

## **Optimalisasi Sistem Informasi Geografis Untuk Prediksi Tanah Longsor Menggunakan Simulasi *Machine Learning Maximum Entropy (MaxEnt)* Desa Batuah, Kecamatan Loa Janan, Kabupaten Kutai Kartanegara, Provinsi Kalimantan Timur**

**Budi Wiyono<sup>1)\*</sup>, Imanuel Kaunang<sup>1)</sup>, Jamaluddin<sup>1)</sup>**

<sup>1)</sup> 1Progam Studi Geologi, Sekolah Tinggi Teknologi Migas Balikpapan

E-mail: [budiw229@gmail.com](mailto:budiw229@gmail.com)

### **ABSTRAK**

Analisis kerawanan longsor ini menerapkan integrasi Sistem Informasi Geografis dan pemodelan *Machine Learning berbasis Maximum Entropy (MaxEnt)* untuk memetakan zona risiko di Desa Batuah, Kecamatan Loa Janan, Kabupaten Kutai Kartanegara, Kalimantan Timur. Tujuan penelitian meliputi evaluasi korelasi antar parameter lingkungan, penilaian kinerja model menggunakan *Area Under the Curve (AUC)*, dan identifikasi zona kerawanan longsor. Data yang digunakan meliputi lima parameter utama sesuai SNI 8291-2016: kemiringan lereng, litologi, jarak dari kelurusan struktur (liniasi), tata guna lahan, dan titik kejadian gerakan tanah. Tahap pra-pemrosesan mencakup reproyeksi, standarisasi resolusi raster ke ukuran sel tiga puluh meter, dan konversi format agar semua layer memiliki dimensi identik. Untuk menghindari multikolinearitas, dilakukan pengujian korelasi Pearson untuk variabel kuantitatif dan Cramer's V untuk variabel kategorikal; hasil menunjukkan tidak ada korelasi tinggi antar parameter sehingga layak dijadikan input *MaxEnt*. Pemodelan dijalankan dengan metode cross-validation untuk membagi data presensi menjadi subset pelatihan dan pengujian bergilir. Kinerja model dievaluasi menggunakan kurva ROC dan AUC dengan nilai AUC uji sebesar 0,825, yang dikategorikan sebagai baik. Peta keluaran *MaxEnt* mengelompokkan wilayah menjadi tiga zona kerawanan: tinggi, sedang, dan rendah; zona berpotensi tinggi terutama terletak di lereng curam dengan litologi rentan dan perubahan tata guna lahan akibat aktivitas manusia. Dalam tahap pra-pemrosesan juga dilakukan pengisian nilai nodata (-9999) dan penyesuaian header raster serta penggunaan *plugin Molusce* untuk analisis korelasi dan validasi dimensi layer. Hasil peta kerawanan diartikan menjadi dasar bagi perencanaan mitigasi, pengendalian lahan, dan penentuan prioritas intervensi pengurangan risiko di wilayah studi.

**Kata Kunci:** *Environmental Parameters, Geographic Information System (GIS), Landslide Susceptibility, Maximum Entropy (MaxEnt), Machine Learning*

### **ABSTRACT**

*This landslide susceptibility analysis applies the integration of Geographic Information Systems and Maximum Entropy (MaxEnt)-based Machine Learning modeling to map risk zones in Batuah Village, Loa Janan District, Kutai Kartanegara Regency, East Kalimantan. The research objectives include evaluating the correlation between environmental parameters, assessing model performance using Area Under the Curve (AUC), and identifying landslide susceptibility zones. The data used includes five main parameters according to SNI 8291-2016: slope gradient, lithology, distance from structural alignment (lineage), land use, and ground motion occurrence points. The pre-processing stage includes reprojection, standardizing raster resolution to a thirty-meter cell size, and format conversion so that all layers have identical dimensions. To avoid multicollinearity, Pearson correlation tests were performed for quantitative variables and Cramer's V for categorical variables; the results showed no high correlation between parameters, making them suitable for use as MaxEnt input. The modeling was run using a cross-validation method to divide the presence data into alternating training and testing subsets. Model performance was evaluated using ROC and AUC curves with a test AUC value of 0.825, which is categorized as good. The MaxEnt output map groups the area into three vulnerability zones: high, medium, and low; the high-potential zone is mainly located on steep slopes with vulnerable lithology and land use changes due to human activities. The pre-processing stage also includes filling in nodata values (-9999) and adjusting raster headers. The Molusce plugin is used for correlation analysis and layer dimension validation. The resulting vulnerability map is expected to be the basis for mitigation planning, land control, and prioritizing risk reduction interventions in the study area.*

*Keyword: Environmental Parameters, Geographic Information System (GIS), Landslide Susceptibility, Maximum Entropy (MaxEnt), Machine Learning*

## 1. Pendahuluan

Tanah longsor merupakan salah satu ancaman geologis global yang paling merusak, mengakibatkan hilangnya nyawa, kerugian ekonomi yang besar, dan kerusakan lingkungan yang parah. Faktor-faktor yang memengaruhi gerakan tanah meliputi kondisi geologi, kemiringan lereng, keberadaan air, aktivitas manusia, serta gempa bumi (Shafapourtehrany, M., Panahi, M., Rezaie, F., Khosravi, K., Kalantari, Z., Bateni, S. M., & Lee, 2025).

Sistem Informasi Geografis (SIG) merupakan sebuah revolusi dalam cara kita memahami dan mengelola informasi geografis. Dengan kemampuan untuk menggabungkan, menganalisis, dan memvisualisasikan data yang terkait dengan lokasi geografis. Di era teknologi informasi saat ini, Sistem Informasi Geografis (SIG) telah menjadi alat penting dalam mengelola dan memvisualisasikan data yang berkaitan dengan lokasi geografis. (Erkamim, 2023).

Dalam pemetaan kerentanan longsor, *Machine Learning* menunjukkan keunggulan yang signifikan dibandingkan metode statistik tradisional, terutama karena kemampuannya memodelkan hubungan non-linear yang kompleks dalam data besar dan di area dengan data geoteknik yang minim (Ado, M., Amitab, K., Maji, A. K., Jasińska, E., Gono, R., Leonowicz, Z., & Jasiński, 2022).

Pada daerah Desa Batuah, yang terletak di Kecamatan Loa Janan, Kabupaten Kutai Kartanegara, Provinsi Kalimantan Timur. Desa ini merupakan salah satu desa yang mengalami pertumbuhan pembangunan yang cukup pesat, baik dari sisi pemukiman, aktivitas ekonomi, hingga eksploitasi sumber daya alam, terutama yang berkaitan dengan sektor pertambangan dan perkebunan. Namun dibalik perkembangan tersebut, terdapat tantangan besar dalam hal kerawanan bencana, terutama bencana banjir saat musim hujan, tanah longsor di sekitar sekitar jalan utama di jalan poros Balikpapan - Samarinda.

## 2. Metode Penelitian

### A. Prepasari Data

Untuk melakukan pemetaan potensi bahaya longsor, terlebih dahulu kita mempersiapkan data. Hal yang pertama yang dilakukan adalah menentukan parameter apa saja yang akan kita gunakan dalam menyusun zonasi kerentanan gerakan tanah. Parameter yang akan digunakan dalam pembuatan peta kerentanan gerakan tanah sesuai dengan SNI 8291-2016 terdiri dari :

- 1) Kemiringan lereng
- 2) Litologi
- 3) Jarak dari kelurusan struktur
- 4) Tata Guna Lahan
- 5) Titik Gerakan Tanah
- 6) Curah hujan

Oleh karena itu untuk kegiatan pre-processing ini perlu untuk mempersiapkan data-data tiap parameter tersebut dengan cara memasukkan semua data tersebut ke QGIS.

### B. Pearson Correlation dan Cramer's Coefficient (Cramer's V)

Dalam menjalankan variabel input dalam pemodelan MaxEnt, penting untuk memeriksa korelasi antar variabel lingkungan guna menghindari kolinieritas sehingga perlu menguji dengan *Pearson Correlation* dan *Cramer's Coefficient (Cramer's V)*. Korelasi Pearson, yang juga dikenal sebagai koefisien korelasi produk-momen Pearson, adalah ukuran statistik dari kekuatan dan arah hubungan linear antara dua variabel kuantitatif (Negi et al., 2025). Diinterpretasikan dalam rumus sebagai berikut.

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

$X_i, Y_i$  : nilai dari dua variabel yang sedang dianalisis.

$X, Y$  : rata-rata dari masing-masing variabel

$R$  : Koefisien korelasi Pearson

Pada metode Pearson Correlation dalam pengklasifikasian penilaian hasil korelasi, dimana nilai koefisien Pearson ( $r$ ) diinterpretasikan menggunakan pendekatan yang lebih fleksibel karena fenomena alam dipengaruhi oleh banyak faktor yang saling berinteraksi sehingga hubungan antar variabel jarang menunjukkan korelasi yang sangat kuat. Nilai  $r$  berada pada rentang  $-1$  hingga  $+1$ , dengan nilai mendekati  $-1$  atau  $+1$  menunjukkan korelasi kuat, sedangkan nilai mendekati  $0$  menunjukkan hubungan yang lemah (Mukaka, 2012). Interpretasi dalam konteks lingkungan menggunakan kategori  $<0,30$  sebagai korelasi lemah,  $0,30-0,70$  sebagai korelasi sedang, dan  $>0,70$  sebagai korelasi kuat baik pada arah positif maupun negatif. Korelasi yang terlalu tinggi ( $>0,70$ ) umumnya dihindari dalam pemodelan prediktif seperti MaxEnt karena berpotensi menimbulkan multikolinearitas, sehingga seleksi variabel dilakukan dengan mempertimbangkan kekuatan hubungan antar variabel lingkungan. Pendekatan ini sejalan dengan rekomendasi literatur ekologis dan geospasial yang menekankan pentingnya pengendalian korelasi antar variabel untuk menjaga keakuratan dan stabilitas model (Dormann, C. F., 2013).

Cramer's  $V$  merupakan ukuran statistik yang digunakan untuk menilai kekuatan asosiasi antara dua variabel kategori dan merupakan pengembangan dari uji Chi-Square. Secara matematis, Cramer's  $V$  dirumuskan sebagai (McHugh, 2013)

$$V = \sqrt{\frac{\chi^2}{n(k-1)}}$$

$\chi^2$  : nilai Chi-Square.

$n$  : jumlah total sampel

$rV$  : Koefisien Cramers's

Pada metode Cramer's  $V$  dalam pengklasifikasian penelian dari korelasi data antara dua variabel katagori di bagi mejadi 3 dimana jika nilai  $<0,2$  (lemah), ini di dasarkan variabel lingkungan bersifat kompleks, dipengaruhi banyak faktor, dan kondisi alam selalu dinamis sehingga hubungan antar-variabel jarang sangat kuat (Akoglu, 2018).  $0,2-0,4$  (sedang), ini di dasarkan dalam studi geospasial, hubungan moderat dianggap wajar karena fenomena alam tidak dikendalikan oleh satu variabel tunggal (Sheskin, 2003).  $>0,4$  (tinggi), ini di dasarkan pada analisi lingkungan nilai ini sudah menunjukkan asosiasi signifikan pada konteks lingkungan, meski tidak setinggi standar statistik konvensional (Cramér, 1946).

### C. Maximum entropy (Maxent)

merupakan teknik machine learning sederhana untuk membangun suatu prediksi atau inferensi dari suatu informasi yang belum lengkap (Phillips et al., 2006). Berdasarkan prinsipnya, distribusi harus memiliki entropi maksimum ketika tidak diketahui sebaran awal dari data tersebut.

$$H(\pi) = -\sum_{x \in X} \pi(x) \ln \pi(x)$$

$\pi$  : Distribusi probabilitas yang belum diketahui

$x$  : Elemen individual dari poin-poin presensi (piksel-piksel presensi).

### D. Area Under the Curve (AUC)

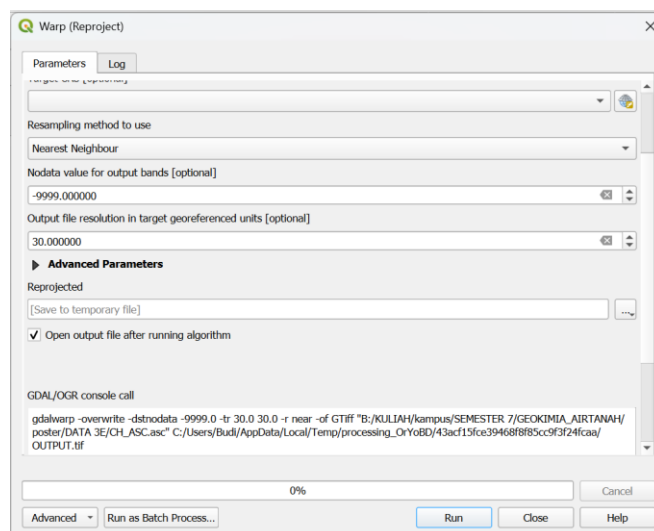
Area Under the Curve (AUC) merupakan indikator kinerja yang digunakan untuk mengevaluasi kemampuan model dalam membedakan kelas positif dan negatif, seperti antara wilayah yang berpotensi longsor dan wilayah yang tidak rawan. Nilai AUC diperoleh dari kurva Receiver Operating Characteristic (ROC), yang menunjukkan hubungan antara True Positive Rate dan False Positive Rate pada berbagai ambang klasifikasi.), interpretasi nilai AUC dibagi ke dalam beberapa kategori, di mana rentang nilai  $0,9$  hingga  $1,0$  mencerminkan bahwa performa model berada pada tingkat yang sangat baik dan mampu melakukan diskriminasi kelas secara hampir sempurna (Mohammed Bekkar, Hamid Kheddar, 2013)

Tabel 1 Perfoema model berdasarkan nilai AUC

| NILAI AUC | PERFORMA MODEL |
|-----------|----------------|
| 0,5-0,6   | Buruk          |
| 0,6-0,7   | Cukup          |
| 0,7-0,8   | Baik           |
| 0,8-0,9   | Sangat Baik    |
| 0,9-1,0   | Sempurna       |

### 3. Hasil dan Pembahasan

Sebelum melakukan analis kerawanan bencana longsor, harus melakukan preparasi data raster yang menjadi parameter dalam analisis kerawanan bencana longsor. Dalam hal ini “*output bands*” dan “*resolution cell*” pada setiap parameter menjadi “-9999” dan “30” dengan menggunakan tools “*warp(reproject)*”.



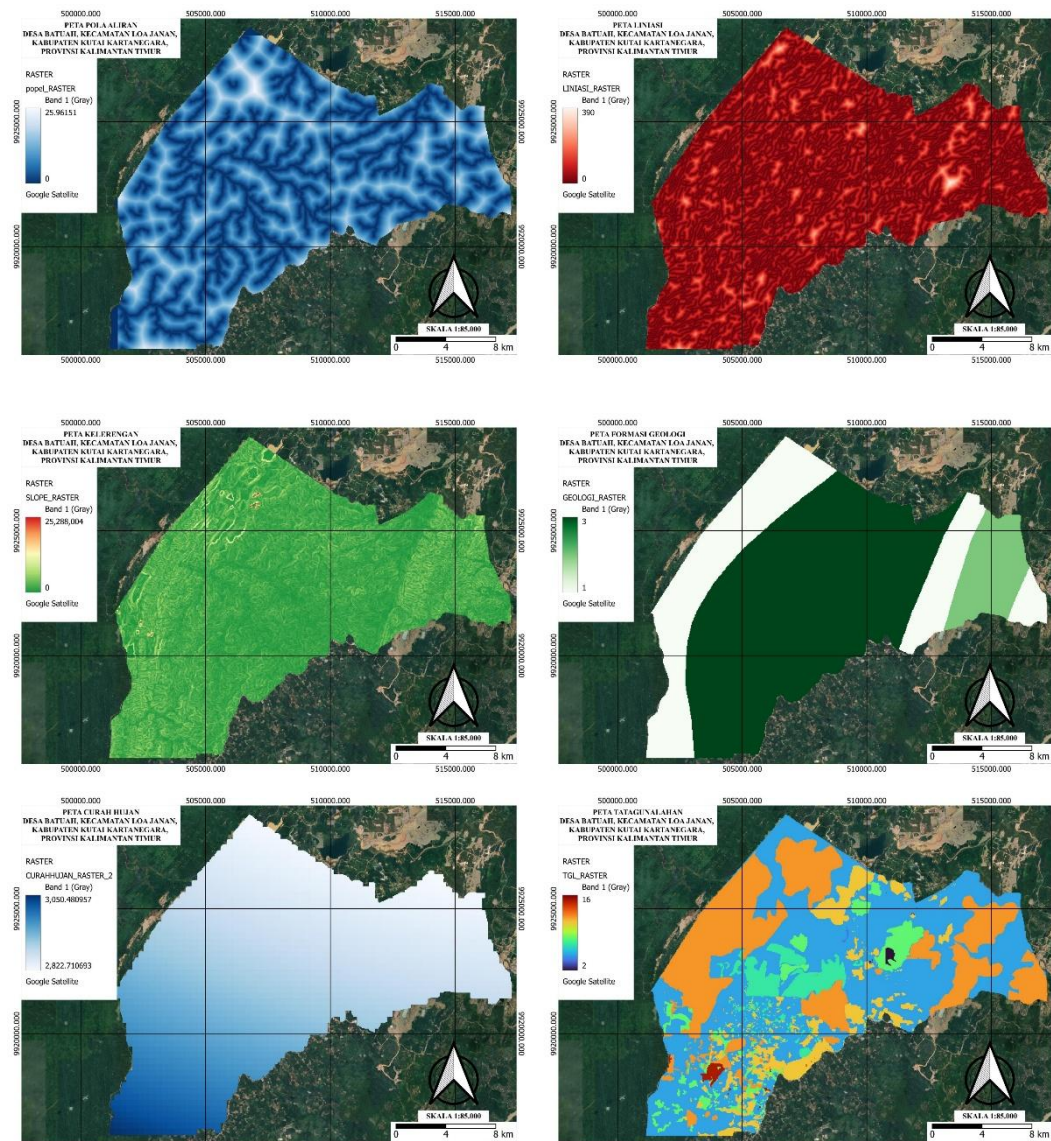
Gambar 1 Tools warp(Reproject)

Dalam setiap pembuatan data raster terdapat masalah terkait dimensi yang berbeda antara parameter lainnya, sehingga setiap parameter harus memiliki dimensi yang sama, dengan mengubah *header* dengan mengubah format raster dari “.tif” menjadi “.asc”. dalam preparasi ini hanya parameter kelerengan yang memiliki dimensi yang berbeda.

```

HEADER
TGL
ncols      538
nrows      438
sllcorner  581124.172699999996
yllycorner  9915896.0960000000034
cellsize   30.0000000000000
NODATA_value -9999
SLOPE
ncols      538
nrows      438
sllcorner  581118.327200000000
yllycorner  9915900.4461000000252
cellsize   30.0000000000000
NODATA_value -9999
LITTOLOGI
ncols      538
nrows      438
sllcorner  581124.172699999996
yllycorner  9915896.0960000000034
cellsize   30.0000000000000
NODATA_value -9999
LINIASI
ncols      538
nrows      438
sllcorner  581124.172699999996
yllycorner  9915896.0960000000034
cellsize   30.0000000000000
NODATA_value -9999
    
```

Gambar 2 Header pada setiap parameter



Gambar 3 Hasil peta pada setiap parameter

Selanjutnya dilakukan korelasi antar parameter-paramter dengan menggunakan metode *Pearson Correlation* dan *Cramer's Coefficient (Cramer's V)*. *Pearson Correlation* menguji korelasi parmeter yang memiliki data yang linier seperti jarak kelurusan struktur dan curah hujan selanjutnya *Cramer's Coefficient (Cramer's V)* menguji korelasi parameter yang memiliki data variabel katagori Kemiringan lereng, Litologi, dan Tata Guna Lahan. Pengujian menggunakan *plugin molusce*. Sehingga menampilkan tabel sebagai berikut.

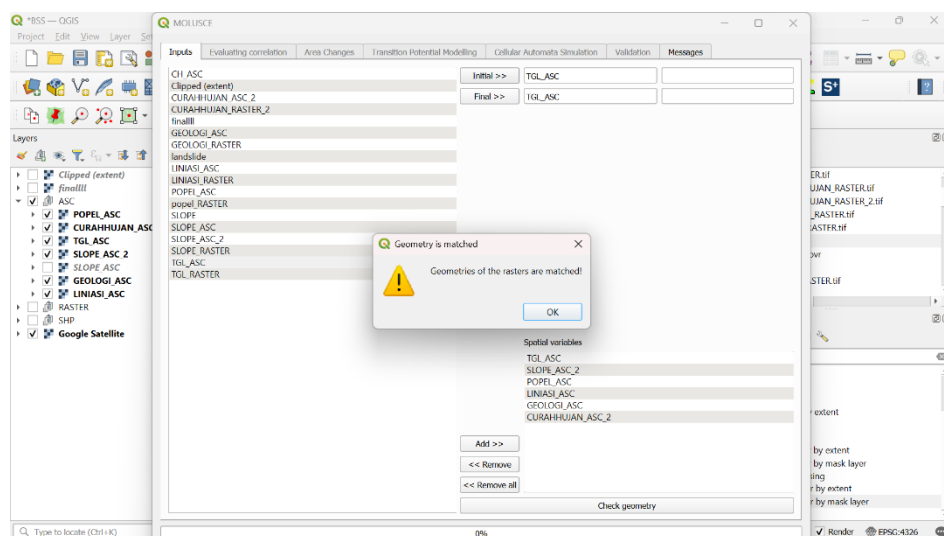
Tabel 2 Korelasi Cramer's Coefficient (Cramer's V)

| Parameter     | Formasi Geologi    |
|---------------|--------------------|
| Tatagunalahan | 0.3002234307331753 |



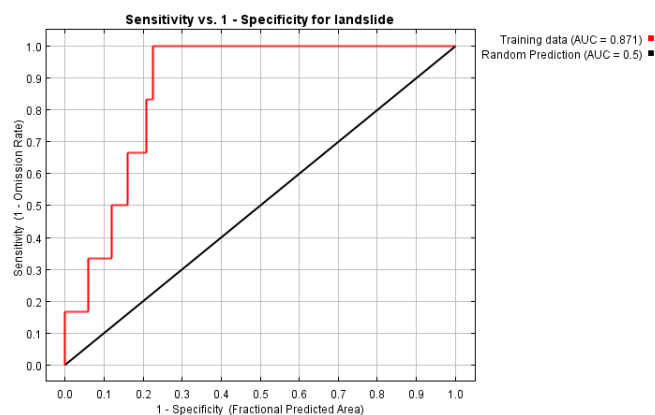
Tabel 3 Korelasi Pearson Correlation

| Parameter          | Kelerengan          | Pola aliran sungai  | Liniasi             | Curah hujan         |
|--------------------|---------------------|---------------------|---------------------|---------------------|
| Kelerengan         | --                  | 0.1726048872187596  | 0.09770646440827507 | 0.02238529714468971 |
| n                  |                     |                     |                     | 7                   |
| Pola aliran sungai | 0.1726048872187596  | --                  | 0.08917174162241587 | 0.04682003773212619 |
|                    |                     |                     |                     | 5                   |
| Liniasi            | 0.02238529714468971 | 0.08917174162241587 | --                  | 0.00684940067459394 |
|                    | 7                   |                     |                     | 9                   |
| Curah hujan        | 0.02238529714468971 | 0.04682003773212619 | 0.00684940067459394 | --                  |
|                    | 7                   | 5                   | 9                   |                     |



Gambar 4 Pengujian dimensi \ pada parameter

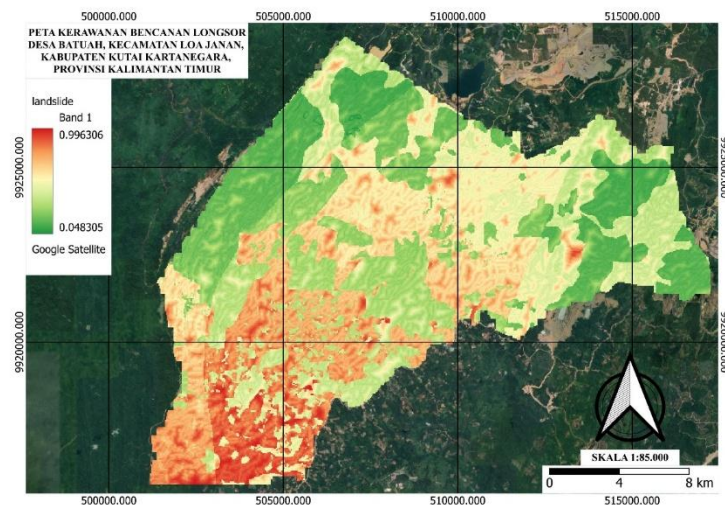
Trakhir pembuatan pemodelan kerawanan bencana dengan menggunakan metode *Maximum entropy* (*Maxent*) pada sistem ini dijalankan dengan pengujian dan pelatihan sampel dengan *Crossvalidate* yang dimana membagi seluruh data presensi menjadi beberapa bagian. Model akan dijalankan berapa kali, di mana setiap bagian secara bergiliran menjadi data uji (*testing*), sisanya menjadi data latih (*training*). Sehingga menghasilkan *outout* sebagai berikut.



Gambar 5 Grafik ROC



Grafik berikut ini adalah kurva karakteristik operasi penerima (ROC) untuk data yang sama. Perhatikan bahwa spesifisitas didefinisikan menggunakan area prediksi, mengartikan bahwa AUC maksimum yang dapat dicapai kurang dari 1. Jika data uji diambil dari distribusi Maxent itu sendiri, maka AUC uji maksimum yang mungkin adalah 0,871.



Gambar 6 Peta kerawanan bencana

Representasi model Maxent untuk longsor. Warna yang lebih hangat menunjukkan area dengan kondisi prediksi yang lebih baik. Titik putih menunjukkan lokasi keberadaan yang digunakan untuk pelatihan, sementara titik ungu menunjukkan lokasi pengujian

#### 4. Kesimpulan

Pada pembuatan analisis kerawanan bencana longsor, dimensi pada setiap parameter memiliki ukuran yang sama berdasarkan analisis menggunakan plug in molusca. Berdasarkan korelasi antara parameter dapat dilihat dengan menggunakan metode Correlation dan Cramer's Coefficient (Cramer's V) dalam pengujian ini memiliki hasil yang bagus dimana dimensi antar parameter memiliki nilai di bawah satu. Pada nilai Area Under the Curve (AUC) berdasarkan gambar 4 memiliki nilai 0,825 dimana menurut tabel 1 memiliki nilai yang baik. Pada hasil korelasi data tersebut menghasilkan zona kerawanan bencana longsor dimana dapat di bagi menjadi 3 yaitu tinggi, sedang, rendah dalam bencana longsor

#### 5. Daftar Pustaka

- Ado, M., Amitab, K., Maji, A. K., Jasińska, E., Gono, R., Leonowicz, Z., & Jasiński, M. (2022). Landslide Susceptibility Mapping Using Machine Learning: A Literature Survey. Remote Sensing. <https://doi.org/10.3390/rs14133029>
- Akoglu, H. (2018). User's guide to correlation coefficients. Turkish Journal of Emergency Medicine, 18, 91–93.
- Cramér, H. (1946). Mathematical Methods of Statistics. Princeton University Press.
- Dormann, C. F., et al. (2013). Collinearity: a review of methods to deal with it and a simulation study. Ecography.
- Erkamim, M. (2023). SISTEM INFORMASI GEOGRAFIS (SIG) TEORI KOMPREHENSIF SIG (E. Rianty (ed.)). PT. GREEN PUSTAKA INDONESIA.
- McHugh, M. L. (2013). The chi-square test of independence. Biochemia Medica.
- Mohammed Bekkar, Hamid Kheddar, T. A. (2013). Evaluation Measures for Models Performance. International Journal of Computer Applications (IJCA), 70, 8.
- Mukaka, M. M. (2012). A guide to appropriate use of correlation coefficient in medical research. Malawi Medical Journal.
- Negi, P. S., Paul, S., & Lata, S. (2025). MaxEnt distribution modelling of Betula utilis D . Don under changing climatic conditions in North Western Himalaya , India. Discover Geoscience, 1–16.



Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *ELSEVIER*, 190, 231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>

Shafapourtehrany, M., Panahi, M., Rezaie, F., Khosravi, K., Kalantari, Z., Bateni, S. M., & Lee, J. A. (2025). Beyond boundaries: AI-optimized global landslide susceptibility mapping. *Natural Hazards and Risk*.

Sheskin, D. J. (2003). *Handbook of Parametric and Nonparametric Statistical Procedures*. Chapman & Hall/CRC.