, Shih, & Cohen, 2000).

Split atribut merupakan proses utama dalam pembentukan pohon keputusan (*Decision Tree*) di C4.5 (Zhang & Jiang, 2012). Ada dua metode pemilihan *split* atribut paling populer yaitu *information gain* dan *gain ratio* (Han, Kamber, & Pei, 2012). Metode *information gain* yang diperkenalkan Quinlan cocok untuk *dataset* dengan variabel diskret akan tetapi metode yang diperkenalkan tidak cocok untuk *dataset* dengan *missing value*, sedangkan metode *gain ratio* dapat bekerja untuk *dataset* dengan variabel kontinu dan *missing value*. Penelitian yang menggunakan *gain ratio* untuk *split* atribut Chaurasia pada tahun 2018 menggunakan *dataset Breast Cancer Wisconsin* dengan jumlah data 683, 10 atribut dan 1 label menghasilkan nilai akurasi sebesar 93.41%, nilai *sensivity* 93.4% dan nilai *specificity* sebesar 90.37% (Chaurasia, Pal, & Tiwari,

2018). Penelitian sebelumnya yang menggunakan *information gain* untuk *split* atribut yang dilakukan oleh Wisti Dwi Septiani pada tahun 2017 menggunakan data penyakit hepatitis yang didapat dari *Machine Learning Repository UCI* (Universitas California Invene) dengan 155 data, 19 atribut dan 1 label menghasilkan nilai akurasi sebesar 77.29% (Septiani, 2017). Berdasarkan uraian diatas penulis akan menggunakan *Information Gain* dan *Gain Ratio* untuk *split* atribut pada Algoritma C4.5 dengan studi kasus klasifikasi lahan kritis di Kabupten Grobogan. Tujuan yang hendak dicapai adalah Mengetahui hasil klasifikasi Algoritma C4.5 jika menggunakan metode pemilihan *split* atribut *Information Gain* dan *Gain Ratio* untuk Klasifikasi Lahan Kritis di Kabupaten Grobogan. Pemilihan metode *split* atribut terbaik antara *Information Gain* dan *Gain Ratio* pada *decision tree* Algoritma C4.5 untuk klasifikasi Lahan Kritis di Kabupaten Grobogan.

TINJAUAN PUSTAKA

1. Algoritma C4.5

Algoritma C4.5 mirip sebuah struktur pohon dimana terdapat node internal yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiao daun menggambarkan kelas (Han et al., 2012). Algoritma C4.5 merupakan generasi baru dari algoritma ID3 (*iterative dichotomiser3*) yang dikembangkan oleh J. Ross Quinlan pada tahun 1983. Algoritma C4.5 menggunakan konsep *information gain* atau *gain ratio* untuk menentukan *split* yang optimal (Larose, 2005).

2. Metode Split Atribut

a. *Information Gain*

Metode penelitian ini diperkenalkan oleh Quinlan dengan berdasar model ID3 (*Iterative Dichotomiser 3*). Metode yang diperkenalkan Quinlan cocok untuk *dataset* dengan variabel diskret akan tetapi metode

yang diperkenalkan tidak cocok untuk *dataset* dengan *missing value*. Metode penelitian Quinlan menggunakan pemilihan *split* atribut yang disebut *Gain*. Sebelum mencari nilai *gain*, hitung nilai *Entropy*. *Entropy* didefinisikan sebagai nilai informasi yang diharapkan. Nilai *Entropy* dapat dihitung melalui rumus persamaan dibawah ini:

$$\text{Entropy (S)} = \sum_{i=1}^n -p_i \times \log_2 p_i$$

Berdasarkan persamaan diatas, berikut merupakan keterangannya:

S = himpunan kasus

n = jumlah partisi S

p_i = proporsi dari S_i terhadap S

Perhitungan nilai *Gain* dapat dirumuskan dengan persamaan dibawah ini:

$$\text{Gain (S, A)} = \text{Entropy (s)} - \sum_{i=1}^m \frac{|S_i|}{|S|} \times \text{Entropy (S}_i)$$

Berdasarkan persamaan diatas, berikut merupakan keterangannya:

S: himpunan dataset

A: kriteria atau atribut

m: jumlah nilai yang mungkin pada kriteria A (jumlah kelas)

S_i : himpunan dataset untuk nilai m

$|S_i|$: jumlah dataset untuk nilai m

$|S|$: jumlah dataset dalam S

b. Gain Ratio

Metode penelitian ini merupakan pengembangan dari metode *Iterative Dichotomiser 3 (ID3)*. Quinlan memperkenalkan metode ini untuk pemilihan *split* atribut menggunakan metode *Info Gain Ratio (IGR)* memperbaiki

Information Gain (IG). C4.5 yang diperkenalkan dapat bekerja pada variabel kontinyu dan *missing value*.

Rumus persamaan *Gain Ratio* seperti berikut:

$$\text{Gain Ratio (S, A)} = \frac{\text{Gain (S, A)}}{\text{Split Info (S, A)}}$$

Dimana *split info* mempunyai persamaan rumus sebagai berikut:

$$\text{Split Information (S, A)} = - \sum_{i=1}^n \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}$$

3. Evaluasi Hasil

a. Sensitivitas (Recall)

Sensitivitas adalah kemampuan pengklasifikasi untuk memilih semua kasus yang harus dipilih. Sensitivitas disebut juga *recall* atau *True Positive Rate (TPRate)*. Persamaan berikut digunakan untuk menghitung sensitivitas.

$$\text{Sensitivity} = \text{recall} = \frac{TP}{TP + FN}$$

b. Precision

Precision digunakan untuk mengukur berapa persen ketepatan prediksi kelas dari suatu data, yaitu berapa banyak kelas yang diprediksi positive dan sesuai dengan kenyataannya.

$$\text{Precision} = \frac{TP}{TP + FP}$$

c. Accuracy

Akurasi didefinisikan sebagai kemampuan pengklasifikasi untuk memilih semua kasus yang harus dipilih dan menolak semua kasus yang harus ditolak. Jika sebuah pengklasifikasi mempunyai akurasi 100% berarti pengklasifikasi tersebut mempunyai False Negative (FN) = False Positive (FP) = 0. Akurasi dihitung menggunakan persamaan:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

d. F-Measure

F-Measure merupakan salah satu perhitungan evaluasi yang perhitungannya mengharmonikan rata-rata dari *precision* dan *recall* (Venkatesan, 2014). Sehingga menghasilkan metrik yang efektif untuk pencarian kembali informasi dalam himpunan yang mengandung masalah ketidakseimbangan. Dapat dihitung dengan persamaan:

$$F - Measure = \frac{2 \times sens \times prec}{sens + prec}$$

METODE PENELITIAN

1. Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari BPDAS Pemali Jratun, karena pelaporan dilakukan empat tahun sekali dan publikasi terakhir adalah tahun 2016 maka penelitian ini menggunakan data lahan kritis tahun 2016. Data ini berupa data kuantitatif yaitu data lahan kritis di Kabupaten Grobogan, dengan empat atribut dan satu label. Jumlah keseluruhan data yang digunakan sebanyak 30492 records.

2. Atribut Penelitian

Variabel atau Atribut yang digunakan antara lain Vegetasi Permannen, Kemiringan Lereng Tingkat Bahaya Erosi, Manajemen Lahan. Sedangkan label yang digunakan adalah kekritisian lahan dengan kategori potensial kritis, kritis, agak kritis dan tidak kritis.

3. Langkah Penelitian

Langkah-langkah eksperimen dan pengujian dari metode split atribut tersebut ditunjukkan pada alur dibawah ini:

1. Masukkan Dataset
2. Menggunakan pemilihan *split* atribut menggunakan *Information Gain*
 - a. Menghitung *Entropy dataset*
 - b. Menghitung *Entropy* setiap *class* pada atribut
 - c. Menghitung *Information Gain* setiap atribut

- d. Nilai *Info Gain* terbesar dijadikan sebagai akar
 - e. Menghitung lagi nilai *Information Gain* tiap-tiap atribut dengan tidak mengikutsertakan atribut yang terpilih menjadi akar (*root*) di tahap sebelumnya.
 - f. Atribut yang memiliki *Information Gain* tertinggi dipilih menjadi cabang (*branches*).
 - g. Mengulangi langkah e dan f sampai dengan dihasilkan nilai *Gain* = 0 untuk semua atribut yang tersisa.
 - h. Pembentukan pohon keputusan (*Decision Tree*) dengan split atribut *Information Gain*.
 - i. Pengklasifikasi akhir, dimana data *testing* diklasifikasikan menjadi *class* Kritis, Potensial Kritis, Agak Kritis dan Tidak Kritis.
3. Metode pemilihan *split* atribut menggunakan *Gain Ratio*
 - a. Menghitung *Entropy dataset*
 - b. Menghitung *Entropy* setiap *class* pada atribut
 - c. Menghitung *Information Gain* setiap atribut
 - d. Menghitung *Split Info*
 - e. Menghitung *Gain Ratio*
 - f. Nilai *Gain Ratio* terbesar dijadikan sebagai akar
 - g. Menghitung lagi nilai *Gain Ratio* tiap-tiap atribut dengan tidak mengikutsertakan atribut yang terpilih menjadi akar (*root*) di tahap sebelumnya.
 - h. Atribut yang memiliki *Gain Ratio* tertinggi dipilih menjadi cabang (*branches*)
 - i. Mengulangi langkah h dan i sampai dengan dihasilkan nilai *Gain* = 0 untuk semua atribut yang tersisa.
 - j. Pembentukan pohon keputusan (*Decision Tree*) dengan split atribut *Gan Ratio*
 - k. Pengklasifikasi akhir, dimana data *testing* diklasifikasikan menjadi *class* Kritis, Potensial Kritis, Agak Kritis dan Tidak Kritis.

4. Membandingkan kedua model dengan evaluasi hasil *precision*, *recall*, *F-measure* dan *accuracy*.
5. Menarik Kesimpulan.

HASIL DAN PEMBAHASAN

1. Metode *Information Gain*

Hasil dari eksperimen dilakukan dengan menggunakan metode pemilihan *split* atribut *Information Gain* menghasilkan *ruleset* sebanyak 70 *ruleset*. Contoh *ruleset* yang dihasilkan sebagai berikut:

```

Rule1: IF Kelas Vegetasi = Sangat Baik
THEN Tidak Kritis
Rule2: IF Kelas Vegetasi= Sangat
Buruk AND Kelas Lereng = Agak
Curam THEN Kritis
Rule3: IF Kelas Vegetasi= Sangat
Buruk AND Kelas Lereng = Datar
AND Kelas Erosi = Berat THEN Agak
Kritis
Rule4: IF Kelas Vegetasi= Sangat
Buruk AND Kelas Lereng = Datar
AND Kelas Erosi = Sangat Berat
THEN Kritis
:
:
:
Rule67: IF Kelas Vegetasi= Buruk
AND Kelas Erosi = Sedang AND Kelas
Lereng = Curam AND Kelas
Manajemen = Sedang THEN Kritis
Rule68: IF Kelas Vegetasi= Buruk
AND Kelas Erosi = Sedang AND Kelas
Lereng = Datar THEN Agak Kritis
Rule69: IF Kelas Vegetasi= Buruk
AND Kelas Erosi = Sedang AND Kelas
Lereng = Landai THEN Agak Kritis
Rule70: IF Kelas Vegetasi= Buruk
AND Kelas Erosi = Sedang AND Kelas
Lereng = Sangat Curam THEN Agak
Kritis

```

Gambar 1. Contoh *Ruleset* Metode *Information Gain*

Berdasarkan *confusion matrix* yang terbentuk dari Metode *Information Gain*

menghasilkan nilai *recall* sebesar 95.46%, nilai *precision* sebesar 97.46%, nilai *F-measure* sebesar 96.46% dan nilai akurasi sebesar 95.24%.

2. Metode *Gain Ratio*

Hasil dari eksperimen dilakukan dengan menggunakan metode pemilihan *split* atribut *Gain Ratio* menghasilkan *ruleset* sebanyak 35 *ruleset*. Contoh *ruleset* yang dihasilkan sebagai berikut:

```

Rule1: IF Kelas Vegetasi = Buruk
THEN Agak Kritis
Rule2: IF Kelas Vegetasi = Sangat Baik
THEN Tidak Kritis
Rule3: IF Kelas Vegetasi= Baik AND
Kelas Manajemen = Baik AND Kelas
Erosi = Berat AND Kelas Lereng =
Agak Curam THEN Potensial Kritis
Rule4: IF Kelas Vegetasi= Baik AND
Kelas Manajemen = Baik AND Kelas
Erosi = Berat AND Kelas Lereng =
Curam THEN Potensial Kritis
:
:
:
Rule31: IF Kelas Vegetasi= Sedang
AND Kelas Manajemen = Sedang
AND Kelas Erosi = Sangat Ringan
THEN Potensial Kritis
Rule32: IF Kelas Vegetasi= Sedang
AND Kelas Manajemen = Sedang
AND Kelas Erosi = Sedang AND Kelas
Lereng = Agak Curam THEN Agak
Kritis
Rule33: IF Kelas Vegetasi= Sedang
AND Kelas Manajemen = Sedang
AND Kelas Erosi = Sedang AND Kelas
Lereng = Curam THEN Agak Kritis
Rule34: IF Kelas Vegetasi= Sedang
AND Kelas Manajemen = Sedang
AND Kelas Erosi = Sedang AND Kelas
Lereng = Datar THEN Potensial Kritis
Rule35: IF Kelas Vegetasi= Sedang
AND Kelas Manajemen = Sedang
AND Kelas Erosi = Sedang AND Kelas
Lereng = Landai THEN Agak Kritis

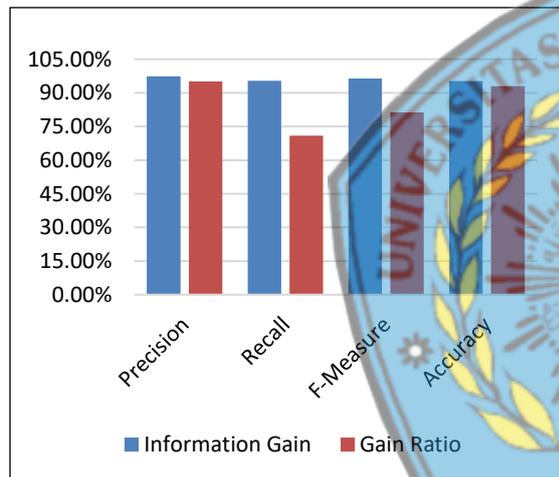
```

Gambar 2. Contoh *Ruleset* Metode *Gain Ratio*

Berdasarkan *confusion matrix* yang terbentuk dari Metode *Gain Ratio* menghasilkan nilai *recall* sebesar 70.97%, nilai *precision* sebesar 95.16%, nilai *F-measure* sebesar 81.30% dan nilai akurasi sebesar 93.04%.

3. Perbandingan Metode *Information Gain* dan *Gain Ratio*

Precision, *Recall*, *F-Measure* dan Tingkat akurasi klasifikasi Algoritma C4.5 dengan menggunakan Metode split atribut *Information Gain* dan *Gain Ratio* akan dibandingkan untuk mengetahui metode terbaik untuk mengklasifikasikan Lahan Kritis di Kabupaten Grobogan.



Gambar 3. Grafik Perbandingan Metode *Information Gain* dan *Gain Ratio*

Berdasarkan Gambar 4.10 menunjukkan bahwa nilai *precision*, *recall*, *f-measure* dan tingkat akurasi dari Metode *Information Gain* lebih tinggi yaitu nilai *precision* sebesar 97.46% di bandingkan dengan *precision* menggunakan *Gain Ratio* yaitu sebesar 95.16%, nilai *recall* sebesar 70.97% di bandingkan dengan *precision* menggunakan *Gain Ratio* yaitu sebesar 70.97%, nilai *F-Measure* sebesar 81.30% di bandingkan dengan *F-Measure* menggunakan *Gain Ratio* yaitu sebesar 81.30%, dan nilai *accuracy* sebesar 93.04% di bandingkan dengan *accuracy* menggunakan *Gain Ratio* yaitu sebesar 93.04%. Hal ini sesuai dengan teori yang

mengatakan bahwa metode split atribut *Information Gain* lebih baik digunakan pada variabel atau atribut kategorik.

KESIMPULAN

1. Hasil klasifikasi menggunakan metode split atribut *Information Gain* menghasilkan *rule* sebanyak 70 *ruleset* dan menghasilkan nilai *recall* sebesar 95.46%, nilai *precision* sebesar 97.46%, nilai *F-measure* sebesar 96.46% dan nilai akurasi sebesar 95.24%.
2. Hasil klasifikasi menggunakan metode split atribut *Gain Ratio* menghasilkan *rule* sebanyak 35 *ruleset* dan menghasilkan nilai *recall* sebesar 70.97%, nilai *precision* sebesar 95.16%, nilai *F-measure* sebesar 81.30% dan nilai akurasi sebesar 93.04%.
3. Berdasarkan evaluasi model kedua metode split atribut tersebut yang dilihat dari nilai *recall*, nilai *precision*, nilai *F-measure* dan nilai akurasi, menyatakan bahwa metode *Information Gain* memiliki nilai lebih unggul daripada metode *Gain Ratio*. Dapat dikatakan metode split atribut *information gain* lebih baik digunakan pada kasus klasifikasi lahan kritis di Kabupaten Grobogan.

DAFTAR PUSTAKA

- Ariadi, R., & Arieshanti, I. (2015). Implementasi Metode Pohon Keputusan untuk Klasifikasi Data Dengan Nilai Fitur yang Tidak Pasti, (June 2015).
- Ariyanto, D. P., Sunarminto, B. H., & Shiddieq, D. (2011). Kajian konservasi tanah kritis berdasarkan satuan lahan di daerah tangkapan air sempor kabupaten kebumen, 8(2).
- Bhavitha, A. M., & Madhuri, S. (2012). A Classification Method using Decision Tree for Uncertain Data. *International Journal of Computer Trends and Technology*, 3.

- BPS. (2016). Luas Lahan Kritis di Luar Kawasan Hutan Menurut Kabupaten/Kota di Provinsi Jawa Tengah, 2016 (ha).
- Chaurasia, V., Pal, S., & Tiwari, B. B. (2018). Prediction of benign and malignant breast cancer using data mining techniques. *Journal of Algorithms & Computational Technology*, 12(2), 119–126. <https://doi.org/10.1177/1748301818756225>
- Dephut. (2014). Lingkup Kementerian Kehutanan.
- Gorunescu, F. (2011). *Data Mining: Concepts, Models and Techniques. Data mining - Concepts, Models and Technique*. Springer. <https://doi.org/10.1007/978-3-642-19721-5>
- Gupta, B., Rawat, A., Jain, A., Arora, A., & Dharmi, N. (2017). Analysis of Various Decision Tree Algorithms for Classification in Data Mining, 163(8), 15–19.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (Third Edit). San Francisco: Morgan Kaufmann Publishers. <https://doi.org/10.1016/B978-0-12-381479-1.00001-0>
- Handarko, J. L. (2015). Implementasi Fuzzy Decision Tree untuk Mendiagnosa Penyakit Hepatitis, 4(2).
- Hendro HS, H., Nahdi, Z., Budiastuti, Mt. S., & Purnomo, D. (2012). Pemetaan Paramater Lahan Kritis Guna Mendukung Rehabilitasi Hutan Dan Lahan Untuk Kelestarian Lingkungan Dan Ketahanan Pangan Dengan Menggunakan Pendekatan Spasial Temporal Di Kawasan Muria. *Prosiding SNST Ke-6 Tahun 2015 Fakultas Teknik Universitas Wahid Hasyim Semarang*, 6(2015), 41–46.
- Indrihastuti, D., Murti Laksono, K., & Tjahjono, B. (2016). Analysis of Critical Land and Recommendation for Land Rehabilitation In The Regional Development In Kendal, Central Java. *Tata Loka*, 18(3), 141–156.
- Larose, D. T. (2005). *Discovering Knowledge in Data*. United States of America: John Wiley & Sons, Inc.
- Lim, T.-S., Loh, W.-Y., Shih, Y.-S., & Cohen, W. W. (2000). A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-three Old and New Classification Algorithms. *Machine Learning*, 40(1992), 203–229. <https://doi.org/10.1023/A:1007608224229>
- Mohankumar, M., Amuthakkani, S., & Jeyamala, G. (2016). Comparative Analysis of Decision Tree Algorithms for the Prediction of Eligibility of a Man for Availing Bank Loan. *International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)*, 2(15), 360–366.
- Nugroho, S. P. (2000). Minimalisasi Lahan Kritis Melalui Pengelolaan Sumberdaya lahan dan Konservasi Tanah dan Air Secara Terpadu. *Jurnal Teknologi Lingkungan*, 1(1, Januari), 73–82.
- Prabandaru, L. H., Nugraha, A. L., & Sukmono, A. (2016). PEMETAAN TINGKAT LAHAN KRITIS KABUPATEN WONOSOBO DENGAN PENGINDERAAN JAUH DAN SISTEM INFORMASI GEOGRAFIS (Studi Kasus: Kec. Kejajar, Kec. Garung, Kec. Mojo Tengah). *Jurnal Geodesi Undip*, 5, 65–72.
- Prasetya, R. N., & Gunawan, T. (2006). Pemanfaatan Citra Penginderaan Jauh dan SIG untuk Pemetaan Lahan Kritis di Daerah Kokap dan Pengasih Kab.

- Kulonprogo, (167), 281–290.
<https://doi.org/10.1057/palgrave.jibs.8490907>
- Rahayu, E. S., Satria, R., & Supriyanto, C. (2015). Penerapan Metode Average Gain , Threshold Pruning dan Cost Complexity Pruning untuk Split Atribut pada Algoritma C4 . 5, 1(2), 91–97.
- Ramayanti, L. A., Yuwono, B. D., & Awaluddin, M. (2015). PEMETAAN TINGKAT LAHAN KRITIS DENGAN MENGGUNAKAN PENGINDERAAN JAUH DAN SISTEM INFORMASI GEOGRAFI (Studi Kasus : Kabupaten Blora). *Jurnal Geodesi Undip*, 4, 200–207.
- Rosyada, M., Prasetyo, Y., & Hani'ah. (2015). Penentuan Tingkat Lahan Kritis Menggunakan Metode Pemboboan Dan Algoritma NDVI. *Jurnal Geodesi Undip*, 4, 85–94.
- Septiani, W. D. (2017). Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 dan Naive Bayes untuk Prediksi Penyakit Hepatitis. *Jurnal Pilar Nusa Mandiri Volume*, 13(1), 76–84.
- Sohn, S. Y., & Kim, J. W. (2012). Decision tree-based technology credit scoring for start-up firms: Korean case. *Expert Systems with Applications*, 39(4), 4007–4012.
<https://doi.org/10.1016/j.eswa.2011.09.075>
- Venkatesan, R. (2014). Multi-Label Classification Method Based on Extreme Learning Machines, 2014(December), 10–12.
- Zhang, X., & Jiang, S. (2012). A Splitting Criteria Based on Similarity in Decision Tree Learning, 7(8), 1775–1782.
<https://doi.org/10.4304/jsw.7.8.1775-1782>
- Zuhud, E. a. M. (2009). POTENSI HUTAN

TROPIKA INDONESIA SEBAGAI PENYANGGA BAHAN OBAT ALAM UNTUK KESEHATAN BANGSA (The Indonesian Tropical Forest as Buffer of Natural Medicine Product for Nation Healthy) The Indonesian tropical forest producing the medicinal plants diversity as. *Jurnal Bahan Alam Indonesia*, 6(6), 227–232.