

## Implementation of C4.5 Algorithm for Critical Land Prediction in Agricultural Cultivation Areas in Pemali Jratun Watershed

<sup>1</sup>Deden Istiawan, <sup>2</sup>Laelatul Khikmah

<sup>1,2</sup>Akademi Statistika Muhammadiyah Semarang

Email: <sup>1</sup>dedenistiawan@gmail.com, <sup>2</sup>aisyah.salsabila17@gmail.com

---

### Article Info

#### Article history:

Received Aug 02<sup>th</sup>, 2019

Revised Aug 12<sup>th</sup>, 2019

Accepted Sep 12<sup>th</sup>, 2019

---

#### Keyword:

C4.5 Algorithm

Classification

Critical Land

Data Mining

Watershed

---

### ABSTRACT

Watershed is a complex system that is built on physical systems, biological systems and human systems that are related to each other. Each component has a distinctive nature and its existence is related to other components so as to form a unified ecosystem. Land use that does not pay attention to the conservation requirements of land and water causes land degradation which ultimately results in critical land. The impact of critical land is not only the withdrawal of soil properties, but also results in a decrease in production functions. Prediction of the critical level of land is needed to reduce the level of damage to the watershed, so that it can be used for policy making by the relevant agencies. In this research C4.5 algorithm will be applied to predictions of critical land in agricultural cultivation areas using critical land parameters. Based on the results of the research on critical land classification of agricultural cultivation areas in the jratun pemali watershed it can be concluded that the C.45 algorithm can be implemented to predict critical land in agricultural cultivation areas with an accuracy rate of 92.47%.

*Copyright © 2019 Puzzle Research Data Technology*

---

### Corresponding Author:

Deden Istiawan

Akademi Statistika Muhammadiyah Semarang

Jl. Prof. Dr. Hamka Km 01 Tambakaji Ngaliyan Semarang

Email: dedenistiawan@gmail.com

DOI: <http://dx.doi.org/10.24014/ijaidm.v2i2.7569>

---

## 1. PENDAHULUAN

Daerah Aliran Sungai (DAS) merupakan suatu system kompleks yang dibangun atas sistem fisik, sistem biologis dan sistem manusia yang terkait satu sama lain. Setiap komponennya memiliki sifat yang khas dan keberadaannya berhubungan dengan komponen lain sehingga membentuk kesatuan ekosistem [1]. Hal ini menggambarkan pentingnya analisis tata guna lahan untuk menghindari resiko degradasi pada area yang lebih luas [2]. Seiring dengan peningkatan jumlah penduduk, konversi lahan merupakan hal yang wajar terjadi. Namun pada kenyataannya konversi lahan menjadi masalah karena terjadi pada Kawasan pertanian yang masih produktif [3]. Pemanfaatan lahan yang tidak memperhatikan syarat konservasi tanah dan air menyebabkan degradasi lahan yang pada akhirnya menimbulkan lahan kritis [4].

Lahan kritis merupakan lahan yang telah mengalami kerusakan secara fisik, kimia dan biologis, sehingga lahan tersebut tidak dapat berfungsi secara baik sesuai dengan peruntukannya sebagai media produksi maupun sebagai media tata air [5]. Dampak dari lahan kritis tidak hanya pemunduran sifat-sifat tanah, namun juga mengakibatkan penurunan fungsi produksi. Fungsi produksi berarti lahan sebagai media tumbuh dan berkembang tanaman pertanian akan terganggu, sehingga akan mempengaruhi kehidupan sosial dan ekonomi masyarakat yang memanfaatkan lahan tersebut [6]. Identifikasi dan pemetaan lahan kritis sangat penting dilakukan untuk perencanaan dan penentuan DAS prioritas yang mengalami kerusakan [7]. Sehingga salah satu faktor penting dalam keberhasilan pemetaan penggunaan lahan terletak pada pemilihan skema klasifikasi lahan [8].

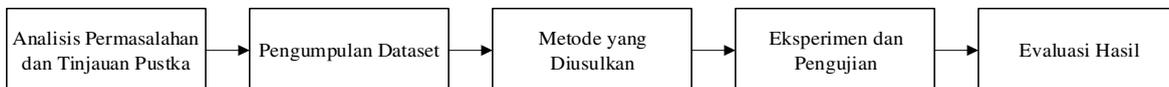
Prediksi tingkat kekritisannya lahan sangat diperlukan untuk mengurangi tingkat kerusakan DAS, sehingga dapat digunakan untuk pengambilan kebijakan oleh instansi terkait [4]. Dalam beberapa tahun terakhir sejumlah pendekatan klasifikasi telah dikembangkan [9] salah satunya adalah pendekatan data mining

[10]. Klasifikasi merupakan *supervise learning* yang membutuhkan label dalam prosesnya untuk mengekstrak model yang digunakan untuk memprediksi label [11]. Penelitian tentang klasifikasi lahan kritis pernah dilakukan oleh [12] algoritma Naïve Bayes diusulkan untuk prediksi lahan kritis, hasil penelitian menunjukkan akurasi algoritma Naïve Bayes mampu memprediksi lahan kritis sebesar 89.45%. Penelitian lain membandingkan algoritma C4.5, K-NN dan Naïve Bayes [8] untuk prediksi lahan kritis, hasil penelitian menunjukkan algoritma C4.5 lebih unggul dalam memprediksi lahan kritis. *Decision Tree* Salah satu algoritma yang sering digunakan dan mempunyai struktur yang sederhana dan mudah untuk diinterpretasikan [13].

Algoritma yang dikembangkan berdasar *Decision Tree* adalah CHAID dimana split tiap nodenya berdasarkan pada *chi-square test*, CART membentuk *Decision Tree* dengan menggunakan *Gini Index* untuk kriteria *split* tiap *nodenya*, ID3 menggunakan *Information Gain* untuk kriteria *splitnya*, C4.5 merupakan pengembangan dari ID3 yang tidak dapat menangani data numerik, C4.5 menggunakan *Gain Ratio* untuk kriteria *Splitnya*. Atribut dengan *Gain Ratio* tertinggi akan dipilih [14]. Seperti pada penelitian [15] membandingkan kinerja algoritma ID3, CART dan C4.5, hasil penelitian menunjukkan tingkat akurasi dan kompleksitas waktu algoritma C4.5 lebih unggul daripada algoritma yang lain. Pada penelitian ini algoritma C4.5 akan diterapkan prediksi lahan kritis pada Kawasan budidaya pertanian dengan menggunakan parameter lahan kritis.

## 2. METODE PENELITIAN

Pada penelitian ini menggunakan metode penelitian eksperimen. Metode penelitian eksperimen adalah uji coba yang dilakukan oleh peneliti untuk melakukan investigasi hubungan sebab akibat. Tahapan pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

### 2.1 Analisis Permasalahan dan Tinjauan Pustaka

Penelitian ini diawali dengan mengidentifikasi masalah pada topik klasifikasi lahan kritis. Selain itu mengidentifikasi dan mengevaluasi metode-metode yang pernah diusulkan untuk topik klasifikasi lahan kritis. Tahap identifikasi dapat dilakukan dengan mengumpulkan jurnal-jurnal, buku, prosiding dan referensi lainnya yang digunakan untuk bahan pertimbangan dalam menganalisis permasalahan yang akan diangkat pada penelitian ini.

### 2.2 Pengumpulan Dataset

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari BPDAS Pemali Jratun tahun 2013 karena pelaporan dilakukan empat tahun sekali dan publikasi terakhir adalah tahun 2013. Data ini berupa data kuantitatif yaitu data lahan kritis pada Kawasan budidaya pertanian, dengan empat atribut dan satu label. Atribut yang digunakan antara lain produktifitas, kemiringan lereng, tingkat erosi dan manajemen lahan. Sedangkan label yang digunakan adalah kekritisian lahan.

Tabel 1 Parameter Lahan Kritis Kawasan Budidaya Pertanian

Indikator	Jenis Data	Keterangan	Kategori
Produktivitas	Ordinal	Atribut	1. Sangat Rendah 2. Rendah 3. Sedang 4. Tinggi 5. Sangat Tinggi
Kemiringan Lereng	Ordinal	Atribut	1. Sangat Curam 2. Curam 3. Agak Curam 4. Landai 5. Datar
Erosi	Ordinal	Atribut	1. Ringan 2. Sedang 3. Berat 4. Sangat Berat
Manajemen Lahan	Ordinal	Atribut	1. Baik 2. Sedang 3. Buruk
Kekritisian Lahan	Ordinal	Kelas	1. Sangat Kritis 2. Kritis

Indikator	Jenis Data	Keterangan	Kategori
			3. Agak Kritis
			4. Potensial Kritis
			5. Tidak Kritis

### 2.3 Metode yang Diusulkan

Setelah mengumpulkan data dan referensi beberapa penelitian terdahulu maka penelitian ini mengusulkan metode Algoritma C4.5 untuk prediksi lahan kritis pada Kawasan budidaya pertanian di daerah aliran sungai pemali jratun. Data mining merupakan suatu proses untuk menemukan informasi tersembunyi dari kumpulan data yang besar [16]. Klasifikasi adalah salah satu bagian dari data mining yang dapat digunakan untuk menggambarkan dan membedakan kelas data [17]. Klasifikasi adalah proses untuk menentukan label dari suatu objek tertentu, untuk menentukan label tersebut, data dibagi menjadi data training dan data testing [14]. Data training digunakan untuk membentuk model klasifikasi, sedangkan data testing digunakan untuk menguji model yang terbentuk untuk mengetahui kinerja dari model klasifikasi tersebut. Apabila akurasi sudah cukup baik memprediksi label, maka model tersebut dapat digunakan untuk memprediksi data yang belum diketahui labelnya [11].

#### 2.3.1 Algoritma C4.5

*Decision Tree* adalah pohon yang digunakan sebagai prosedur penalaran untuk mendapatkan jawaban dari masalah yang dimasukan. Metode ini sangat fleksibel karena memberikan keuntungan berupa visualisasi dalam bentuk *Decision Tree* yang membuat prosedur prediksinya dapat diamati [14]. Algoritma C4.5 mirip sebuah struktur pohon dimana terdapat node internal yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji dan setiap daun menggambarkan kelas [11]. Algoritma C4.5 merupakan generasi baru dari algoritma ID3 yang dikembangkan oleh J. Ross Quinlan pada tahun 1996 [18]. Dalam ID3 induksi *Decision Tree* hanya bisa dilakukan pada atribut bertipe kategorikal. Sedangkan tipe numerik tidak dapat digunakan.

Algoritma C4.5 menggunakan konsep *entropy* untuk memilih atribut yang optimal secara rekursif sampai tidak ada cabang lagi yang mungkin dihasilkan [19]. *Entropy* dapat dihitung dengan menggunakan persamaan berikut:

$$E(s) = - \sum_{i=1}^m p(\omega_i|s) \log_2 p(\omega_i|s) \quad (1)$$

$p(\omega_i|s)$  adalah proporsi kelas- $i$  dalam semua data latih yang diproses di node  $s$ .  $p(\omega_i|s)$  didapatkan dari jumlah semua baris data dengan label kelas dibagi dengan jumlah baris semua data. Sementara  $m$  adalah jumlah nilai berbeda dalam data. Nilai *entropy* yang lebih tinggi akan meningkatkan potensi klasifikasi. Yang perlu diperhatikan adalah jika *entropy* untuk node bernilai 0 berarti semua data *vector* berada pada label kelas yang sama dan node tersebut menjadi daun yang berisi keputusan [20].

*Gain* digunakan untuk memperkirakan pemilihan atribut yang tepat untuk menjadi pemecah pada node tersebut. *Gain* sebuah atribut ke- $j$  dihitung menggunakan persamaan berikut:

$$Gain(s, j) = E(s) - \sum_{i=1}^n p(v_i|s) \times E(S_i) \quad (2)$$

$p(v_i|s)$  adalah proporsi nilai  $v$  muncul pada kelas dalam node.  $E(S_i)$  adalah *entropy* komposisi nilai  $v$  dari kelas ke- $j$  dalam data ke- $i$  node tersebut.  $n$  adalah jumlah nilai berbeda dalam node. Kriteria yang paling banyak digunakan untuk memilih fitur sebagai pemecah dalam algoritma C4.5 adalah *gain ratio*, yang diformulasikan oleh persamaan berikut:

$$Gain Rasio(s, j) = \frac{Gain(s, j)}{SplitInfo(s, j)} \quad (3)$$

Persamaan 3 menyatakan nilai *gain ratio* pada atribut ke- $j$ .  $SplitInfo(s, j)$  didapat dari

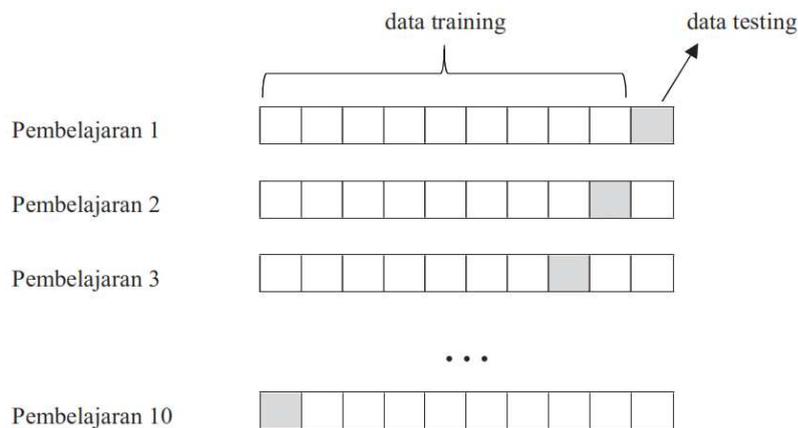
$$SplitInfo(s, j) = - \sum_{i=1}^k p(v_i|s) \log_2 p(v_i|s) \quad (4)$$

## 2.4 Eksperimen dan Pengujian

Pada tahap ini dilakukan eksperimen dengan menggunakan algoritma C.45 untuk prediksi lahan kritis pada Kawasan budidaya pertanian di Daerah Aliran Sungai Pemali Jratun. *Tools* yang digunakan dalam penelitian ini adalah Rapidminer 9.3

### 2.4.1 10-Fold Cross Validation

Model validasi klasifikasi yang sering digunakan adalah *x-fold cross validation* [21]. *x-fold cross validation* adalah Teknik validasi dengan membagi data secara acak ke dalam *x* bagian dan masing-masing bagian akan dilakukan proses klasifikasi [11]. Dengan menggunakan *cross validation* akan dilakukan percobaan sebanyak *x*. *10-fold cross validation* merupakan *cross validation* yang direkomendasikan untuk perhitungan akurasi [22]. Data yang digunakan dalam penelitian ini adalah data training untuk mencari error rate secara keseluruhan, secara umum pengujian nilai *x* dilakukan sebanyak 10 kali untuk memperkirakan akurasi estimasi seperti pada Gambar 2 ilustrasi *10-Cross Fold Validation*



Gambar 2. 10-Fold Cross Validation

## 2.5 Evaluasi Model

Evaluasi hasil dari eksperimen merupakan sebuah alat ukur yang dapat digunakan untuk menilai atau mengukur seberapa baik metode yang diusulkan terhadap metode lainnya dan apakah metode yang diusulkan mempunyai perbedaan hasil yang signifikan terhadap model lainnya. Suatu sistem yang akan melakukan klasifikasi diharapkan dapat melakukan klasifikasi di semua dataset dengan benar. Akan tetapi, tidak dapat dipungkiri bahwa kinerja suatu sistem tidak bisa bekerja 100% benar. Oleh karena itu, sebuah sistem klasifikasi juga diukur kinerjanya. Umumnya cara mengukur kinerja klasifikasi menggunakan matriks confusion [20].

### 2.5.1 Akurasi

Akurasi merupakan salah satu indikator yang dipakai dalam penelitian ini untuk mengukur performa dari metode yang diusulkan. Nilai akurasi dihitung dengan mengambil persentase prediksi yang benar dari keseluruhan data. Prediksi yang tepat berarti hasil dimana label hasil prediksi adalah sama dengan label dari data. Perhitungan akurasi dapat dilihat pada Tabel 2.

Tabel 2. Perhitungan Akurasi

Label	Kenyataan Benar	Kenyataan Salah
Prediksi Benar	TP ( <i>True Positive</i> )	FP ( <i>False Positive</i> )
Prediksi Salah	FN ( <i>False Negative</i> )	TN ( <i>True Negative</i> )

Dengan mengetahui jumlah data yang akan diklasifikasikan secara benar dapat diketahui akurasi hasil prediksi dan dengan mengetahui jumlah data yang akan diklasifikasikan secara salah maka dapat diketahui laju *error* dari prediksi yang dilakukan. Untuk menghitung akurasi digunakan rumus sebagai berikut:

$$Akurasi = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \quad (5)$$

Semua algoritma klasifikasi berusaha untuk membentuk model yang mempunyai akurasi tinggi. Umumnya model yang dibangun dapat memprediksi dengan benar pada semua data yang menjadi data latihnya, tetapi ketika model berhadapan dengan data uji barulah kinerja model dari algoritma klasifikasi ditentukan.

### 2.5.2 Sensitivitas atau Recall

Sensitivitas atau *Recall* dalam bidang ilmu pencarian informasi mengukur proporsi positif asli yang diprediksi secara benar sebagai positif. Sensitivitas berhubungan dengan kemampuan pengujian untuk mengidentifikasi hasil yang positif dari sejumlah data yang sebenarnya positif. Persamaan yang digunakan untuk menghitung adalah sebagai berikut:

$$\text{Sensitivitas} = \frac{TP}{TP+FN} \quad (6)$$

### 2.5.3 Precision

Dalam bidang pencarian informasi, *Precision* atau disebut juga *positive prediction value* merupakan matrik untuk mengukur kinerja sistem dalam mendapatkan data yang relevan. Sementara *Recall* merupakan matrik untuk mengukur kinerja dalam mendapatkan data yang terbaca. Dalam data mining, *Precision* adalah jumlah data yang *true positive* dibagi dengan jumlah data yang dikenali sebagai positif. Persamaan berikut digunakan untuk menghitung *Precision*:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

## 3. HASIL DAN PEMBAHASAN

Hasil implementasi algoritma C4.5 untuk prediksi lahan kritis pada Kawasan budidaya pertanian di Daerah Aliran Sungai Pemali Jratun. Pada penelitian ini digunakan *10-fold cross validation* untuk mencari *error rate* secara keseluruhan. Pada Tabel 3 dapat dilihat hasil *stratified 10-fold cross validation*, bahwa tingkat kebenaran dari hasil klasifikasi berjumlah 102.641 dari 111.003 data dengan tingkat akurasi model sebesar 92.47%. Tingkat kesalahan dari hasil klasifikasi 8.362 dari 111.003 data dengan tingkat kesalahan klasifikasi sebesar 7.53%.

Tabel 3. *Confusion Matrix* Algoritma C4.5

	Tidak Kritis	Potensial Kritis	Agak Kritis	Kritis	Sangat Kritis	Precision
Tidak Kritis	72727	5127	32	33	8	93.33%
Potensial Kritis	2197	17831	519	1	2	86.77%
Agak Kritis	28	263	10217	67	0	96.61%
Kritis	7	0	39	1651	23	95.99%
Sangat Kritis	3	0	0	13	215	93.07%
Recall	97.02%	76.79%	94.54%	93.54%	86.69%	

Tingkat akurasi tersebut didukung dengan hasil yang ditunjukkan pada Tabel 3 yaitu hasil *Precision* dan *Recall*. *Precision* merupakan matrik untuk mengukur kinerja sistem dalam mendapatkan data yang relevan. Pada penelitian ini diperoleh nilai *Precision* yang tinggi disemua label lahan kritis yaitu 93.33% pada label tidak kritis, 86.77% pada label potensial kritis, 96.61% pada label agak kritis, 95.99% pada label kritis dan 93.07% pada label sangat kritis. Sedangkan *Recall* adalah mengukur proporsi positif asli yang diprediksi secara benar sebagai positif. Pada penelitian ini diperoleh nilai 97.02 pada label tidak kritis, 76.79% pada label potensial kritis, 94.54% pada label agak kritis, 93.54 pada label kritis dan 86.69 pada label sangat kritis.

Dengan menggunakan algoritma C.45 dapat diketahui informasi lain yaitu berupa *rule* atau aturan dari pohon keputusan yang dapat digunakan untuk prediksi lahan kritis pada Kawasan budidaya pertanian. Hasil *rule* atau aturan struktur pohon sebagai berikut:

```

MANAJEMEN = Baik
| LERENG = Agak Curam: Tidak Kritis {Tidak Kritis=4397, Potensial Kritis=3738, Agak Kritis=16, Kritis=0, Sangat Kritis=8}
| LERENG = Curam: Potensial Kritis {Tidak Kritis=525, Potensial Kritis=2205, Agak Kritis=279, Kritis=0, Sangat Kritis=2}
| LERENG = Datar: Tidak Kritis {Tidak Kritis=53728, Potensial Kritis=4273, Agak Kritis=6, Kritis=9, Sangat Kritis=0}
| LERENG = Landai: Tidak Kritis {Tidak Kritis=14562, Potensial Kritis=2710, Agak Kritis=16, Kritis=22, Sangat Kritis=0}
| LERENG = Sangat Curam: Potensial Kritis {Tidak Kritis=42, Potensial Kritis=571, Agak Kritis=86, Kritis=1, Sangat Kritis=0}
MANAJEMEN = Buruk
| PRODUKTIFITAS = Rendah: Kritis {Tidak Kritis=33, Potensial Kritis=8, Agak Kritis=8, Kritis=118, Sangat Kritis=28}
| PRODUKTIFITAS = Sangat Rendah: Kritis {Tidak Kritis=2, Potensial Kritis=0, Agak Kritis=0, Kritis=167, Sangat Kritis=110}

```

```

| PRODUKTIFITAS = Sangat Tinggi: Agak Kritis {Tidak Kritis=0, Potensial Kritis=0, Agak
Kritis=8, Kritis=0, Sangat Kritis=0}
| PRODUKTIFITAS = Sedang: Kritis {Tidak Kritis=0, Potensial Kritis=0, Agak Kritis=25,
Kritis=149, Sangat Kritis=100}
| PRODUKTIFITAS = Tinggi: Agak Kritis {Tidak Kritis=0, Potensial Kritis=0, Agak
Kritis=59, Kritis=11, Sangat Kritis=0}
MANAJEMEN = Sedang
| PRODUKTIFITAS = Rendah: Agak Kritis {Tidak Kritis=0, Potensial Kritis=0, Agak
Kritis=913, Kritis=541, Sangat Kritis=0}
| PRODUKTIFITAS = Sangat Rendah: Kritis {Tidak Kritis=0, Potensial Kritis=0, Agak
Kritis=0, Kritis=34, Sangat Kritis=0}
| PRODUKTIFITAS = Sangat Tinggi: Potensial Kritis {Tidak Kritis=336, Potensial
Kritis=4789, Agak Kritis=935, Kritis=0, Sangat Kritis=0}
| PRODUKTIFITAS = Sedang: Agak Kritis {Tidak Kritis=17, Potensial Kritis=2351, Agak
Kritis=6674, Kritis=700, Sangat Kritis=0}
| PRODUKTIFITAS = Tinggi: Potensial Kritis {Tidak Kritis=1320, Potensial Kritis=2576,
Agak Kritis=1782, Kritis=13, Sangat Kritis=0}

```

#### 4. KESIMPULAN

Berdasarkan hasil penelitian klasifikasi lahan kritis Kawasan budidaya pertanian di daerah aliran sungai pemali jratan dapat disimpulkan bahwa Algoritma C.45 dapat diimplementasikan untuk prediksi lahan kritis pada Kawasan Budidaya Pertanian dengan tingkat akurasi sebesar 92.47% dari 111.003 dataset dengan empat parameter lahan kritis. Hasil penelitian dapat membantu pengelola Daerah Aliran Sungai Pemali Jratan untuk memprediksi lahan kritis pada Kawasan Budidaya Pertanian, sehingga dapat diambil kebijakan yang sesuai.

#### REFERENSI

- [1] A. Susetyaningsih, "Pengaturan Penggunaan Lahan di Daerah Hulu DAS Cimanuk Sebagai Upaya Optimalisasi Pemanfaatan Sumberdaya Air," *Kontruksi*, vol. 10, no. 01, pp. 1–8, 2012.
- [2] R. H. Anasiru, "Analisis Spasial Dalam Klasifikasi Lahan Kritis Di Kawasan Sub-DAS Lange Gorontalo," *Inform. Pertan.*, vol. 25, no. 2, pp. 261–272, 2016.
- [3] A. Huzaini and S. Rahayu, "Tingkat Kekritisan Lahan di Kecamatan Gunngpati Kota Semarang," *J. Tek. PWK*, vol. 2, no. 2, pp. 270–280, 2013.
- [4] M. Rosyada, Y. Prasetyo, and Hani'ah, "Penentuan Tingkat Lahan Kritis Menggunakan Pembobotan dan Algoritma NDVI," *J. Geod. Undip*, vol. 4, no. 1, pp. 85–94, 2015.
- [5] H. Hendro, Z. Nadhi, S. Budiastuti, and D. Purnomo, "Pemetaan Lahan Kritis di Kawasan Muria untuk Meningkatkan Daya Dukung Lingkungan yang Berbasis pada Sistem Informasi Geografis," *J. Ilmu Pertan.*, vol. 17, no. 1, pp. 46–51, 2014.
- [6] M. K. Nasution, "Tingkat Kekritisn dan Rehabilitasi Lahan di DAS Krueng Aceh," Institut Pertanian Bogor, 2018.
- [7] I. M. Parsa, U. S. Wiradisastra, and H. Pawitan, "Identifikasi dan Pemetaan Lahan Kritis Menggunakan Teknik Pengindraan Jauh dan Sistem Informasi Geografi," *Manaj. Hutan Trop.*, vol. IX, no. 2, pp. 63–77, 2003.
- [8] N. Khotimah and D. Istiawan, "Perbandingan Algoritma C4.5, Naive Bayes dan K-Nearest Neighbour untuk Prediksi Lahan Kritis di Kabupaten Pemalang," *Univ. Res. Colloquium*, pp. 41–50, 2018.
- [9] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *Int. J. Remote Sens.*, vol. 28, no. 5, pp. 823–870, 2007.
- [10] J. R. Otukei and T. Blaschke, "Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 125, pp. 527–531, 2010.
- [11] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, Thrid. Wyman Street, Waltham: Morgan Kaufmann, 2012.
- [12] D. Istiawan, "Implementasi Algoritma Naive Bayes untuk Klasifikasi Lahan Kritis pada Kawasan Budidaya Pertanian," *Univ. Res. Colloquium*, vol. 9, pp. 369–372, 2019.
- [13] C. J. Mantas and J. Abellán, "Credal-C4.5: Decision tree based on imprecise probabilities to classify noisy data," *Expert Syst. Appl.*, vol. 41, no. 10, pp. 4625–4637, 2014.
- [14] F. Gorunescu, *Data Mining Concepts, Model and Techniques*, 12th ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.
- [15] B. Hssina, A. Merbouha, H. Ezzikouri, and M. Erritali, "A Comparative Study of Decision Tree ID3 and C4.5," *Int. J. Adv. Comput. Sci. Appl.*, vol. 4, no. 2, 2014.
- [16] M. J. Zaki and W. J. Meira, *Data Mining and Analysis: Fundamental Concepts and Algorithms*. 2014.

- [17] D. M. Farid, L. Zhang, C. M. Rahman, M. A. Hossain, and R. Strachan, "Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1937–1946, 2014.
- [18] J. R. Quinlan, "C4.5: Programs for Machine Learning.pdf," *Morgan Kaufmann*, vol. 5, no. 3. The Morgan Kaufman, United States of America, p. 302, 1993.
- [19] I. Rahmayuni, "Perbandingan Performansi Algoritma C4.5 dan CART Dalam Klasifikasi Data Nilai Mahasiswa Prodi Teknik Komputer Politeknik Negeri Padang," *Teknoif*, vol. 2, no. 1, pp. 40–46, 2014.
- [20] E. Prastyo, *Data Mining Mengolah data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: ANDI, 2014.
- [21] C. A. Floudas and P. M. Pardalos, *Encyclopedia of Optimization*. United States of America: Springer US, 2009.
- [22] L. Torgo, *Data Mining with R: Learning with Case Studies*, vol. 20105341. New York: Chapman & Hall, 2010.

### BIBLIOGRAFI PENULIS



**Deden Istiawan.** Menempuh pendidikan S1 di Institut Sains & Teknologi AKPRIND Yogyakarta, S2 Magister Teknik Informatika di Universitas Dian Nuswantoro. Menjadi Pengajar di Akademi Statistika (AIS) Muhammadiyah Semarang. Minat Penelitian saat ini pada bidang data mining dan machine learning.



Laelatul Khikmah. Memperoleh gelar S.Si dari Prodi Studi Statistika Fakultas MIPA Universitas Islam Indonesia, dan gelar M.Si dari Program Pascasarja Statistika IPB. Dia saat ini sebagai dosen program D3 Statika di Akademi Statistika Muhammadiyah Semarang, Indonesia. Selain itu dia tergabung di Data Science Indonesia Regional Jawa Tengah. Minat Penelitian saat ini di bidang data mining, Regresi Logistik dan Data Kategorik.