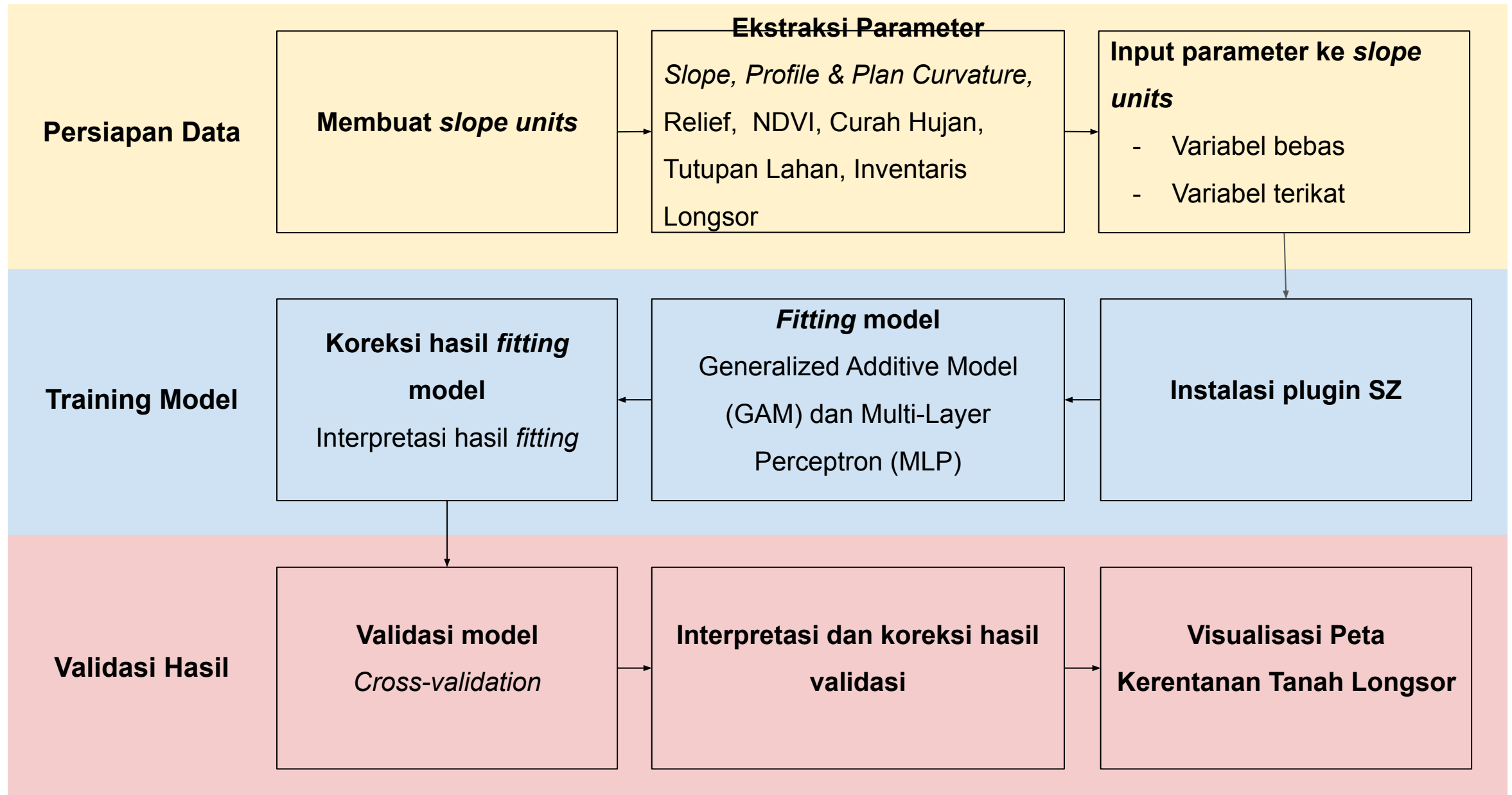


Persiapan Pengolahan Data

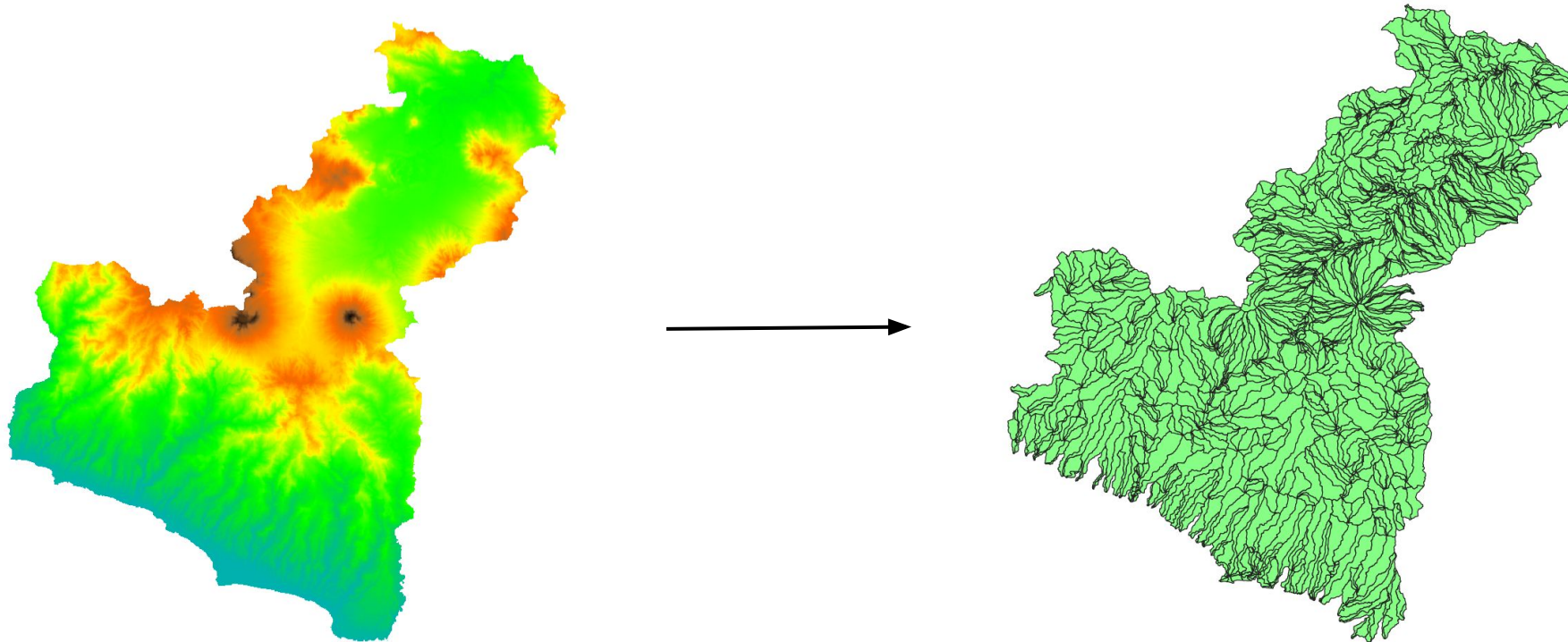
Konsep Pengolahan Data

Disusun oleh **Martines Pasaribu, S.Kel**



Slope Units (SU)

- Merupakan **unit pemetaan** (vektor) hasil pembentukan dari data DEM berupa **daerah aliran sungai (*watershed*)** yang digunakan sebagai **area perhitungan kerentanan tanah longsor**.
- Tools yang digunakan → GRASS GIS (Plugin) : *r.watershed*



Ekstraksi Parameter

- Parameter yang digunakan sebagai input :
 - Rata-rata kemiringan lereng (**slope mean**)
 - Rata-rata *profile* dan *plan curvature* (**profile curvature mean** dan **plan curvature mean**)
 - Rata-rata relief (**relief mean**)
 - Rata-rata tahunan NDVI (**NDVI mean**) → Temporal : 2015 - 2025
 - Rata-rata maksimum harian curah hujan per tahun (**Rain mean**) → Temporal : 2015 - 2025
 - **Kategori tutupan lahan** (Bangunan/gedung, Pemukiman, Lahan Kosong, Hutan)
 - **Titik dan waktu** kejadian **tanah longsor** → Temporal : 2005 - 2023
- Seluruh parameter harus dalam bentuk vektor, sebelum diinput ke *slope units*.

Input Parameter

- Setiap parameter diinput ke *slope units* dengan tools : **Overlap Analysis**.
- Hasilnya → **Setiap SU memiliki nilai setiap parameter input**.
- Variabel Independen (bebas) dan variabel dependen (terikat) ditentukan sebagai input model.
- **Variabel bebas** : *slope*, *plan* dan *profile curvature*, relief, NDVI, curah hujan, tutupan lahan.
- **Variabel terikat** : *slope units* yang terdapat kejadian longsor (diberi nilai 1, jika tidak = 0)

Generalized Additive Model (GAM)

- Model statistik yang memperluas Generalized Linear Model (GLM) dengan memungkinkan hubungan non-linear antara variabel independen dan variabel dependen menggunakan fungsi smooth (splines).

$$g(\mu) = \alpha + \sum_{j=1}^p f_j(X_j) \quad (\text{Titti et al. 2024})$$

α = Intercept (nilai dasar sebelum mempertimbangkan variabel independen).

$f_j(X_j)$ = Fungsi smooth yang merepresentasikan hubungan linear atau non-linear antara faktor predisposisi X_j dengan respons Y

p = Jumlah variabel independen yang digunakan dalam model.

$g(\mu)$ = Fungsi link GAM

- Terdapat 2 pendekatan menggunakan GAM :
 - Binomial → berfungsi untuk klasifikasi **kerentanan** longsor
 - Gaussian → berfungsi untuk identifikasi **intensitas** (luas atau volume) longsor

Generalized Additive Model (GAM)

Binomial Likelihood

- Memprediksi kerentanan longsor pada suatu lokasi berdasarkan faktor-faktor lingkungan.
- Jika Y adalah variabel biner (1 jika terjadi longsor, 0 jika tidak), maka model GAM menggunakan distribusi Bernoulli :

$$Y \sim \text{Bernoulli}(\mu)$$

- Fungsi link **log** digunakan untuk mengubah probabilitas menjadi skala linear :

$$\log \left(\frac{\mu}{1 - \mu} \right) = \alpha + \sum_{j=1}^p f_j(X_j)$$

dimana, μ adalah probabilitas rentan longsor, $1 - \mu$ adalah probabilitas tidak rentan longsor.

- Model ini menghasilkan peta **probabilitas kerentanan longsor**, di mana nilai mendekati 1 berarti lokasi tersebut sangat rentan longsor, sementara mendekati 0 tidak rentan longsor.

Generalized Additive Model (GAM)

Gaussian Likelihood

- Regresi digunakan untuk memprediksi luas atau volume longsor berdasarkan faktor lingkungan.
- Jika Y adalah luas longsor dalam satuan meter persegi, maka model GAM menggunakan distribusi Gaussian :

$$Y \sim \mathcal{N}(\mu, \sigma^2)$$

- Fungsi link identitas digunakan :

$$\mu = \alpha + \sum_{j=1}^p f_j(X_j)$$

- Model ini menghasilkan indeks **intensitas kejadian longsor**.

Multi-Layer Perceptron

- Multi-Layer Perceptron (MLP) termasuk model dari Artificial Neural Network (ANN) yang menggunakan beberapa lapisan (*layers*) neuron untuk menangkap hubungan kompleks antara variabel prediktor dan variabel respons.
- MLP terdiri dari 3 lapisan :
 - **Input Layer** : Menerima nilai dari variabel prediktor.
 - **Hidden Layer** : Menerapkan bobot dan fungsi aktivasi untuk menangkap pola non-linear.
 - **Output Layer** : Menghasilkan output berupa probabilitas (untuk klasifikasi) atau nilai kontinu (untuk regresi).
- MLP bekerja dengan *forward propagation* (perhitungan maju) dan *backpropagation* (pembelajaran balik) untuk menyesuaikan bobot dan bias dalam proses pembelajaran.

Multi-Layer Perceptron

- Perhitungan dasar *forward propagation* : Setiap neuron dalam *hidden layer* menerima masukan dari neuron di lapisan sebelumnya. Neuron ini melakukan perhitungan menggunakan bobot (w) dan bias (b)

$$z_i^{(k)} = h_i^{(k)} \left(b_i^{(k)} + \sum_{j=1}^{n_{k-1}} w_{ij}^{(k)} z_j^{(k-1)} \right)$$

dimana,

$z_i^{(k)}$ = nilai aktivasi neuron ke- i di lapisan ke- k

$w_{ij}^{(k)}$ = bobot antara neuron ke- j di lapisan ke- $k-1$ dengan neuron ke- i di lapisan ke- k

$b_i^{(k)}$ = bias neuron ke- i di lapisan ke- k

h = fungsi aktivasi

Multi-Layer Perceptron

MLP Classification

- MLP klasifikasi digunakan untuk menentukan apakah suatu lokasi rentan atau tidak rentan terhadap longsor berdasarkan faktor lingkungan seperti curah hujan, elevasi, dan kemiringan lereng.
- Lapisan output menggunakan fungsi aktivasi sigmoid :

$$h(a) = \frac{1}{e^{-a} + 1}$$

- Fungsi Loss yang digunakan adalah **Binary Cross-Entropy** :

$$\mathcal{L}(y, p) = -(y \log(p) + (1 - y) \log(1 - p))$$

dimana,

$y \in \{1,0\}$ = slope units yang terjadi longsor (1) atau tidak (0).

$p \in [0,1]$ = probabilitas hasil prediksi *space-time* MLP

Multi-Layer Perceptron

MLP Regression

- MLP regresi dapat digunakan untuk memprediksi intensitas berupa luas atau volume longsor berdasarkan faktor lingkungan.
- Lapisan output menggunakan fungsi identitas (aktivasi linear)
- Fungsi Loss yang digunakan adalah **Mean Squared Error (MSE)** :

$$\mathcal{L}(y, \hat{y}) = \frac{(\hat{y} - y)^2}{N}$$

dimana,

$\hat{y}_i$ = prediksi luas intensitas longsor oleh MLP

N = jumlah sampel training

Fitting Model

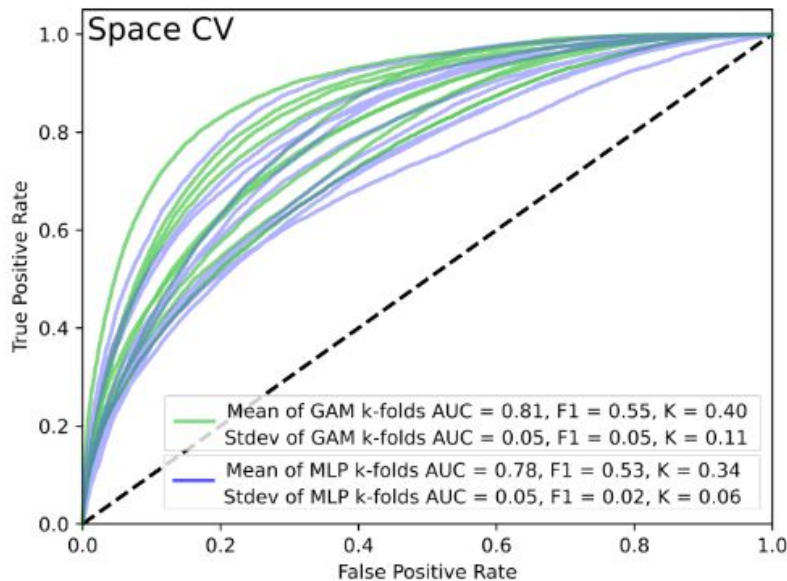
- Tahap pengujian model dalam sekali iterasi ($k = 1$) bertujuan melihat kesesuaian model dengan seluruh parameter input sebagai *training data* (100% data) agar dapat memprediksi atau mengklasifikasikan kerentanan tanah longsor.
- Hasilnya dengan metode klasifikasi – prediksi kerentanan :
 - *Receiving Operative Characteristic (ROC) curve, Area Under Curve (AUC), F1-score, dan Cohen's Kappa.*
- Hasilnya dengan metode regresi - prediksi intensitas :
 - *Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), R^2 , Koefisien Korelasi Pearson.*

Cross-Validation

- Metode untuk mengevaluasi kinerja model dengan membagi data menjadi beberapa subset (*folds*), di mana model dilatih pada sebagian data dan diuji pada bagian lainnya. Metode ini membantu menghindari *overfitting* dan memberikan estimasi akurat tentang performa model.
- Jenis *cross-validation* :
 - **Standard K-Fold Cross-Validation** : Dataset secara acak dibagi menjadi k bagian (lipatan atau folds) yang ukurannya kira-kira sama. Model dilatih sebanyak k kali. Setiap kali, satu lipatan digunakan sebagai set validasi, dan k-1 lipatan lainnya digunakan sebagai set pelatihan.
 - **Spatial Cross-Validation** : Pembagian dataset dilakukan berdasarkan lokasi geografis. Tujuannya adalah memastikan bahwa set pelatihan dan set validasi terpisah secara geografis
 - **Temporal Cross-Validation** : Metode ini menangani dependensi temporal. Pembagian data harus menggunakan urutan waktu. Data masa lalu digunakan untuk melatih model, dan data masa depan digunakan untuk memvalidasinya.
 - **Spatio-Temporal Cross-Validation** : Metode yang paling kompleks karena harus mengatasi kedua dependensi, spasial dan temporal, secara bersamaan. Tujuannya adalah memastikan independensi spasial dan temporal antara set pelatihan dan validasi

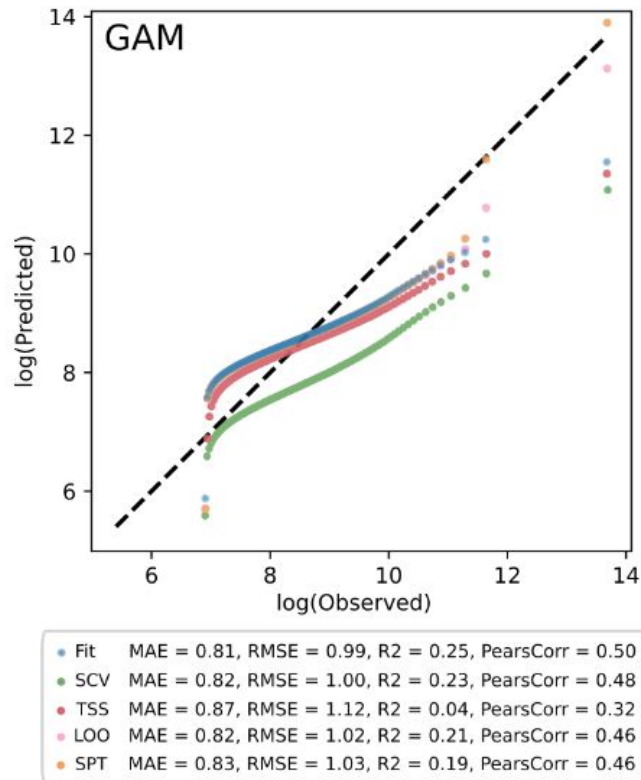
Interpretasi Hasil Kerentanan (Metode Klasifikasi)

Grafik ROC



- Grafik yang memvisualisasikan kemampuan sebuah model klasifikasi biner untuk membedakan antara kelas positif dan negatif pada berbagai threshold (ambang batas) klasifikasi.
- **AUC** : Nilai tunggal yang mengukur keseluruhan kemampuan diskriminatif model klasifikasi biner, dihitung sebagai luas area di bawah kurva ROC.
 - Nilai AUC memiliki rentang 0 - 1. Semakin tinggi nilai AUC (mendekati 1), semakin baik model secara keseluruhan dalam memisahkan kelas positif dan negatif, terlepas dari threshold yang dipilih.
- **F1-Score** adalah metrik yang menggabungkan Precision dan Recall dalam satu nilai harmonik dengan rentang 0 - 1. F1 tinggi (1.0) berarti model memiliki keseimbangan antara Precision dan Recall.
- **Cohen's Kappa (K)** berfungsi menilai akurasi model dengan mempertimbangkan tebakan acak. Rentang 0 - 1, semakin mendekati 1 semakin baik kinerja hasil model terhadap data sebenarnya.

QQ-Plots



Interpretasi Hasil Intensitas (Metode Regresi)

- Grafik yang digunakan untuk membandingkan distribusi dua set data, biasanya untuk memeriksa apakah data sampel (misalnya, residual/error model) berasal dari distribusi teoritis tertentu (seringkali distribusi normal).
- **RMSE (Root Mean Squared Error):** Memberikan gambaran rata-rata besarnya error dalam unit yang sama dengan variabel target. Lebih sensitif terhadap error besar. Nilai lebih rendah lebih baik.
- **MAE (Mean Absolute Error):** Memberikan gambaran rata-rata besarnya error, kurang sensitif terhadap *outlier* dibandingkan RMSE. Nilai lebih rendah lebih baik.
- **R² (Koefisien Determinasi):** Proporsi varians dalam variabel dependen yang dapat dijelaskan oleh variabel independen dalam model. Berkisar antara 0 dan 1 (atau bisa negatif untuk model yang sangat buruk). Nilai lebih tinggi lebih baik (mendekati 1).
- **Korelasi Pearson** mengukur hubungan linear antara dua variabel kontinu. Nilai -1 sampai 1.

Sumber

- ¹ Titti G, Sarretta A, Lombardo L, Crema S, Pasuto A and Borgatti L (2022) Mapping Susceptibility With Open-Source Tools: A New Plugin for QGIS. Front. Earth Sci. 10:842425. doi: 10.3389/feart.2022.842425.
- ² Titti G, Hu L, Festi P, Elia L, Borgatti L, Lombardo L. 2025. An updated version of the SZ-plugin: from space to space-time data-driven modeling in QGIS. EarthArxiv, Preprint. <https://doi.org/10.31223/X5JD9X>

TERIMA KASIH