

## Perbandingan Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbour untuk Prediksi Lahan Kritis di Kabupaten Pemalang

Nur Khotimah<sup>1\*</sup>, Deden Istiawan<sup>2</sup>

Akademi Statistika Muhammadiyah Semarang

\*Email: nurkhotimah111@gmail.com

---

### Abstrak

**Keywords:**

Lahan Kritis;  
Klasifikasi; C4.5;  
Naïve Bayes; k-NN.

*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. Data dan informasi mengenai degradasi hutan dan lahan seringkali tidak mengacu kepada format dan struktur database yang dapat dipertanggungjawabkan. Sehingga salah satu faktor penting dalam menentukan kesuksesan pemetaan penggunaan lahan guna meningkatkan fungsi hutan lindung dan hutan konservasi terletak pada pemilihan skema klasifikasi lahannya. Klasifikasi dapat diselesaikan dengan menggunakan teknik data mining. Klasifikasi termasuk dalam tipe supervised learning yang artinya dibutuhkan data pelatihan untuk membangun suatu model klasifikasinya. Terdapat lima kategori klasifikasi yaitu berbasis statistik, berbasis jarak, berbasis pohon keputusan, berbasis jaringan saraf, dan berbasis aturan. Tiap kategori klasifikasi memiliki banyak pilihan algoritma. Algoritma yang paling sering digunakan adalah algoritma C4.5, Naïve Bayes, dan k-Nearest Neighbour (k-NN). Penelitian ini akan membandingkan hasil klasifikasi dari tiga algoritma tersebut untuk mengetahui algoritma mana yang memiliki kinerja paling baik dalam memprediksi berdasarkan nilai akurasi. Data yang digunakan yaitu data lahan lahan kritis daerah hutan lindung dan hutan konservasi Kabupaten Pemalang. Hasil dari penelitian ini diperoleh C4.5 memiliki akurasi paling tinggi sebesar 77,75 % disusul Naïve Bayes 77,49% dan terakhir k-NN memiliki akurasi sebesar 73,91%.*

### 1. PENDAHULUAN

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 (Prabandaru, Lailanugraha, & Sukmono, 2016). Lahan kritis merupakan salah satu indikator terjadinya degradasi lingkungan sebagai dampak dari berbagai jenis pemanfaatan sumber daya lahan yang kurang bijaksana (Indrihastuti, 2016) sehingga penggunaan dan kemampuannya mengalami proses kerusakan fisik, kimia, dan biologi yang pada akhirnya membahayakan fungsi hidrologis, orologis, produksi pertanian, pemukiman dan kehidupan sosial ekonomi dari daerah tersebut (Hendro, Nahdi, Budiastuti, & Purnomo, 2015). Namun, tidak tersedianya data spasial dan informasi berpengaruh terhadap penilaian validitas data lahan kritis (Candra, 2011). Oleh karena itu, data dan informasi obyektif kondisi hutan dan lahan sasaran RHL sangat diperlukan terutama dalam menunjang formulasi strategi RHL yang berdaya guna, sehingga diharapkan dapat diperoleh acuan dalam pengalokasian sumber daya secara proporsional (Pemali Jratun, 2013).

Jawa Tengah memiliki hutan lindung seluas 84,430 Ha. Persentase luasan tersebut adalah sebesar 11,15% dari luas kawasan hutan di Jawa Tengah, atau sekitar 2,59% dari luas daratan di Jawa Tengah sedangkan berdasarkan laporan lahan kritis Badan Pengelolaan Daerah Aliran Sungai (BPDAS) Pemali Jratun 2013, Kabupaten Pemalang memiliki daerah kritis untuk hutan lindung dan hutan konservasi terluas di Jawa Tengah yakni sebesar 7.01% atau setara dengan 2871.5Ha. Parameter yang menjadi faktor kekritisannya lahan, yaitu tingkat bahaya erosi, produktivitas lahan, erosi lahan, manajemen lahan, singkapan batuan, kemiringan lereng dan curah hujan (Hendro et al., 2015). Namun data dan informasi mengenai degradasi hutan dan lahan seringkali tidak mengacu kepada format dan struktur *database* yang dapat dipertanggungjawabkan. Sehingga salah satu faktor penting dalam menentukan kesuksesan pemetaan penggunaan lahan terletak pada pemilihan skema klasifikasi lahannya (Pemali Jratun, 2013).

Metode sebelumnya yang pernah digunakan yaitu Sistem Informasi Geografis (SIG) (Hendro et al., 2015). Namun SIG memiliki kelemahan membutuhkan biaya yang mahal untuk menangkap citra satelit yang beresolusi baik, serta untuk mengetahui kondisi curah hujan harus berasal dari minimal tiga stasiun pemantauan (Amri, Halim, & Barchia, 2014). Sehingga untuk memperoleh informasi terpercaya dari data satelit secara efektif diperlukan teknik klasifikasi yang tepat. Sejumlah pendekatan klasifikasi telah dikembangkan selama beberapa dekade terakhir (Lu & Weng, 2007) termasuk juga dengan pendekatan data mining (Otukey & Blaschke, 2010) (Pal & Mather, 2003).

Data mining merupakan suatu proses untuk menemukan informasi dari jumlah data yang besar (Zaki & Meira, 2014). Data mining mempunyai lima peran utama yaitu estimasi, prediksi, klasifikasi, kluster dan asosiasi (P.-N. Tan, Steinbach, & Kumar, 2006). Peran data mining yang sering digunakan adalah klasifikasi dan kluster karena dapat digunakan untuk atribut yang banyak (Fan, Wallace, Rich, & Zhang, 2006). Kluster merupakan proses pengelompokan data ke dalam beberapa kelompok sehingga data dalam kelompok tersebut memiliki tingkat kemiripan karakteristik antara data yang satu dengan yang lainnya (Hosseini, Maleki, & Gholamian, 2010). Sedangkan klasifikasi adalah bentuk analisis data untuk mengekstrak model yang akan digunakan untuk memprediksi label kelas (Han, Kamber, & Pei, 2012).

Klasifikasi adalah salah satu bagian dari data mining yang dapat digunakan untuk menggambarkan dan membedakan kelas data (Farid, Zhang, Rahman, Hossain, & Strachan, 2014). Klasifikasi termasuk dalam tipe supervised learning yang artinya dibutuhkan data pelatihan untuk membangun suatu model klasifikasinya. Terdapat lima kategori klasifikasi yaitu berbasis statistik, berbasis jarak, berbasis pohon keputusan, berbasis jaringan saraf, dan berbasis aturan. Tiap kategori klasifikasi memiliki banyak pilihan algoritma. Algoritma yang paling sering digunakan adalah algoritma C4.5, Naïve Bayes, dan k-Nearest Neighbor (k-NN) (Wu et al., 2008).

Algoritma C4.5 adalah suatu cara klasifikasi yang mudah digunakan dan diinterpretasikan (Abellán, 2013) (Mantas & Abellán, 2014). C4.5 memiliki beberapa kelebihan antara lain, mudah dimengerti, mudah diimplementasikan, membutuhkan sedikit waktu, mampu menangani data numerik dan kategorik dan dapat mengolah *dataset* yang besar dan rumit (Farid et al., 2014). Namun, algoritma C4.5 bias terhadap distribusi kecil (Mulholland, 1995).

Algoritma k-NN merupakan sebuah metode untuk melakukan klasifikasi terhadap objek baru berdasarkan  $k$  tetangga terdekatnya dan kelas yang paling banyak muncul yang akan menjadi kelas hasil klasifikasi (Witten, Frank, & Hall, 2011). Kelebihan dari k-NN adalah dapat digunakan untuk memecahkan permasalahan *multiclass* (Aburomman, Bin, & Reaz, 2015) namun, k-NN memiliki masalah untuk menemukan tetangga terdekat pada titik *query* dari *dataset* yang digunakan (Liaw, Wu, & Leou, 2010).

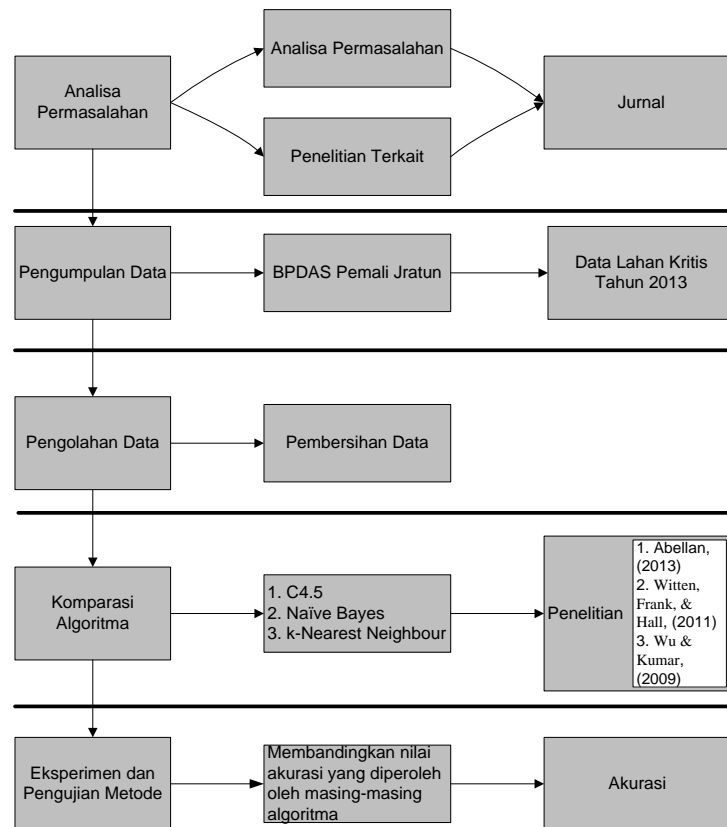
*Naïve Bayes* adalah algoritma klasifikasi probabilitas sederhana yang berdasarkan pada teorema Bayes, asumsi bebas yang kuat (Farid & Rahman, 2010). *Naïve Bayes* juga

merupakan algoritma klasifikasi yang utama pada *data mining* dan banyak diterapkan dalam masalah klasifikasi di kehidupan sehari-hari karena memiliki performa klasifikasi yang tinggi. Algoritma *Naïve Bayes* juga memiliki beberapa keunggulan seperti mudah serta biaya perhitungan kecil (Wu & Kumar, 2009), dapat menangani data missing (Kang, Yoo, & Han, 2012), memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam *database* dengan data yang besar (Escalante, Morales, & Sucar, 2016). Namun algoritma *Naïve Bayes* memiliki asumsi independensi atribut (Zhang, Liu, & Almpandis, 2017).

Penelitian yang sudah dilakukan masing-masing algoritma memiliki kekurangan dan kelebihan masing-masing oleh karena itu penulis akan melakukan perbandingan terhadap algoritma *C4.5*, *Naïve bayes* dan *k-Nearest Neighbor*, untuk prediksi lahan kritis pada kawasan hutan lindung dan hutan konservasi Kabupaten Pematang Jaya.

## 2. METODE

Metode penelitian yang digunakan pada penelitian ini adalah metode eksperimen. Metode eksperimen yang dimaksud adalah uji coba yang dikontrol atau dikendalikan oleh peneliti sendiri untuk melakukan peninjauan hubungan kausal (hubungan sebab akibat). Tahapan penelitian dapat dilihat pada Gambar 1 yang dapat dijelaskan sebagai berikut:



Gambar 1. Tahapan Penelitian

### 2.1 Analisa Permasalahan dan Tinjauan Pustaka

Penelitian ini diawali dengan mengumpulkan jurnal yang berhubungan dengan metode klasifikasi dengan algoritma *C4.5*, *Naïve Bayes* dan *k-Nearest Neighbour*. Selanjutnya mencari referensi metode penelitian terdahulu dengan topik prediksi lahan kritis.

### 2.2 Pengumpulan Data

*Dataset* yang digunakan dalam penelitian ini adalah menggunakan *dataset* parameter lahan kritis. Parameter lahan kritis yang digunakan dalam penelitian ini adalah data

tingkat bahaya erosi, kemiringan lereng dan produktifitas yang didapatkan dari Balai Pengelolaan Daerah Aliran Sungai (BPDAS) Pemali Jratun. Teknik dalam pengambilan data DAS itu sendiri menggunakan Peta Citra Satelit Landsat 7 tahun 2011 dan 8 tahun 2013. Interpretasi citra landsat tentang kondisi biofisik wilayah yang memberikan indikator terhadap bentuk dan tingkat kekritisian wilayah. Hasil interpretasi peta citra adalah berupa peta penutupan lahan sebagai *base map* penyusunan lahan kritis. Adapun kelompok kekritisian lahan menurut pengelompokan dengan metode K-modes yaitu kelompok satu yang menjadi prioritas ketiga dengan karakteristik penutupan lahan yang sangat baik, kemiringan lereng yang datar, tingkat bahaya erosi yang berat dan manajemen yang buruk. Selanjutnya kelompok dua dengan karakteristik penutupan lahan yang sedang, kemiringan lereng yang landai, tingkat bahaya erosi yang sangat berat, dan manajemen yang buruk dan kelompok 3 yang menjadi prioritas kedua dengan karakteristik penutupan lahan yang baik, kemiringan lereng yang landai, tingkat bahaya erosi yang sangat berat, dan manajemen yang buruk (Yulianita & Istiawan, 2017). Kriteria ini didasarkan pada parameter yang terdiri dari, kondisi tutupan vegetasi, kemiringan lereng, tingkat erosi, manajemen, dan produktivitas lahan.

#### 1. **Tingkat Bahaya Erosi**

Tingkat bahaya erosi dapat dihitung dengan cara membandingkan tingkat erosi di suatu satuan lahan dan kedalaman tanah efektif pada satuan lahan. Dalam hal ini tingkat erosi dihitung dengan menghitung perkiraan rata-rata tanah hilang tahunan akibat erosi lapis dan alur yang dihitung dengan rumus *Universal Soil Loss Equation* (USLE) (Hendro et al., 2015).

#### 2. **Kemiringan Lereng**

Kemiringan lereng adalah perbandingan antara jarak vertikal suatu lahan beberapa satuan, diantaranya adalah dengan persen dan derajat. Data spasial kemiringan lereng dapat disusun dari hasil pengolahan data garis kontur dengan bersumber pada peta topografi atau peta rupa bumi (Sunartomo, 2011).

#### 3. **Manajemen Lahan**

Manajemen lahan merupakan salah satu kriteria yang dipergunakan untuk menilai lahan kritis di kawasan hutan lindung, yang dinilai berdasarkan kelengkapan aspek pengelolaan yang meliputi keberadaan tata batas kawasan, pengamanan dan pengawasan serta dilaksanakan atau tidaknya penyuluhan (Kurnia, Sutrisno, & Sungkana, 2007).

#### 4. **Produktivitas**

Produktivitas merupakan salah satu kriteria yang dipergunakan untuk menilai kekritisian lahan di kawasan budidaya pertanian, yang dinilai berdasarkan rasio terhadap produksi komoditi umum optimal pada pengelolaan tradisional. Sehingga terdapat deliniasi wilayah kerja sesuai dengan tingkat kekritisian lahan dengan memperhatikan fungsi kawasan yaitu kawasan hutan lindung, kawasan budidaya pertanian, kawasan lindung di luar kawasan hutan, kawasan hutan konservasi dan kawasan hutan produksi (Wibowo & Gintings, 2010) (Ramayanti, Yuwono, & Awaluddin, 2015).

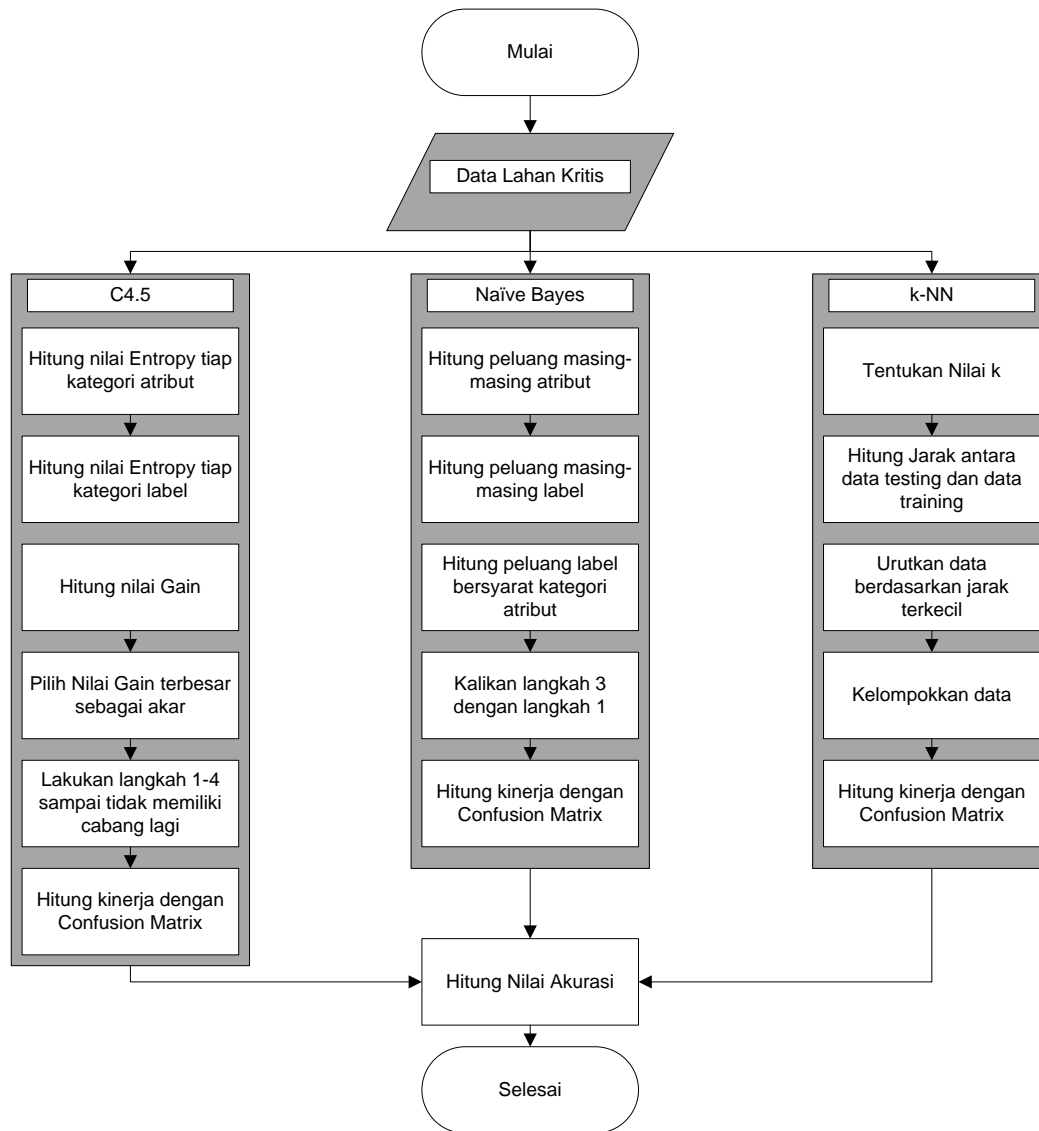
### 2.3 **Pengolahan Data**

Pada tahap pengolahan data, masing-masing dataset yang masih berisi data informasi pendukung perlu dilakukan pengolahan dengan membersihkan data informasi pendukung tersebut. Data yang diolah hanya tiga parameter karena berdasarkan data yang ada untuk atribut manajemen lahan memiliki nilai yang sama dan tidak memengaruhi nilai klasifikasi. Data penelitian ini bertipe kategorik serta kabupaten untuk dijadikan objek, dalam penelitian ini adalah Kabupaten Pematang.

### 2.4 **Komparasi Algoritma Klasifikasi**

Klasifikasi termasuk dalam tipe *supervised learning* yang artinya dibutuhkan data pelatihan untuk membangun suatu model klasifikasinya. Terdapat lima kategori

klasifikasi yaitu berbasis statistik, berbasis jarak, berbasis pohon keputusan, berbasis jaringan saraf, dan berbasis aturan. Namun untuk penelitian ini hanya membandingkan algoritma C4.5, *Naïve Bayes*, dan *k-Nearest Neighbor* (k-NN). Berikut tahapan dalam perhitungan penelitian ini:



Gambar 2 Langkah Perhitungan

#### 2.4.1 C4.5

Algoritma *C4.5* merupakan algoritma yang umum digunakan untuk pengambilan keputusan. *C4.5* akan mencari solusi permasalahan dengan menjadikan kriteria sebagai *node* yang saling berhubungan membentuk seperti struktur pohon (Mantas & Abean, 2014). *C4.5* adalah model prediksi terhadap suatu keputusan menggunakan struktur hirarki atau pohon. Setiap pohon memiliki cabang, cabang mewakili suatu atribut yang harus dipenuhi untuk menuju cabang selanjutnya hingga berakhir di daun (tidak ada cabang lagi). Konsep data dalam *C4.5* adalah data dinyatakan dalam bentuk tabel yang terdiri dari atribut dan *record*. Atribut digunakan sebagai parameter yang dibuat sebagai kriteria dalam pembuatan pohon, untuk pohon pertama adalah nilai *Gain*

tertinggi dan berulang sampai tidak ada cabang lagi. Persamaan yang digunakan untuk menghitung *Gain* seperti berikut:

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad 2.1$$

#### 2.4.2 Naïve Bayes

*Naïve Bayes* adalah algoritma yang digunakan dalam statistika untuk menghitung peluang dari suatu hipotesis, *Naïve Bayes* menghitung peluang suatu label berdasarkan pada atribut yang dimiliki dan menentukan label yang memiliki peluang paling tinggi (Lu, Ling, & Huang, 2003). *Naïve Bayes* mengklasifikasikan label berdasarkan pada probabilitas sederhana dengan mengasumsikan bahwa setiap atribut dalam data tersebut bersifat saling terpisah. *Naïve Bayes* merupakan salah satu metode yang banyak digunakan berdasarkan beberapa sifatnya yang sederhana. *Naïve Bayes* mengklasifikasikan data berdasarkan peluang  $P$  atribut  $X$  dari setiap label  $Y$  data. Model peluang setiap label dan jumlah atribut yang dapat dituliskan seperti persamaan berikut.

$$P(C_i|X_1, X_2, \dots, X_n) \quad 2.2$$

Penghitungan *Naïve Bayes* dapat dijelaskan dengan  $C_i$  adalah hipotesis data  $X$  yang merupakan suatu label.  $P(C_i|X)$  adalah peluang hipotesis label berdasarkan kategori  $X$  (*posteriori probability*).  $P(C_i)$  adalah peluang hipotesis label (*prior probability*).  $P(X|C_i)$  adalah peluang data  $X$  berdasarkan kategori pada hipotesis label.  $P(X)$  adalah probabilitas data  $X$ . Sehingga didapatkan rumus penghitungan *Naïve Bayes* dituliskan pada persamaan.

$$P(C_i|X) = \frac{P(X|C_i) P(C_i)}{P(X)} \quad 2.3$$

#### 2.4.3 k-Nearest Neighbour

*k-Nearest Neighbour* adalah algoritma pengklasifikasian yang didasarkan pada analogi, yaitu membandingkan data uji dengan data pelatihan yang berada dekat dengan dan memiliki kemiripan dengan data uji tersebut (S. Tan, 2006). Kemiripan data uji dengan data pelatihan didasarkan pada jaraknya. Banyak persamaan yang dapat digunakan untuk menghitung jarak antara data uji dan data pelatihan. Perhitungan selanjutnya adalah persamaan untuk mencari kemiripan dengan *k-nearest neighbour* yaitu:

$$Similarity(T,S) = \frac{\sum_{i=1}^n Sim(K_i(T), K_i(S)) w_i}{\sum_{i=1}^n w_i} \quad 2.4$$

#### 2.5 Eksperimen Pengujian

Pada tahap ini menjelaskan mengenai langkah-langkah eksperimen serta proses dari perbandingan klasifikasi dengan tiga algoritma. Proses analisis diawali dengan menghitung klasifikasi tiap algoritma dan mencari nilai evaluasi dengan menggunakan nilai akurasi. Setelah didapatkan nilai akurasi maka dibandingkan dengan melihat akurasi mana yang paling tinggi. Program yang dipakai untuk analisis ini yaitu *Rapidminer*. digunakan program *Rapidminer*.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Data Uji

Penelitian ini menggunakan *dataset* parameter kekritisian hutan lindung dan hutan konservasi Kabupaten Pemalang tahun 2013. *Dataset* ini merupakan data laporan terakhir yang dilakukan oleh BPDAS Pemali Jratun. *Dataset* ini berjumlah 391 data dan memiliki tiga atribut serta satu label. Tiga atribut tersebut adalah parameter kekritisian yang terdiri dari serta satu label merupakan kategori kekritisian suatu lahan.

Tabel 1. Keterangan Atribut dan Label

No	Indikator	Jenis Data	Keterangan	Jumlah Kategori	Kategori
1	Tingkat Bahaya Erosi	Ordinal	Atribut	4	1. Sangat Berat 2. Berat 3. Sedang 4. Ringan
2	Kemiringan Lereng	Ordinal	Atribut	5	1. Sangat Curam 2. Curam 3. Agak Curam 4. Landai 5. Datar
3	Produktifitas	Ordinal	Atribut	5	1. Sangat Buruk 2. Buruk 3. Sedang 4. Baik 5. Sangat baik
4	Kekritisian	Ordinal	Label	2	1. Kritis 2. Tidak Kritis

#### 3.2 Pengujian Algoritma

Pengujian yang dilakukan pada penelitian ini adalah pengujian dengan *use training dan cross-validation*. *Use training* yaitu mengevaluasi seberapa baik algoritma mampu memprediksi label dari instance setelah dilakukan pelatihan. Data pelatihan akan digunakan untuk data uji. Sedangkan *Cross-validation* mengevaluasi algoritma menggunakan nilai *folds* yang dimasukkan. Pada penelitian ini nilai *folds* adalah 10. Evaluasi pengujian yang dilakukan terhadap algoritma klasifikasi *Naïve Bayes*, *k-Nearest Neighbour*, dan C4.5 dalam penelitian ini menggunakan tingkat akurasi yang dihasilkan algoritma tersebut dalam memprediksi kekritisian lahan kawasan hutan lindung dan hutan konservasi Perhitungan akurasi dapat dilihat ditabel 2.

Tabel 2. Perhitungan Akurasi

Label	Kenyataan Benar	Kenyataan Salah
Prediksi Benar	TP (True Positive)	FP (False Positive)
Prediksi Salah	FN (False Negative)	TN (True Negative)

$$\text{Akurasi} = \frac{TP+TN}{TP+FN+FP+TN} \times 100 \quad 3.1$$

### 3.3 Eksperimen Pengujian

Pada pengujian ini semua *dataset* sejumlah 391 data akan dijadikan data pelatihan dan data pengujian. Hasil dari mode pengujian ini akan mencapai nilai tingkat akurasi yang besar hingga mencapai 100%, dikarenakan data pelatihan dan data uji yang digunakan adalah data yang sama. Hasil dari mode pengujian ini dapat dilihat pada tabel berikut:

Tabel 3. Hasil Pengujian *Use Training* dan *Cross-validation*

No	Algoritma	Prediksi Benar	Prediksi Salah	Classification Accuracy%
1	C4.5	304	87	77.75
2	Naïve Bayes	303	88	77.49
3	k-NN	289	102	73.91

## 4. KESIMPULAN

Kesimpulan yang diperoleh pada penelitian ini adalah dari hasil perbandingan algoritma klasifikasi *k-Nearest Neighbour*, *Naïve Bayes* dan C4.5 yang digunakan pada studi kasus prediksi lahan kritis kawasan hutan lindung dan hutan konservasi Kabupaten Pemalang menyatakan bahwa algoritma klasifikasi C4.5 merupakan algoritma klasifikasi yang memiliki tingkat akurasi paling tinggi dalam memprediksi kekritisian lahan kritis daerah kawasan hutan lindung dan hutan konservasi Kabupaten Pemalang dibandingkan algoritma klasifikasi *Naïve Bayes* dan *k-Nearest Neighbour* yaitu mencapai 77.75% disusul *Naïve Bayes* 77,49% dan terakhir *k-NN* memiliki akurasi sebesar 73,91%.

## REFERENSI

- Abellán, J. (2013). Ensembles of decision trees based on imprecise probabilities and uncertainty measures. *Information Fusion*. <https://doi.org/10.1016/j.inffus.2012.03.003>
- Aburomman, A. A., Bin, M., & Reaz, I. (2015). A novel SVM-kNN-PSO ensemble method for intrusion detection system. *Applied Soft Computing Journal*, 1–13. <https://doi.org/10.1016/j.asoc.2015.10.011>
- Candra, D. S. (2011). Analysis Of Critical Land In The Musi Watershed. *International Journal of Remote Sensing and Earth Sciences*, 8, 13–18.
- Escalante, H. J., Morales, E. F., & Sucar, L. E. (2016). A naive Bayes baseline for early gesture recognition. *Pattern Recognition Letters*. <https://doi.org/10.1016/j.patrec.2016.01.013>
- Fan, W., Wallace, L., Rich, S., & Zhang, Z. (2006). Tapping the power of text mining. *Communications of the ACM*, 49(9), 76–82. <https://doi.org/10.1145/1151030.1151032>
- Farid, D. M., & Rahman, M. Z. (2010). Anomaly Network Intrusion Detection Based on Improved Self Adaptive Bayesian Algorithm. *Journal Of Computer*, 5(1), 23–31. <https://doi.org/10.4304/jcp.5.1.23-31>
- Farid, D. M., Zhang, L., Rahman, C. M., Hossain, M. A., & Strachan, R. (2014). Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks. *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2013.08.089>



- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. San Francisco, CA, itd: Morgan Kaufmann. <https://doi.org/10.1016/B978-0-12-381479-1.00001-0>
- Hendro, H., Nahdi, Z., Budiastuti, S., & Purnomo, D. (2015). Pemetaan Parameter 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*, 3, 41–46.
- Hosseini, S. M. S., Maleki, A., & Gholamian, M. R. (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. *Expert Systems with Applications*, 37(7), 5259–5264. <https://doi.org/10.1016/j.eswa.2009.12.070>
- Indrihastuti, D. (2016). Analisis lahan kritis dan arahan rehabilitasi lahan dalam pengembangan wilayah kabupaten kendal jawa tengah. *Repository Institut Pertanian Bogor*.
- Kang, H., Yoo, S. J., & Han, D. (2012). Expert Systems with Applications Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. *Expert Systems With Applications*, 39(5), 6000–6010. <https://doi.org/10.1016/j.eswa.2011.11.107>
- Kurnia, U., Sutrisno, N., & Sungkana, I. (2007). *Perkembangan lahan kritis*. Lembaga Penelitian Indonesia.
- Liaw, Y., Wu, C., & Leou, M. (2010). Fast k -nearest neighbors search using modified principal axis search tree. *Digital Signal Processing*, 20(5), 1494–1501. <https://doi.org/10.1016/j.dsp.2010.01.009>
- Lu, J., Ling, C. X., & Huang, J. (2003). Comparing Naive Bayes , Decision Trees , and SVM with AUC and Accuracy. *The Third IEEE International Conference on DataMining*, 11–14.
- Mantas, C. J., & Abean, J. (2014). Credal-C4.5: Decision tree based on imprecise probabilities to classify noisy data. *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2014.01.017>
- Mulholland, M. (1995). Application of the C4 . 5 classifier to building an expert system for ion chromatography, 27, 95–104.
- Pemali Jratun, B. (2013). *Laporan Penyusunan Review Data dan Peta Lahan Kritis Wilayah Kerja BPDAS Pemali Jratun 2013*. BPDAS Pemali Jratun.
- Prabandaru, L. H., Lailanugraha, A., & Sukmono, A. (2016). Pemetaan Tingkat Lahan Kritis Kabupaten Wonosobo Dengan Penginderaan Jauh Dan Sistem Informasi Geografis. *Jurnal Geodesi Undip Oktober 2016*, 5, 65–72.
- 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 April 2015*, 4(April), 200–207.
- Sunartomo, A. F. (2011). Inventarisasi dan Sebaran Lahan Kritis di Kabupaten Situbondo. *J-SEP*, 5(1), 12–22.

- Tan, P.-N., Steinbach, M., & Kumar, V. (2006). Introduction to Data Mining. *Journal of School Psychology, 19*(1), 51–56. [https://doi.org/10.1016/0022-4405\(81\)90007-8](https://doi.org/10.1016/0022-4405(81)90007-8)
- Tan, S. (2006). An effective refinement strategy for KNN text classifier, *30*, 290–298. <https://doi.org/10.1016/j.eswa.2005.07.019>
- Wibowo, A., & Gintings, A. N. (2010). Degradasi dan Upaya Pelestarian Hutan. *Membalik Kecenderungan Degradasi Sumber Daya Lahan Dan Air*, 67–87. Retrieved from <http://www.litbang.pertanian.go.id/buku/membalik-kecenderungan-degrad/BAB-III-3.pdf>
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining. [Morgan Kaufmann series in data management systems]*. [https://doi.org/10.1002/1521-3773\(20010316\)40:6<9823::AID-ANIE9823>3.3.CO;2-C](https://doi.org/10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C)
- Wu, X., & Kumar, V. (2009). *The Top Ten Algorithm in Data Mining*. Taylor & Francis Group, LLC.
- Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., ... Dan, J. H. (2008). *Top 10 algorithms in data mining*. <https://doi.org/10.1007/s10115-007-0114-2>
- Yulianita, T., & Istiawan, D. (2017). Implementasi Algoritma K-modes untuk Penentuan Prioritas Rehabilitasi Daerah Aliran Sungai Berdasarkan Parameter Lahan Kritis. In *The 6th University Research Colloquium 2017* (pp. 429–440).
- Zaki, M. J., & Meira, W. J. (2014). *Data Mining and Analysis: Fundamental Concepts and Algorithms*, 562. Retrieved from <https://books.google.com/books?hl=en&lr=&id=Gh9GAwAAQBAJ&pgis=1>
- Zhang, C., Liu, C., Zhang, X., & Alpanidis, G. (2017). An Up-to-Date Comparison of State-of-the-Art Classification Algorithms. *Expert Systems With Applications*. <https://doi.org/10.1016/j.eswa.2017.04.003>