

Penerapan Algoritma *Gaussian Naive Bayes* Dalam Penentuan Prioritas Rehabilitasi Daerah Aliran Sungai Berdasarkan Parameter Lahan Kritis

Tiara Destiana¹, Yuyun Umidah², Ultach Enri³

^{1,2}Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang

²Program Studi, Fakultas, Universitas

Email: tiara.destiana19237@student.unsika.ac.id¹

ABSTRACT

The quality of watersheds (DAS) is declining due to reduced water availability, erosion of fertile soil layers, and landslides. The Pemali Jratun watershed has a large critical area of 559,492,530 hectares. In this work, a classification approach can be used in data mining with the *Gaussian Naive Bayes* algorithm to identify areas that are considered important land. Five categories—Very Critical, Critical, Slightly Critical, Potentially Critical, and Not Critical—will be used to categorize the research. Three parts make up the scenario distribution of the dataset: the 70:30, 80:20 and 90:10 data. Three scenarios made with a ratio of 70:30 produce more accurate test results. Accuracy of 71%, Recall 0.71, and Precision 0.56, resulting in an F1-Score of 0.61. The final results of the classification show that there are two important classes of vital land: class 1 has a relatively large amount of potentially crucial land, and class 2 has a relatively significant amount of critical land.

Keywords: Classification, *Gaussian Naive Bayes*, DAS, Critical Land

ABSTRAK

Berkurangnya sumber air, lapisan tanah yang subur mengalami erosi, longsor, dimana hal ini berdampak pada perubahan lahan kritis, yang menyebabkan penurunan kualitas Daerah Aliran Sungai (DAS). DAS Pemali Jratun memiliki lahan kritis seluas 559.492.530 hektar. Pendekatan klasifikasi dapat digunakan untuk mengidentifikasi wilayah yang memiliki lahan kritis dengan menggunakan algoritma *Gaussian Naive Bayes* selama proses data mining, dengan metodologi yang digunakan dalam penelitian ini adalah KDD. Skenario pembagian dataset terbagi menjadi 3 yaitu data 70:30, 80:20, dan 90:10, penelitian ini akan dibagi menjadi 5 klasifikasi, yaitu Sangat Kritis, Kritis, Agak Kritis, Kritis Potensial, dan Tidak Kritis. Hasil pengujian dari 3 skenario yang dibuat, pemodelan dari rasio 70:30, memiliki akurasi yang unggul. Nilai F1-Score 0,61, Precision 0,56, Recall 0,71, dan Accuracy 71%. Berdasarkan kesimpulan akhir klasifikasi, terdapat dua kelas lahan kritis penting yaitu kelas 1 dengan tingkat kekritisan lahan berpotensi kritis dan kelas 2 dengan tingkat kekritisan lahan agak kritis.

Kata Kunci: Klasifikasi, *Gaussian Naive Bayes*, DAS, Lahan Kritis

Riwayat Artikel :

Tanggal diterima : 17-08-2023

Tanggal revisi : 21-08-2023

Tanggal terbit : 22-08-2023

DOI :

<https://doi.org/10.31949/infotech.v9i2.6501>

INFOTECH journal by Informatika UNMA is licensed under CC BY-SA 4.0

Copyright © 2023 By Author



1. PENDAHULUAN

1.1. Latar Belakang

Sumberdaya hutan memberikan segudang manfaat yang cukup baik dari sisi sosial maupun ekonomi. Tetapi, satu sisi manfaat ekologis kerap terabaikan yang menyebabkan ekosistem hutan menjadi tidak seimbang yang akan berdampak buruk terhadap kerusakan di sekitar Daerah Aliran Sungai (DAS) (Kristiano et al., 2021). Kerusakan yang terjadi membuat penurunan kualitas Daerah Aliran Sungai (DAS), penurunan tersebut disebabkan dengan adanya alih fungsi hutan yang salah mengakibatkan rendahnya jumlah hutan pada lokasi tersebut, menurunnya sumber mata air, lapisan tanah yang subur mengalami erosi, terjadinya tanah longsor, serta terjadi pendangkalan sungai hal tersebut berdampak perubahan ke arah lahan kritis (Yulianita & Istiawan, 2017).

Menurut Kementerian Kehutanan pada tahun 2006 berdasarkan kriteria sangat kritis dan kritis memperlihatkan bahwa luas lahan kritis yang dimiliki adalah seluas 30,19 juta hektar, kegiatan Rehabilitasi Hutan dan Lahan (RHL) yang berhasil mampu membuat penurunan terus terjadi. Tahun 2011 lahan kritis kembali turun menjadi 27,29 juta hektar, dan turun kembali ditahun 2013 menjadi 24,30 juta hektar. Mendapatkan hasil dari kegiatan Rehabilitasi Hutan dan Lahan (RHL) yaitu terjadinya penurunan jumlah lahan kritis, tetapi hal tersebut belum mampu mengurangi jumlahnya karena jumlah lahan kritis masih cukup luas yaitu 14,006 juta hektar pada tahun 2018.

Berdasarkan Badan Pusat Statistika (BPS) Jawa Tengah (2021) pada tahun 2020 mencatat bahwa luas lahan di Jawa Tengah mencapai 3,2 juta hektare (ha). Terdapat 1,49 juta atau 46% lahan dengan kondisi normal sedangkan sekitar 54% lahan dengan kondisi sangat kritis, kritis, agak kritis dan potensial kritis. Provinsi Jawa Tengah sendiri memiliki dua Balai Pengelolaan Daerah Aliran Sungai (BPDAS) yaitu BPDAS Pemali Jratun dan Solo.

Pada Daerah Aliran Sungai (DAS) Pemali Jratun memiliki 10 Satuan Wilayah Pengelolaan (SWP) terdiri dari 133 DAS dan 69 Sub DAS. DAS sendiri memiliki salah satu fungsinya sebagai pengendali sedimentasi waduk, penyedia air, dan pengendali banjir, tetapi pada beberapa desa di kabupaten semarang di tahun 2020 mengalami banjir serta longsor seperti pada Desa Ngrapah, Rowoboni dan Banyubiru.

Perlu diketahui bahwa lahan kritis yang dimiliki DAS Pemali Jratun seluas 559 492,530 ha. Oleh karena itu program Rehabilitasi Hutan dan Lahan diperlukan guna melindungi daya dukung dan memperbaiki kerusakan yang terjadi pada DAS sehingga fungsi DAS tersebut tetap terjaga sampai dimasa yang akan datang menurut Kementerian Kehutanan.

Untuk menentukan wilayah yang memiliki lahan kritis diperlukan cara yang efektif salah satunya

adalah menggunakan metode klasifikasi yang telah dikembangkan beberapa dekade terakhir, termasuk dengan pendekatan *Data mining*. Menurut Purwanto (Purwanto et al., 2018) mengatakan bahwa proses menemukan korelasi atau pola hubungan dari ratusan bahkan ribuan *field* pada sebuah *database* secara teknis disebut dengan *Data mining*.

Salah satu dari bagian *Data mining* adalah klasifikasi tujuannya adalah memisahkan atribut ke dalam kelas atau kategori yang sesuai untuk menentukan model yang tepat dimana model tersebut akan digunakan dalam proses pengklasifikasian pada atribut yang kelasnya belum diketahui. Terdapat beberapa algoritma yang dapat digunakan dalam metode klasifikasi, salah satunya adalah variasi dari algoritma *Naive Bayes* yaitu *Gaussian Naive Bayes*.

Berdasarkan latar belakang di atas, akan dilakukan penelitian untuk penentuan prioritas rehabilitasi daerah aliran sungai dengan menerapkan metodologi *Data mining* dan algoritma *gaussian naive bayes*.

1.2. Tinjauan Pustaka

1. Daerah Aliran Sungai (DAS)

Tanah, air, tumbuh-tumbuhan dan sumber daya manusia sebagai pelaku pemanfaat sumber daya alam tersebut adalah komponen utama sebuah unit ekosistem dari Daerah Aliran Sungai (DAS) (Ariyani et al., 2020). Menurut Peraturan Menteri Lingkungan Hidup dan Kehutanan Republik Indonesia Nomor 10 Tahun 2022.

DAS singkatan dari Daerah Aliran Sungai mengacu pada terhubungnya wilayah daratan dengan sungai dan anak sungainya yang bermanfaat sebagai penampung, penyimpan serta mengalirkan air yang berasal dari curah hujan ke danau atau ke laut dengan batas darat pemisah topografis dan batas laut sampai daerah perairan yang terpengaruh aktivitas daratan. DAS terdiri dari beberapa daerah didalamnya, yaitu DAS Hulu, DAS Tengah, DAS Hilir.

2. Parameter Lahan Kritis

Terdapat parameter penentu lahan kritis dalam kawasan hutan lindung yang meliputi beberapa ketentuan menurut Permenhut Nomor P.32/Menhut-II/2009 :

a. Penutupan Lahan

Aspek utama yang mempengaruhi betapa vitalnya suatu lahan adalah penutupan lahan, sehingga setiap perubahan pada tutupan lahan harus mempertimbangkan kemampuannya agar lahan kritis dapat terkendali. Penutupan lahan diklasifikasikan menjadi lima kelas dengan nilai parameternya berdasarkan prosentase penutupan tajuk. Klasifikasi penutupan lahan untuk masing-masing kelas ditunjukkan pada **Tabel 2.2**

Tabel 2.1 Klasifikasi dan Skoring Penutupan Lahan Untuk Penentuan Lahan Kritis

| Kelas (Bobot 50) | Prosentase Penutupan Tajuk (%) | Skor |
|------------------|--------------------------------|------|
| Sangat Baik | >80 | 5 |
| Baik | 61-80 | 4 |
| Sedang | 41-60 | 3 |
| Buruk | 21-40 | 2 |
| Sangat Buruk | >20 | 1 |

(Sumber : Dirjen Pengendalian DASHL, 2018)

b. Kemiringan Lereng

Perbandingan ketinggian (jarak vertikal) antara suatu lahan dengan jarak horizontalnya disebut sebagai kemiringan lereng. Dalam persen dan derajat, masing-masing adalah dua cara untuk menyatakan ukuran satuan lereng. Lahan tidak akan mengalami kekritisian jika kerataan lereng meningkat. Klasifikasi Kemiringan Lereng untuk masing-masing kelas ditunjukkan pada **Tabel 2.3**

Tabel 2.2 Klasifikasi dan Skoring Kemiringan Lereng Untuk Penentuan Lahan Kritis

| Kelas (Bobot 20) | Kemiringan Lereng (%) | Skor |
|------------------|-----------------------|------|
| Datar | <8 | 5 |
| Landai | 8-15 | 4 |
| Agak Curam | 16-25 | 3 |
| Curam | 26-40 | 2 |
| Sangat Curam | >40 | 1 |

(Sumber : Dirjen Pengendalian DASHL, 2018)

c. Tingkat Erosi

Dengan membandingkan jumlah erosi pada suatu lahan dan kedalaman tanah efektif di unit lahan tersebut, seseorang dapat menentukan Tingkat Bahaya Erosi (TBE). Jika risiko erosi berkurang maka lahan tidak akan mengalami kekritisian. Klasifikasi Tingkat Erosi untuk masing-masing kelas ditunjukkan pada **Tabel 2.4**

Tabel 2.3 Klasifikasi dan Skoring Tingkat Bahaya Erosi Untuk Penentuan Lahan Kritis

| Kelas (Bobot 20) | Jumlah Erosi (ton/ha/tahun) | Skor |
|------------------|-----------------------------|------|
| Sangat Ringan | <15 | 0 |
| Ringan | 15-60 | 1 |
| Sedang | 61-180 | 2 |
| Berat | 181-480 | 3 |
| Sangat Berat | >480 | 4 |

(Sumber : Dirjen Pengendalian DASHL, 2018)

d. Manajemen

Salah satu faktor yang digunakan untuk mengevaluasi lahan krusial di kawasan hutan lindung adalah manajemen. Pengelolaan dievaluasi berdasarkan ketelitian komponen-komponennya, yang meliputi adanya batas wilayah, pengamanan dan pengawasan, serta dilakukan atau tidaknya penyuluhan. Semakin lahan tidak mengalami kekritisian maka dapat dikatakan semakin baik pengelolaannya. Klasifikasi Manajemen untuk masing-masing kelas ditunjukkan pada **Tabel 2.5**

Tabel 2.4 Klasifikasi dan Skoring Manajemen Untuk Penentuan Lahan Kritis

| Kelas (Bobot 10) | Prosentase Penutupan Tajuk (%) | Skor |
|------------------|--------------------------------|------|
| Baik | Lengkap | 5 |
| Sedang | Tidak Lengkap | 3 |
| Buruk | Tidak Ada | 1 |

(Sumber : Dirjen Pengendalian DASHL, 2018)

Selanjutnya adalah mengelompokan tingkat kekritisian lahan sesuai dengan total skor pada fungsi kawasan. Penentuan tingkat kekritisian lahan dapat dilakukan dengan cara manual yaitu dengan cara penjumlahan nilai yang diperoleh pada masing-masing lahan, yaitu:

$$\begin{aligned}
 & \text{Tingkat Kekritisian Lahan} & (2.1) \\
 & = \sum_{i=1}^n (\% \text{Bobot} \\
 & \quad \times \text{Skor})
 \end{aligned}$$

Untuk penilaiannya, masing-masing kelas diberi bobot, besaran/deskripsi serta skor. Tingkat kekritisian lahan dihitung dengan mengalikan persentase bobot dan skor sehingga diperoleh suatu jumlah total skor masing-masing kawasan. Hasil dari skoring selanjutnya diklasifikasikan seperti pada **Tabel 2.6**.

Tabel 2.5 Klasifikasi Kriteria Berdasarkan Total Skor

| Total Skor | | | |
|-----------------------|----------------------------|--------------------------------------|----------------------|
| Kawasan Hutan Lindung | Kawasan Budidaya Pertanian | Kawasan Lindung diluar Kawasan Hutan | Tingkat Lahan Kritis |
| 120-180 | 115-200 | 110-200 | Sangat Kritis |
| 181-270 | 201-275 | 201-275 | Kritis |
| 271-360 | 276-350 | 276-350 | Agak Kritis |
| 361-450 | 351-425 | 351-425 | Potensial Kritis |
| 451-500 | 426-500 | 426-500 | Tidak Kritis |

(sumber : Dirjen Pengendalian DASHL, 2018)

3. Data Mining

Data mining merupakan pemahaman mendalam mengenai data yang dicapai melalui operasi studi yang melibatkan proses pengumpulan, pembersihan, pemrosesan dan analisis sekumpulan data (Purwanto et al., 2018). Menurut Jayasree, Balan (2013) proses mendapatkan informasi rahasia, belum ditemukan, akurat, dan dapat digunakan dari *database* besar dan menggunakan informasi tersebut untuk membuat pilihan bisnis penting dikenal sebagai *Data mining*.

4. Knowledge Discovery in Database (KDD)

KDD adalah metode terstruktur untuk mengekstraksi pola yang valid, baru, berguna dari kumpulan data yang besar dan rumit agar dapat dipahami (Ramadhan, 2017).

Proses menemukan informasi tersembunyi dalam basis data yang cukup besar disebut sebagai *Knowledge Discovery in Database* (KDD) atau *Data mining*. *Data mining* adalah tahapan dalam keseluruhan proses KDD, meskipun kedua nama tersebut memiliki definisi yang berbeda tetapi saling terkait. Proses atau tahapan KDD digambarkan pada **Gambar 2.6** terdiri dari urutan berulang dari langkah berikut:

a. Pembersihan Data (Data Cleaning)

Data yang menjadi fokus *Knowledge Discovery in Databases* (KDD) harus terlebih dahulu menjalani pembersihan agar dapat digunakan dalam proses *Data mining*. Prosedur ini memerlukan sejumlah langkah, termasuk menghilangkan data berganda, memeriksa data yang inkonsisten, dan memperbaiki masalah data seperti kesalahan pencetakan.

b. Integrasi Data (Data Integration)

Proses menggabungkan data dari beberapa *database* menjadi satu *database* baru dikenal sebagai integrasi data. karena data yang diperlukan untuk proses penambangan data terkadang berasal dari banyak basis data, bukan hanya satu. Proses pengintegrasian data perlu kehati-hatian karena jika ada kesalahan, hasilnya dapat disalahtafsirkan atau bahkan menyesatkan untuk ditindaklanjuti.

c. Seleksi Data (Data Selection)

Sebelum tahap *Knowledge Discovery in Database* (KDD) proses ekstraksi informasi dimulai dengan memilih data yang berasal dari sekumpulan data yang harus diselesaikan. Hasil proses tersebut akan digunakan dalam proses *Data mining* dan disimpan dalam file terpisah.

d. Transformasi Data (Data Transformation)

Proses dimana data dikonsolidasikan atau diubah menjadi format yang sesuai dengan *Data mining* dengan menjalankan prosedur ringkasan atau agregasi. Transformasi data ini dikatakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

e. Data mining

Untuk mengekstraksi pola data akan terjadi terlebih dahulu proses pencarian pola/informasi menarik dalam data dengan metode cerdas seperti algoritma dan teknik lainnya.

f. Evaluasi Pola (Pattern Evaluation)

Pemeriksaan dilakukan untuk melihat kesesuaian pola atau informasi yang didapatkan apakah sejalan dengan hipotesis yang ada sebelumnya atau bahkan sebaliknya.

g. Presentasi Pengetahuan

Tahapan terakhir bagaimana teknik visualisasi dan representasi pengetahuan akan digunakan untuk mempresentasikan pengetahuan mudah dimengerti serta dipahami oleh pengguna.

5. Gaussian Naive Bayes

Salah satu varian dari *Naive Bayes* adalah *Gaussian Naive Bayes*. Metode ini merupakan distribusi yang memiliki sifat kontinu dengan ciri-ciri dari rata-rata variansinya. *Gaussian naive bayes* bekerja pada tipe data angka (Cahyaningrum et al.,

2022), berikut adalah bentuk pendistribusian dari *gaussian naive bayes* :

$$P(X_i) = X_i|Y = Y_j$$

$$= \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(X_i-\mu_{ij})^2}{2\sigma_{ij}^2}} \quad (1)$$

Dimana :

P: Peluang

X_i : Atribut ke i

x_i : Nilai atribut ke i

Y: Kelas yang dicari

y_j : Sub kelas Y yang dicari

μ : Mean, menyatakan rata rata dari seluruh atribut

σ : Deviasi standar, menyatakan varian dari seluruh atribut.

6. Tuning Hyperparameter

Hyperparameter merupakan perkiraan parameter nilai yang dapat meningkatkan model kinerja dalam pendekatan *machine learning*. *Hyperparameter* biasanya digunakan untuk meningkatkan kinerja dari suatu algoritma, yang cukup untuk mempengaruhi banyak model pengujian (Herni Yulianti et al., 2022).

Beberapa jenis *hyperparameter* yaitu *GridSearchCV*, *RandomSearchCV*, *bayesian optimization*, dan *evolutionary optimization* (Putri et al., 2023). *Grid search cross validation* bertujuan untuk memilih kombinasi model dan hyperparameter yang menguji setiap kombinasi secara otomatis (Nugraha & Sasongko, 2022). Selain itu *cross validation grid search* (*GridSearchCV*) digunakan untuk mengoptimalkan nilai akurasi, dan merupakan modul *scikit-learn* yang melakukan validasi pada lebih dari satu model.

7. Python

Python sering digunakan untuk scripting, analisis statistik, dan aplikasi dan pengembangan Web serba guna. Ini dapat digunakan secara interaktif atau terprogram. Banyaknya pustaka basis data, matematika, dan grafik yang cocok untuk proyek dalam eksplorasi data dan masalah *Data mining* merupakan kekuatan pemrograman *Python*. Terhubungnya *Python* dengan sebagian besar *database* SQL dan NoSQL memungkinkan *Python* dapat memuat serta memproses data mentah dan tabular dengan relatif mudah. Scipy dan Numpy pustaka yang ada pada *Python* yang mengimplementasi model matematika dan statistik. Selain itu pustaka *matplotlib* pada *Python* dapat menghasilkan visualisasi data yang kompleks serta bernuansa. *Python* menjadi lebih banyak digunakan dalam *Data mining* dan komunitas ilmu data baru-baru ini (Dean, 2014).

8. Confusion Matrix

Pengukuran kinerja yang dapat diterapkan pada metode klasifikasi salah satunya adalah *Confusion matrix*, menggambarkan performa dari sebuah model atau algoritma menggunakan table. Berikut Tabel 1. Menampilkan *confusion matrix*.

Tabel 1 Confusion Matrix

| | | |
|--|---|---|
| | Predict Positive (Prediksi Positif) | Predict Negative (Prediksi Negatif) |
| Actual Positive (Aktual Positif) | (TP = True Positive) | (FN = False Negative) |
| Actual Negative (Aktual Negatif) | (FP = False Positive) | (TN = True Negative) |

Berikut rumus untuk mencari nilai *accuracy*, *precision*, *recall*, dan *f1-score*

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

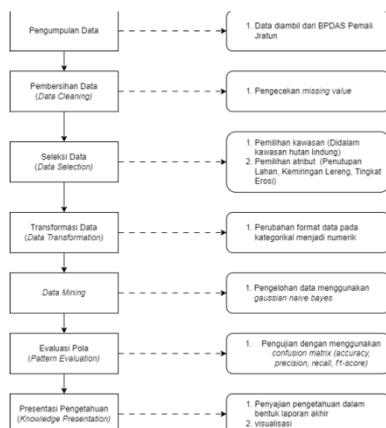
$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 - Score = \frac{2(Recall * Precision)}{Recall + Precision} \quad (5)$$

1.3 Metodologi Penelitian

Data daerah aliran sungai berdasarkan parameter lahan kritis pada Balai Pengelolaan Daerah Aliran Sungai (BPDAS) Pemali Jratun adalah objek penelitian yang menjadi pembahasan dalam penelitian ini. Sub Das yang berada dalam wilayah administrasi dari BPDAS Pemali Jratun akan dianalisis dengan menggunakan algoritma *Naive Bayes* agar dapat menentukan DAS prioritas di Kabupaten Semarang.

Pendekatan yang digunakan dalam penelitian ini dikenal dengan KDD atau *Knowledge Discovery in Database*. Terdapat 7 tahapan dalam KDD, tetapi dalam penelitian ini hanya melakukan 6 tahap yaitu Pembersihan Data (*Data Cleaning*), Integrasi Data (*Data Integration*), Seleksi Data (*Data Selection*), Transformasi Data (*Data Transformation*), *Data mining*, Evaluasi Pola (*Pattern Evaluation*), Presentasi Pengetahuan (*Knowledge Presentation*) (Han & Kamber, 2006). Berikut tahapan yang akan dilakukan pada Gambar 1.



Gambar 1 Metodologi Penelitian

2. PEMBAHASAN

Dalam Penelitian ini digunakan dataset parameter lahan kritis di Kabupaten Semarang tahun 2020 yang diperoleh dari Balai Pengelolaan Daerah Aliran Sungai (BPDAS) Pemali Jratun.

2.1. Data Uji

Dataset yang diperoleh dari BPDAS Pemali Jratun untuk Kabupaten Semarang yang berada di dalam kawasan hutan lindung sebanyak 2035 dari 18547 data, setelah dilakukan proses pengecekan *missing value* dan seleksi data

Data yang sudah dilakukan proses pengecekan *missing value* serta seleksi data mendapatkan 3 atribut parameter lahan kritis dan 1 label kekritisian lahan.

Tabel 2 Hasil Seleksi Data

| KELAS_VEG | KELAS_E ROSI | KELAS_LE RENG | KRITIS |
|--------------|---------------|---------------|------------------|
| sangat buruk | sangat ringan | landai | potensial kritis |
| sangat buruk | Ringan | curam | agak kritis |
| sangat buruk | Ringan | agak curam | kritis |
| sangat buruk | Berat | sangat curam | agak kritis |
| sangat buruk | Sedang | datar | potensial kritis |

Sebelum masuk kedalam tahapan *data mining*, penelitian ini akan menunjukkan gambaran umum dari lahan kritis yang ada di Kabupaten Semarang berdasarkan parameter lahan kritis.

Tabel 3 Perhitungan Tingkat Kekritisian

| Parameter | % Bobot | Skor | Tingkat Kekritisian Lahan | Klasifikasi Tingkat Kekritisian Lahan |
|----------------------|---------|------|---------------------------|---------------------------------------|
| Penutupan Lahan | 50 | 1 | 50 | Kritis |
| Kemiringan Lereng | 20 | 3 | 60 | |
| Tingkat Bahaya Erosi | 20 | 5 | 100 | |
| Jumlah | | | | |

Hasil perhitungan manual akan mempunyai nilai hasil penggabungan dari beberapa parameter yang

digunakan. Skor kekritisan lahan pada kawasan hutan lindung pada Kabupaten Semarang adalah Kritis dengan skor total tingkat kekritisan lahan sebesar 210.

Selanjutnya akan dilakukan transformasi data. Tahap ini menggunakan model *GaussianNB*, dimana model tersebut hanya dapat memproses data dalam bentuk angka (*numeric*) sedangkan *values* dari dataset berbentuk *object*. Oleh karena itu perlu dilakukan perubahan jenis tipe data yang semula adalah data berupa kategorik kin menjadi angka (*numeric*).

Tabel 4 Hasil Transformasi Data

| KELAS_VEG | KELAS_LERENG | KELAS_EROSI | KRITIS |
|-----------|--------------|-------------|--------|
| 4 | 4 | 1 | 0 |
| 4 | 3 | 0 | 0 |
| 3 | 2 | 0 | 1 |
| 3 | 4 | 1 | 0 |
| 3 | 4 | 1 | 0 |

2.2. Pengujian Algoritma *Gaussian Naive Bayes*

Pengujian yang dilakukan dalam penelitian ini adalah menggunakan *tuning hyperparameter GridSearchCV* bertujuan untuk memilih kombinasi model dan hyperparameter yang menguji setiap kombinasi secara otomatis (Nugraha & Sasongko, 2022). Selain itu digunakan untuk meningkatkan nilai akurasi daripada model yang akan digunakan.

Evaluasi pengujian *gaussian naive bayes* menggunakan *confussion matrix* untuk melihat nilai akurasi, presisi, recall dan f1-score yang dihasilkan dari algoritma tersebut dalam menentukan kekritisan lahan yang berada di dalam kawasan hutan lindung.

Pemodelan *Gaussian Naive Bayes* akan di uji dengan beberapa skenario pembagian data, tahapan awal menentukan variabel independen dan variabel dependen. Untuk variabel independen terdiri dari KELAS_VEG, KELAS_EROSI, KELAS_LERENG. Serta kelas dependen yaitu KRITIS.

2.3. Hasil Pengujian *Gaussian Naive Bayes* dengan *Confusion Matrix*

Tabel di bawah merupakan hasil dari pemodelan *Gaussian Naive Bayes* yang dilakukan menggunakan skenario pembagian data dilakukan sebanyak 3 kali dengan skenario 70:30, 80:20, dan 90:10. Hasil dari pemodelan diukur dengan menggunakan parameter pada *confusion matrix*, yaitu nilai *Accuracy*, *Precision*, *Recall*, dan *F1-Score*.

Tabel 5 Hasil *Confusion Matrix*

| Skenario | <i>Confusion Matrix</i> | | | |
|----------|-------------------------|------------------|---------------|-----------------|
| | <i>Accuracy</i> | <i>Precision</i> | <i>Recall</i> | <i>F1-Score</i> |
| 70:30 | 71% | 0.56 | 0.71 | 0.61 |
| 80:20 | 70% | 0.55 | 0.70 | 0.61 |
| 90:10 | 67% | 0.52 | 0.67 | 0.57 |

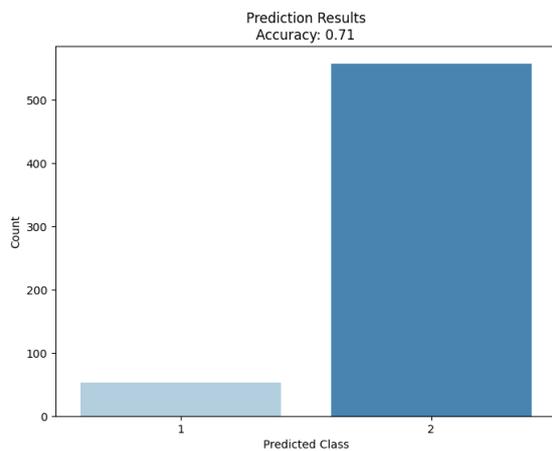
Berdasarkan hasil evaluasi model diatas, maka dapat disimpulkan bahwa pemodelan dengan rasio pembagian data 70:30 merupakan pemodelan yang optimal dengan nilai *Accuracy* sebesar 71%, nilai *Precision* sebesar 0.56, nilai *Recall* sebesar 0.71 dan nilai *F1-Score* sebesar 0.61 jika dibandingkan dengan skenario 2 maupun skenario 3. Berikut tampilan hasil evaluasi skenario 1.

Tabel 6 Hasil Prediksi

| KRITIS | Prediksi |
|------------------|------------------|
| potensial kritis | agak kritis |
| agak kritis | agak kritis |
| kritis | agak kritis |
| agak kritis | agak kritis |
| potensial kritis | agak kritis |
| agak kritis | agak kritis |
| kritis | agak kritis |
| tidak kritis | potensial kritis |
| kritis | agak kritis |
| agak kritis | agak kritis |

2.4. Presentasi Pengetahuan

Setelah tahapan pengujian telah dilakukan dan sudah mendapatkan skenario yang optimal, proses selanjutnya adalah presentasi pengetahuan dengan menampilkan visualisasi hasil dari pemodelan yang telah dilakukan. Visualisasi akan ditampilkan dalam bentuk diagram batang dengan tujuan agar dapat mudah dipahami. Berdasarkan gambar dibawah ini menunjukkan bahwa terdapat beberapa hasil klasifikasi yaitu kelas 1 dengan kriteria Potensial Kritis, kelas 2 dengan kriteria Agak Kritis. Dengan kelas terbanyak yaitu kelas 2.



Gambar 2 Visualiasi Hasil Prediksi

3. KESIMPULAN

Terdapat tiga skenario yang diuji dalam penelitian ini dengan melakukan pembagian rasio dataset. Skenario 1 perbandingan 70:30 skenario 2 perbandingan 80:20 dan skenario 3 perbandingan 90:10. Berdasarkan hasil evaluasi dari ketiga skenario, pemodelan dari rasio 70:30, memiliki akurasi yang lebih tinggi jika dibandingkan dengan skenario lainnya. Nilai *Accuracy* sebesar 71%, nilai *Precision* sebesar 0.56, nilai *Recall* sebesar 0.71 dan nilai *F1-Score* sebesar 0.61. Skenario 1 menjadi skenario terbaik dalam menerapkan klasifikasi dengan menggunakan algoritma *Naive Bayes*, hal ini dibuktikan dengan kelas yang menjadi prioritas untuk dilakukannya rehabilitasi. Hasil presentasi pengetahuan menampilkan dimana terdapat dua kelas yaitu kelas 1 dan kelas 2. Hasil klasifikasi menunjukkan bahwa pemodelan dari Kelas 1 dengan tingkat kekritisan lahan potensial kritis memiliki jumlah lebih sedikit sedangkan kelas 2 dengan tingkat kekeritisan agak kritis memiliki jumlah lebih banyak. Berdasarkan hal tersebut DAS yang menjadi prioritas pertama untuk di rehabilitasi adalah DAS dengan tingkat kekritisan agak curam, setelah itu DAS dengan tingkat kekeritisan potensial kritis menjadi prioritas kedua untuk direhabilitasi.

PUSTAKA

- Ariyani, N., Ariyanti, D. O., & Ramadhan, M. (2020). Pengaturan Ideal Tentang Pengelolaan Daerah Aliran Sungai di Indonesia (Studi di Sungai Serang Kabupaten Kulon Progo) *. 27(3), 592–614.
- BPS Jawa Tengah. (2021). *J a w a T e n g a h Province in Figures*. 1–1003.
- Cahyaningrum, H., Arifianto, D., & Abdurrahman, G. (2022). Analisis Perbandingan Metode K Nearest Neighbor Dan Gaussian Naive Bayes Pada Klasifikasi Jurusan Siswa (Studi Kasus pada Siswa SMA Muhammadiyah 3 Jember) A Comparative Analysis of K Nearest Neighbor and Gaussian Naive Bayes methods in students major clas. *Jurnal Smart Teknologi*, 3(3), 2774–1702.
- <http://jurnal.unmuhjember.ac.id/index.php/JS T>
- Dean, J. (2014). *Big Data, Data Mining, and Machine Learning Value Creation for Business Leaders and Practitioners* (pp. 49–50).
- Dirjen Pengendalian DASHL. (2018). Peraturan Direktur Jenderal Pengendalian Daerah Aliran Sungai dan Hutan Lindung Nomor P.3/PDASHL/SET/KUM.1/7/2018 Tentang Petunjuk Teknis Penyusunan Data Spasial Lahan Kritis. 1–20.
- Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques Second Edition*.
- Herni Yulianti, S. E., Oni Soesanto, & Yuana Sukmawaty. (2022). Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit. *Journal of Mathematics Theory and Application*, 4(1), 21–26. <https://doi.org/10.31605/jomta.v4i1.1792>
- Kristiano, D. H., Noertjahyani, N., & Sufiadi, E. (2021). Analisis Kekritisan Lahan Pada Fungsi Kawasan Konservasi DAS Citarum Dengan Software Arcgis. *OrchidAgro*, 1(1), 14. <https://doi.org/10.35138/orchidagro.v1i1.237>
- Nugraha, W., & Sasongko, A. (2022). Hyperparameter Tuning pada Algoritma Klasifikasi dengan Grid Search. *SISTEMASI: Jurnal Sistem Informasi*, 11(2), 391–401.
- Purwanto, A., Darmadi, E. A., Trimitra, P., Mandiri, K., By, J., Jomin -Blok, P., & Barat -Kotabaru -Cikampek -Karawang, S.-J. (2018). *Perbandingan Minat Siswa Smu Pada Metode Klasifikasi Menggunakan 5 Algoritma*. 2(1), 43–47.
- Putri, T. A. E., Widiharih, T., & Santoso, R. (2023). Penerapan Tuning Hyperparameter Randomsearchcv Pada Adaptive Boosting Untuk Prediksi Kelangsungan Hidup Pasien Gagal Jantung. *Jurnal Gaussian*, 11(3), 397–406. <https://doi.org/10.14710/j.gauss.11.3.397-406>
- Ramadhan, M. I. (2017). Penerapan Data Mining untuk Analisis Data Bencana Milik BNPB Menggunakan Algoritma K-Means dan Linear Regression. *Jurnal Informatika Dan Komputer*, 22(1), 57–65.
- Yulianita, T., & Istiawan, D. (2017). Implementasi Algoritma K-modes untuk Penentuan Prioritas Rehabilitasi Daerah Aliran Sungai Berdasarkan Parameter Lahan Kritis. *The 6th University Research Colloquium 2017, September*, 429–440.