ISSN 0125-1790 (print), ISSN 2540-945X (online)
Majalah Geografi Indonesia Vol 38, No 2 (2024) : 169-180
DOI: 10.22146/mgi.95857
©2024 Fakultas Geografi UGM dan Ikatan Geograf Indonesia (IGI)

MAJALAH GEOGRAFI INDONESIA

NASKAH PENELITIAN

# Alur kerja pembelajaran mesin pada pemodelan spasial kerawanan longsor

**Guruh Samodra**[1*]
[1]*Departemen Geografi Lingkungan, Fakultas Geografi , Universitas Gadjah Mada, Yogyakarta, Indonesia*
** Email: guruh.samodra@ugm.ac.id*

Submitted: 2024-05-02 Revisions: 2024-09-11 Accepted: 2024-09-27 Published: 2024-09-30
*©2024 Fakultas Geografi UGM dan Ikatan Geograf Indonesia (IGI)*

©2024 by the authors and Majalah Geografi Indonesia.
This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution(CC BY SA) https://creativecommons.org/licenses/by-sa/4.0/

**Abstrak** Salah satu instrumen pengurangan risiko bencana longsor adalah peta kerawanan longsor yang dihasilkan dari pemodelan spasial. Alur kerja pemodelan spasial kerawanan longsor menggunakan model pembelajaran mesin belum terakomodasi dalam Standar Nasional Indonesia (SNI) yang berlaku saat ini (tahun 2024). Penelitian ini berusaha menjelaskan variasi langkah-langkah dalam alur kerja pembelajaran mesin dan menunjukkan perbedaannya dengan alur kerja model statistik. Model statistik regresi logistik dan model pembelajaran mesin *random forest* dipilih untuk menjelaskan perbedaan alur kerja pemodelan spasial kerawanan longsor. Formulasi alur kerja pemodelan spasial kerawanan longsor diterapkan untuk memetakan kerawanan longsor di Kabupaten Pacitan. Pada tanggal 27-29 November 2017, 743 longsor terjadi di Kabupaten Pacitan dipicu oleh hujan yang sangat lebat akibat Siklon Tropis Cempaka. Kabupaten Pacitan merupakan salah satu wilayah rawan longsor di Provinsi Jawa Timur. Alur kerja pemodelan spasial kerawanan longsor terbagi atas beberapa langkah yaitu penyiapan data, pra-pemrosesan data (*pre-processing*), melatih dan menyetel model, memvalidasi model, pemodelan spasial, dan uji akurasi. Hasil uji akurasi model RF dan LR yang diterapkan di Kabupaten Pacitan masing-masing sebesar 0,75 dan 0,73. Penelitian ini diharapkan dapat memberikan masukan dalam penyusunan SNI pemetaan kerawanan longsor di masa mendatang serta dapat digunakan sebagai acuan dalam pemetaan kerawanan longsor secara umum di Indonesia.

**Kata kunci:** kerawanan, longsor, pembelajaran mesin, pemodelan spasial, alur kerja

***Abstract*** *One of the landslide risk reduction instruments is landslide susceptibility maps which can be produced by spatial modeling. The landslide susceptibility modeling based on machine learning workflows have not been accommodated in the current verison of Indonesian National Standard (SNI). This study seeks to explain the variation of machine learning workflows and show how they differ from statistical learning workflows. Logistic regression model and random forest machine learning models were selected to explain variations in landslide susceptibility modeling workflows. The modeling workflows were applied to map landslide susceptibility in Pacitan Regency. On 27-29 November 2017, 743 landslides occurred in Pacitan Regency triggered by very heavy rain due to Tropical Cyclone Cempaka. Pacitan Regency is one of the landslide-prone areas in East Java Province. The landslide susceptibility modeling workflow is divided into several steps, i.e. data preparation, data pre-processing, training and tuning the model, validating the model, spatial modeling, and accuracy testing. The accuracy test results of the RF and LR models applied in Pacitan Regency were 0.75 and 0.73 respectively. This research is expected to provide a benchmark for landslide susceptibility mapping in Indonesia.*

***Keywords:*** *landslide, susceptibility, machine learning, spatial modeling, workflow*

## PENDAHULUAN

Indonesia merupakan wilayah yang memiliki risiko longsor tinggi. Bertambahnya penduduk yang bermukim di daerah perbukitan dapat meningkatkan jumlah kejadian risiko longsor. Jumlah kejadian longsor dan kematian akibat longsor di Indonesia dari tahun 2004 hingga 2023 tercatat sejumlah 9349 kejadian longsor dan 2774 korban meninggal akibat longsor (https://dibi.bnpb.go.id/, 2024). Rata-rata jumlah kejadian longsor dan korban meninggal akibat longsor adalah 467 kejadian/tahun dan 138 jiwa/tahun. Pengurangan risiko longsor perlu dilakukan mengingat banyaknya korban jiwa yang diakibatkan oleh kejadian longsor. Upaya pengurangan risiko longsor dapat dilakukan melalui berbagai macam tindakan mitigasi seperti perlindungan elemen berisiko dengan struktur fisik, pembatasan pengembangan bangunan/infrastruktur di daerah rawan, penegakan aturan, serta pengembangan sistem peringatan dini (Schuster & Kockelman, 1996). Salah satu instrumen tindakan mitigasi pembatasan pengembangan bangunan/infrastruktur di daerah rawan adalah peta kerawanan longsor (Fell et al., 2008).

Peta kerawanan longsor merupakan peta yang menunjukkan kemungkinan terjadinya longsor di suatu lokasi tertentu (Corominas & Moya, 2008). Peta kerawanan (*susceptibility*) hanya berisi tentang prediksi lokasi atau dengan kata lain hanya dapat menjawab pertanyaan di mana kemungkinan longsor akan terjadi. Berbagai macam teknik pemodelan spasial telah banyak diterapkan untuk pembuatan peta kerawanan longsor. Secara umum, teknik pemodelan spasial kerawanan longsor dapat dibagi menjadi heuristik (*knowledge driven*), statistik (*data driven*), dan deterministic/fisik (Samodra et al., 2017; Soeters & Westen, 1996). Indonesia mengadopsi ketiga teknik pemodelan spasial tersebut dan tertuang dalam dokumen SNI 8291:2016 tentang penyusunan dan penentuan zona kerentanan gerakan tanah (Badan

Standardisasi Nasional, 2016). Teknik pemodelan spasial yang paling sesuai untuk wilayah luas dalam skala menengah hingga kecil adalah dengan menggunakan model statistik (Cascini, 2008; Soeters & Westen, 1996; van Westen et al., 2008).

Prosedur pemetaan kerawanan longsor menggunakan model statistik berdasarkan Standar Nasional Indonesia (SNI) adalah dengan mengekstrak informasi setiap faktor pengontrol longsor (peta faktor pengontrol) dari peta sebaran longsor, memodelkan dengan model statistik, dan menyajikan dalam bentuk peta kerawanan. Akurasi pemodelan tidak dibahas secara eksplisit dalam SNI tersebut. Model statistik dalam pemodelan spasial kerawanan longsor diantaranya adalah *frequency ratio*, *weight of evidence*, *discriminant analysis*, dan regresi logistik. Saat ini, model statistik telah dikembangkan menjadi model pembelajaran mesin (*machine learning*). Kemajuan teknologi komputer dan Sistem Informasi Geografis (SIG) turut memiliki andil dalam pengembangan model statistik menjadi model pembelajaran mesin (*machine learning*) dalam pemetaan kerawanan longsor.

Prosedur pemetaan kerawanan longsor menggunakan model pembelajaran mesin cukup bervariasi dibandingkan dengan model statistik. Terapan model pembelajaran mesin untuk pemetaan kerawanan longsor memerlukan beberapa langkah yang sering disebut sebgai alur kerja pembelajaran mesin (*machine learning workflow*). Alur kerja dalam pemodelan kerawanan longsor berbasis pembelajaran mesin secara umum dapat dibagi menjadi beberapa langkah yaitu mengekstrak informasi setiap faktor pengontrol longsor dari sampel titik/area tanah longsor dan non-longsor (Lovelace et al., 2020), memodelkan dengan metode tertentu, dan menguji akurasi. Variasi setiap langkah dalam alur kerja pembelajaran mesin seperti cara inventarisasi longsor, cara pemodelan, dan cara menguji akurasi dapat menyebabkan perbedaan hasil peta kerawanan longsor. Sebagai contoh, penanganan data inventarisasi longsor, yaitu penempatan sampel longsor pada poligon longsor dan jumlah sampel, dapat mempengaruhi tampilan dan akurasi peta kerawanan longsor (Abraham et al., 2021; Hussin et al., 2016; Steger et al., 2016). Model pembelajaran mesin juga memiliki banyak variasi seperti *decision tree*, *random forest*, *artificial neural network*, *XGBoost,* dll. Variasi langkah-langkah dalam alur kerja pembelajaran mesin dapat dipilih dan disusun untuk menghasilkan peta kerawanan longsor yang memiliki akurasi tinggi.

Alur kerja pemetaan kerawanan longsor menggunakan model pembelajaran mesin belum terakomodasi dalam SNI tentang pemetaan atau zonasi kerawanan longsor. Penelitian ini berusaha menjelaskan variasi langkah-langkah dalam alur kerja pembelajaran mesin dan menunjukkan perbedaannya dengan alur kerja model statistik. Model statistik regresi logistik (LR) dan model pembelajaran mesin *random forest* (RF) dipilih untuk menjelaskan alur kerja pemodelan spasial kerawanan longsor. Formulasi alur kerja pemodelan spasial kerawanan longsor diterapkan untuk memetakan kerawanan longsor di Kabupaten Pacitan yang pernah mengalami kejadian longsor dengan jumlah banyak akibat Siklon Tropis Cempaka. Keberadaan data inventarisasi longsor yang lengkap di Kabupaten Pacitan akibat Siklon Tropis cempaka menjadi alasan dipilihnya Kabupaten Pacitan sebagai area studi pada penelitian ini. Penelitian ini diharapkan dapat memberikan masukan dalam penyusunan SNI pemetaan kerawanan longsor di masa mendatang serta dapat digunakan sebagai acuan dalam pemetaan kerawanan longsor secara umum di Indonesia.

## METODE PENELITIAN

### Daerah penelitian

Alur kerja pembelajaran mesin untuk pemodelan spasial kerawanan longsor diterapkan untuk wilayah Kabupaten Pacitan, Provinsi Jawa Timur (Gambar 1). Luas wilayah Kabupaten Pacitan adalah 1.390 km$^2$ dan didominasi oleh morfologi perbukitan hingga pegunungan. Morfologi datar dengan ketinggian 0 – 10 mdpal dapat ditemui di sisi selatan Kabupaten Pacitan dan merupakan teluk sempit yang diapit oleh morfologi perbukitan di sisi timur dan barat. Relief kasar dengan lereng 15° hingga 70° dapat dijumpai pada lembah sungai Girindulu yang mengalami penorehan yang besar akibat proses pelapukan, erosi, dan longsor.

Tatanan geomorfologi Kabupaten Pacitan merupakan hasil dari empat fase utama yaitu tahap awal perkembangan busur vulkanisme selama Eosen Tengah hingga Oligosen Awal, pertumbuhan dan penghentian busur vulkanisme selama Oligosen akhir hingga Miosen Awal, pertumbuhan karbonat disertai dengan erosi dan pengendapan endapan vulkaniklastik berumur Miosen tengah, dan yang terakhir adalah proses pengangkatan diikuti oleh erosi dan pengendapan selama Pliosen hingga sekarang (Smyth et al., 2008). Fase terakhir mengakibatkan pelapukan, erosi, dan longsor menjadi proses dominan yang terjadi di permukaan lahan saat ini.

Pada tanggal 27-29 November 2017, Kabupaten Pacitan dilanda hujan yang sangat lebat akibat Siklon Tropis Cempaka. Hujan kumulatif sebesar 235,97 mm selama satu setengah hari memicu 743 kejadian longsor di Kabupaten Pacitan (Samodra et al., 2020). Data 743 kejadian longsor tersebut digunakan dalam penelitian ini untuk pemodelan spasial kerawanan longsor.

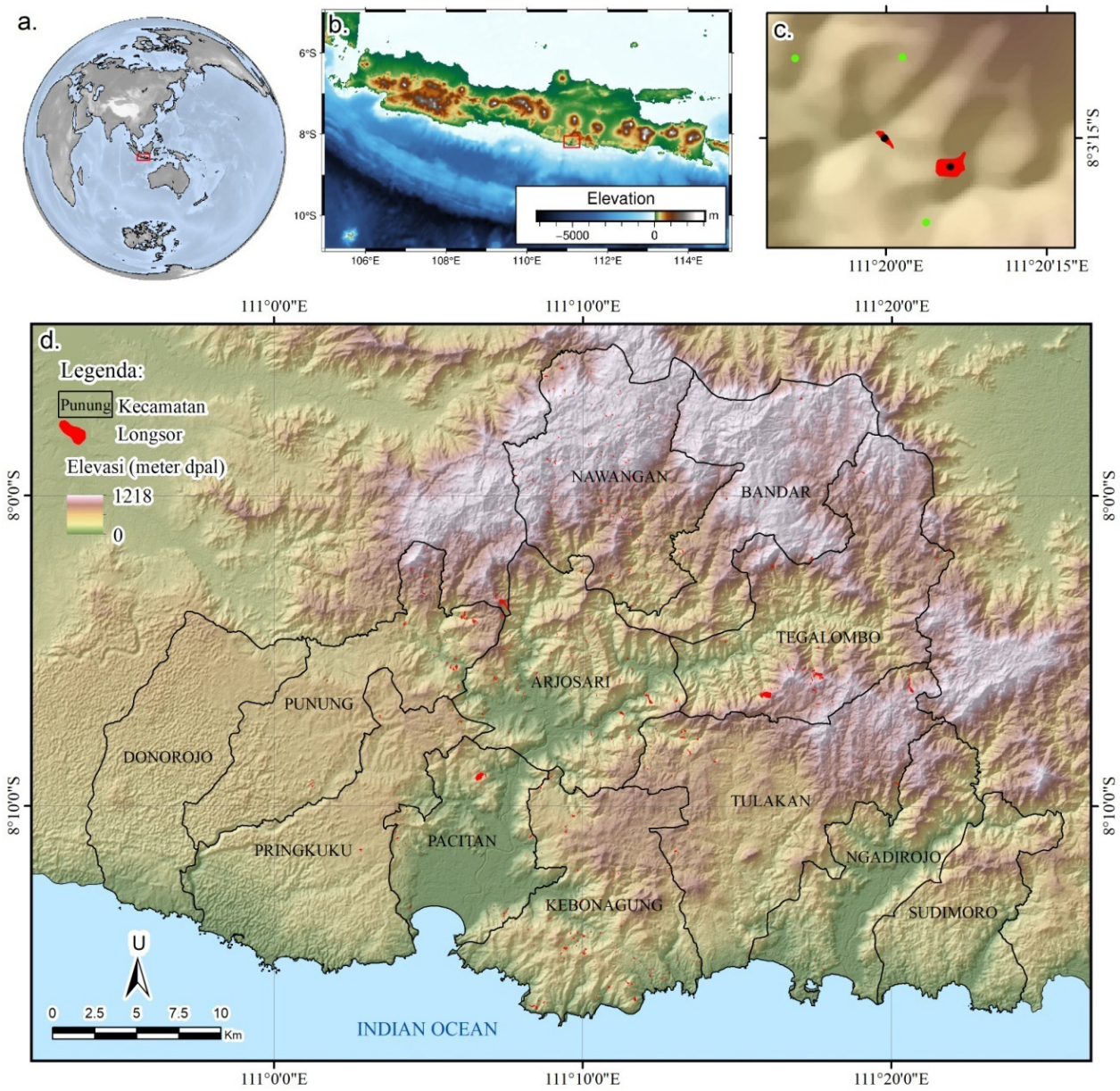
### Pengumpulan data

Pengumpulan data penelitian meliputi data primer dan data sekunder. Pengumpulan data primer yaitu melakukan pengambilan foto udara dan survei lapangan. Pengambilan foto udara dilakukan menggunakan drone *UAV* untuk inventarisasi longsor. Survei lapangan dilakukan untuk memperoleh informasi tambahan mengenai tipologi longsor, aktivitas longsor, dan kondisi lingkungan lokasi penelitian (Samodra et al., 2020).

Pengumpulan data sekunder meliputi studi literatur pada *database open access* dan penghimpunan data geospasial pada *website InaGeoportal*. Pengumpulan data geospasial terdiri dari data spasial vector Rupa Bumi Indonesia (RBI), dan data raster *digital elevation model* (DEM) yaitu *digital elevation model* nasional (DEMNAS) yang disediakan oleh Badan Informasi Geospasial (BIG).

Peta faktor pengontrol longsor berasal dari *National Digital Elevation Model of Indonesia* (DEMNAS) dengan resolusi spasial sekitar 8 m, yang tersedia di https://tanahair.indonesia.go.id/demnas/#/, dan peta rupabumi digital pada skala 1:25.000 yang tersedia di https://tanahair.indonesia.go.id/portal-web. Dalam penelitian ini, faktor pengontrol longsor diwakili oleh atribut medan digital, seperti kemiringan lereng, aspek, jarak ke sungai, jarak ke jalan, ketinggian, penggunaan lahan, plan curvature, profile curvature, indeks kekuatan aliran (SPI), dan indeks kebasahan medan (TWI) (Gambar 2). Atribut medan digital diproses dari DEMNAS dengan System for Automated Geoscientific Analysis (SAGA) GIS (Conrad et al., 2015).

Signifikansi faktor pengontrol terhadap longsor, secara hipotetis, dapat dijelaskan sebagai berikut. Elevasi

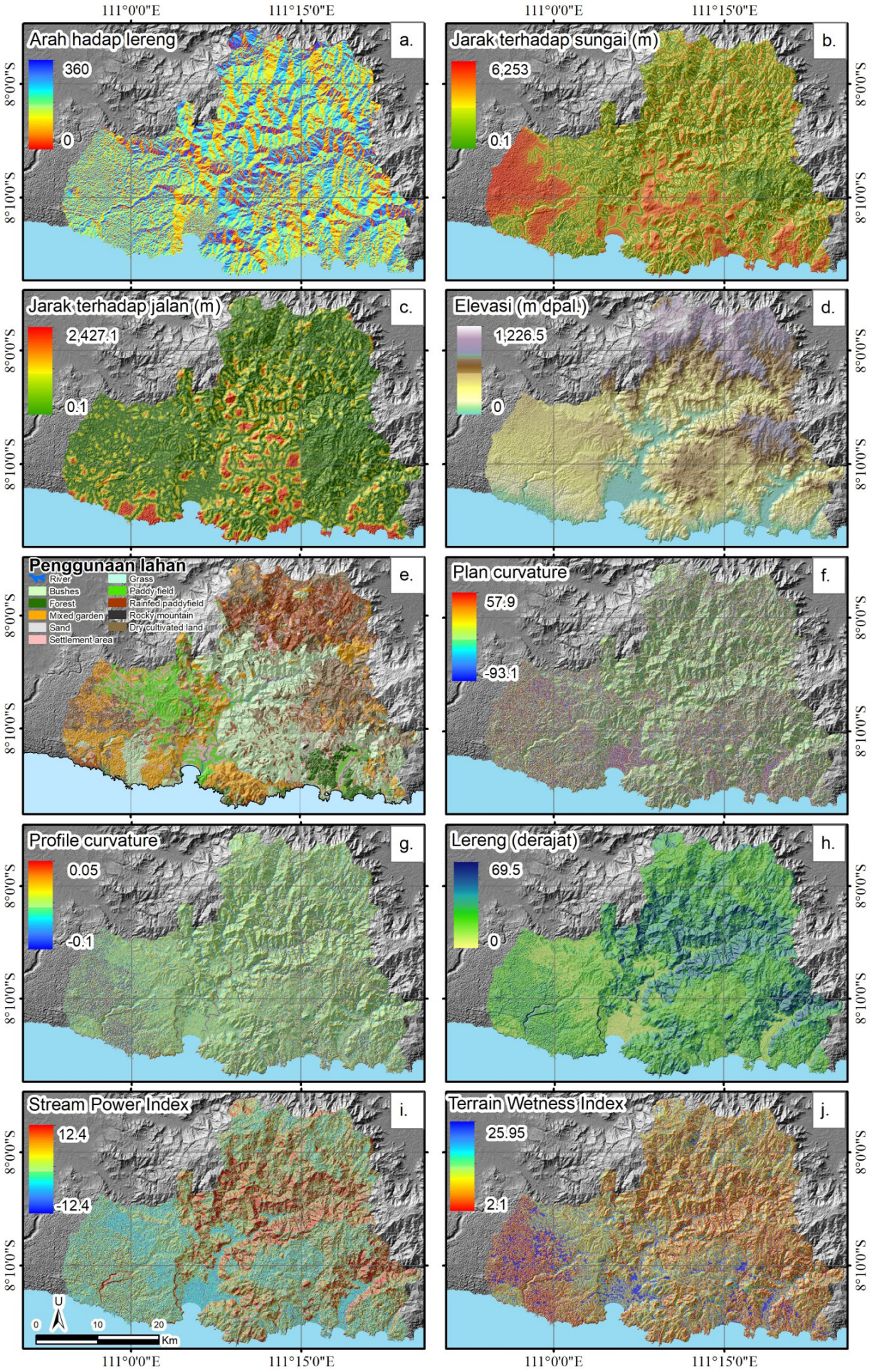



**Gambar 1.** Lokasi penelitian, kotak merah menunjukkan Pulau Jawa (a), Elevasi Pulau jawa (kotak merah menunjukan lokasi Kabupaten Pacitan) (b), area longsor hasil inventarisasi (titik hitam merupakan titik *centroid* area longsor dan titik hijau merupakan sampel titik non-longsor (c), wilayah penelitian Kabupaten Pacitan (d) (data elevasi diunduh dari https://tanahair.indonesia.go.id/demnas/#/)

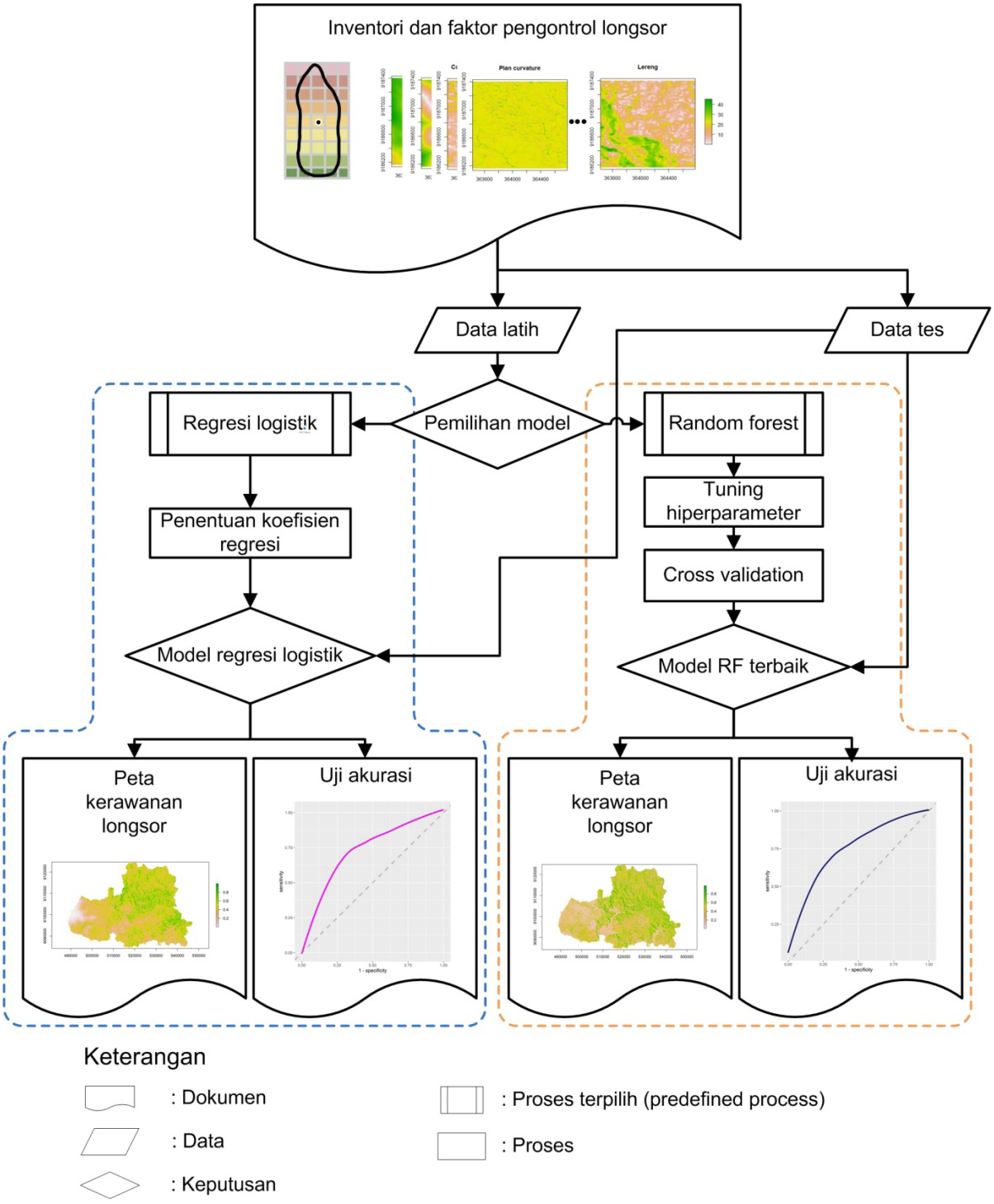
(Gambar 2d) merupakan representasi kondisi topografi dan menggambarkan relief lokal yang menunjukkan posisi ketinggian minimum dan maksimum di atas permukaan laut. Kemiringan lereng (Gambar 2h) merepresentasikan hubungan keseimbangan antara kuat geser dan tegangan geser yang bekerja pada sebuah lereng. Arah hadap lereng (Gambar 2a) mencerminkan perbedaan tingkat pelapukan dan kelembaban tanah yang berkaitan dengan insolasi matahari. *Curvature* (Gambar 2f dan 2h) mencerminkan bentuk lereng yang dapat mempengaruhi arah aliran permukaan dan arah aliran air tanah. *Stream Power Index* (SPI) (Gambar 2i) mencerminkan kekuatan aliran yang dihitung dari luas daerah tangkapan air dan kecuraman lereng. Daerah dengan tangkapan hujan yang luas dan lereng yang curam akan menghasilkan kekuatan aliran yang tinggi (jumlah air yang banyak dan kecepatan aliran yang tinggi). *Terrain Wetness Index* (TWI) (gambar 2j) mencerminkan kelembaban tanah atau kecenderungan suatu tempat untuk mengakumulasi air (Beven & Kirkby, 1979; Moore et al., 1991; Quinn et al., 1991). Jarak terhadap sungai (Gambar 2b) sebagai faktor pengontrol mengasumsikan bahwa tempat yang lebih dekat terhadap sungai akan menyebabkan lebih banyak longsor karena lereng yang curam dan erosi. Penggunaan lahan (Gambar 2e) dan jarak terhadap jalan (gambar 2c) sebagai pengontrol longsor mengasumsikan bahwa aktivitas manusia dapat meningkatkan ketidakstabilan lereng.

### Metode analisis data

Alur kerja pemodelan spasial kerawanan longsor dapat dibagi menjadi beberapa langkah yang lebih rinci yaitu penyiapan data, pra-pemrosesan data (*pre-processing*), melatih dan menyetel model, memvalidasi model, pemodelan spasial, dan uji akurasi (Gambar 3). Penyiapan data dapat dilakukan dengan pengumpulan data primer maupun sekunder seperti yang telah dijelaskan pada sub bab pengumpulan data.




**Gambar 2.** Faktor-faktor pengontrol longsor: arah hadap lereng (a), jarak terhadap sungai (b), jarak terhadap jalan (c), elevasi (d), penggunaan lahan (e), *plan curvature* (f), *profile curvature* (g), lereng (h), *stream power index* (i), dan *terrain wetness index* (j)




**Gambar 3.** Diagram alir alur kerja pemodelan spasial kerawanan longsor statistik (garis putus-putus biru) dan pembelajaran mesin (garis putus-putus oranye)

Pra-pemrosesan merupakan tahapan untuk menyiapkan data agar dapat diolah termasuk memberi nilai atribut inventori longsor dan faktor pengontrol longsor. Titik longsor dan non-longsor sejumlah masing-masing 743 non-longsor dipersiapkan dalam penelitian ini. Sampel titik longsor pada penelitian ini diletakkan pada titik tengah poligon longsor sedangkan titik non-longsor dibuat secara acak diluar poligon longsor. Titik longsor diberikan atribut "Yes" dan titik non-longsor diberi atribut "No". Keduanya dikategorikan sebagai tipe data *Factors* yang memiliki level sehingga model yang digunakan dalam penelitian ini adalah model klasifikasi. Semua faktor pengontrol longsor diperoleh dari peta raster dan memiliki atribut sebagai data kontinyu kecuali data raster penggunaan lahan. Data penggunaan lahan merupakan data spasial vektor diskrit yang dikonversi menjadi data raster dan setiap tipe penggunaan lahan diberikan atribut bilangan asli. Setelah semua data diberikan atribut, proses selanjutnya adalah proses tabulasi silang antara data longsor dan non-longsor dengan faktor pengontrol longsor. Selanjutnya, data tabulasi silang dibagi menjadi data latih dan data tes sebanyak 75% (longsor, non-longsor) dan 25% (longsor, non-longsor). Data latih digunakan untuk pembuatan model sedangkan data tes untuk menguji akurasi model.

Tahap melatih dan menyetel model adalah tahap pembuatan model menggunakan data training. Khusus untuk model pembelajaran mesin yang memiliki hiperparameter, pembuatan model memerlukan penyetelan hiperparameter untuk memilih model yang paling baik. Model regresi

yang digunakan dalam penelitian ini tidak memiliki nilai hiperparameter sedangkan model random forest memiliki hiperparameter mtry dan min_n. Proses pemilihan model random forest terbaik dengan melibatkan penyetelan parameter yang optimal sering disebut sebagai proses validasi (*validation*). Pada penelitian ini penyetelan hiperparameter dilakukan dengan menggunakan validasi silang 10-fold berdasarkan blok spasial (10-fold *spatial block cross validation*). Model terbaik digunakan untuk memprediksi nilai probabilitas spasial kerawanan longsor menggunakan 10 data spasial faktor pengontrol. Model terbaik diuji akurasinya menggunakan data tes yang merupakan data independen yang tidak digunakan dalam pembuatan model.

Model yang digunakan dalam penelitian ini adalah model statistik (regresi logistik) dan pembelajaran mesin (random forest). Dua model tersebut dipilih untuk menunjukkan perbedaan alur kerja keduanya. Model regresi logistik menghubungkan kejadian longsor sebagai sebagai variabel independen dengan faktor-faktor pengontrol longsor sebagai variabel dependen (Samodra, 2023). Probabilitas terjadinya longsor dihitung menggunakan estimasi *maximum likelihood* dengan cara mentransformasi variabel dependen menjadi variabel logit dengan persamaan sebagai berikut:

$$f(z) = \pi(S = 1|X_1, X_2, \dots . X_n) = \frac{1}{1+e^{-(\beta_0+\sum_{i=1}^{n}(\beta_i X_i)}} \quad (1)$$

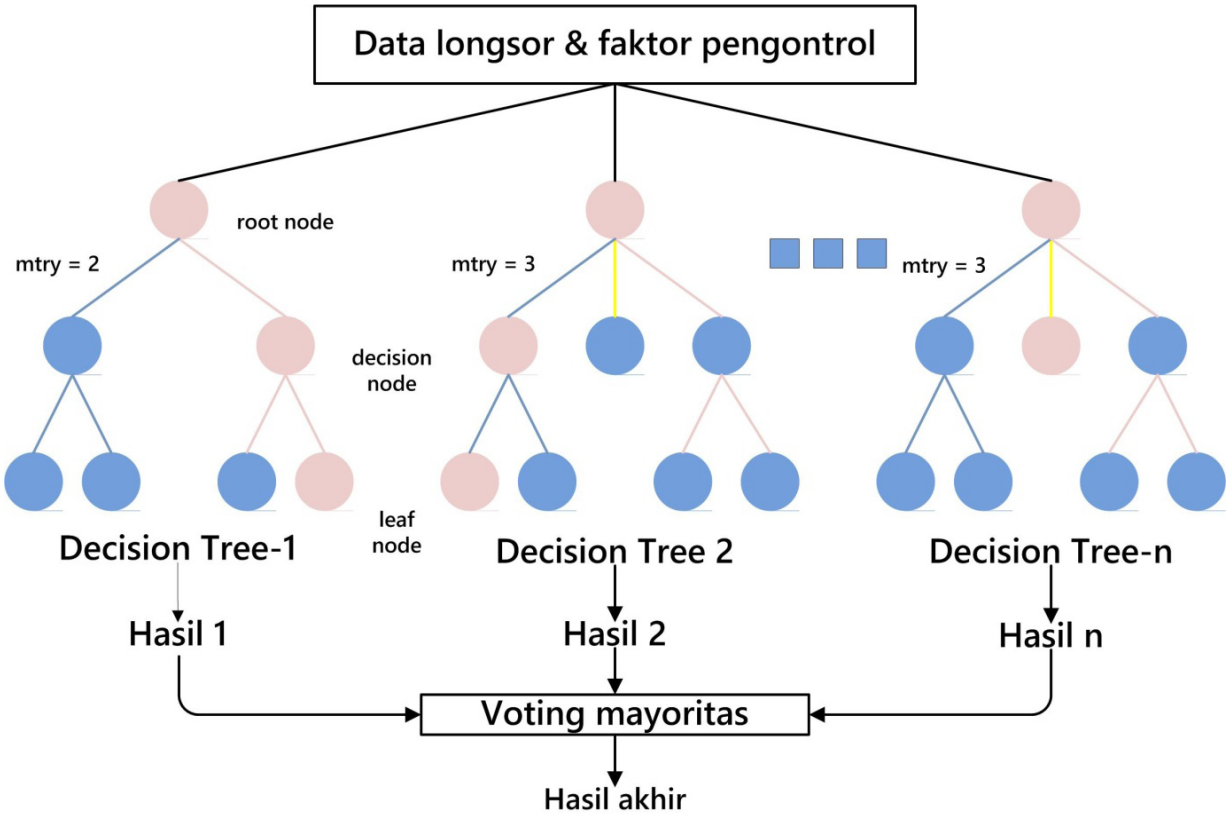
dimana $\pi(S = 1|X_1, X_2, \dots . X_n)$ merupakan sebuah piksel yang diprediksi apakah akan terjadi longsor atau tidak yang dipengaruhi oleh variabel independen terhadap faktor pengontrol longsor dari $X_1$ ke $X_n$. $\beta_0$ adalah konstanta dari persamaan, dan $\beta_i$, $\beta_2$, $\beta_n$merupakan koefisien yang tidak diketahui dan akan diestimasi berdasarkan variabel independen menggunakan teknik *maximum likelihood*.

Random forest merupakan salah satu algoritma pembelajaran mesin yang dibangun dengan cara membuat *decision tree* secara acak dan memilih beberapa *decision tree* untuk tujuan prediksi ataupun klasifikasi (Breiman, 2001). Algoritma *random forest* diawali dengan cara membuat sejumlah sampel *tree* ($n_{tree}$) (titik longsor dan piksel faktor-faktor pengontror longsor yang berada pada longsor), menumbuhkan pohon klasifikasi dengan cara mengambil sampel secara acak faktor pengontrol sebagai kandidat untuk setiap *split* ($m_{try}$) dan memilih split terbaik. Penumbuhan pohon klasifikasi akan berhenti sampai *leaf node* apabila sudah mencapai jumlah minimum observasi (min_n) yang telah ditentukan (Probst et al., 2019). Selanjutnya adalah memprediksi data baru dengan mengagregasikan prediksi dari sejumlah sampel *tree* ($n_{tree}$) dengan cara suara mayoritas (*majority vote*) (Liaw & Wiener, 2002) (Gambar 4). Nilai min_n dan $m_{try}$ merupakan hiperparameter yang perlu dilakukan penyetelan (*tuning*) untuk memperoleh hasil prediksi yang optimal. Kedua model dalam penelitian ini dijalankan menggunakan bahasa pemrograman R (R Core Team, 2021) dengan package utama Tidymodels (Kuhn & Wickham, 2020).

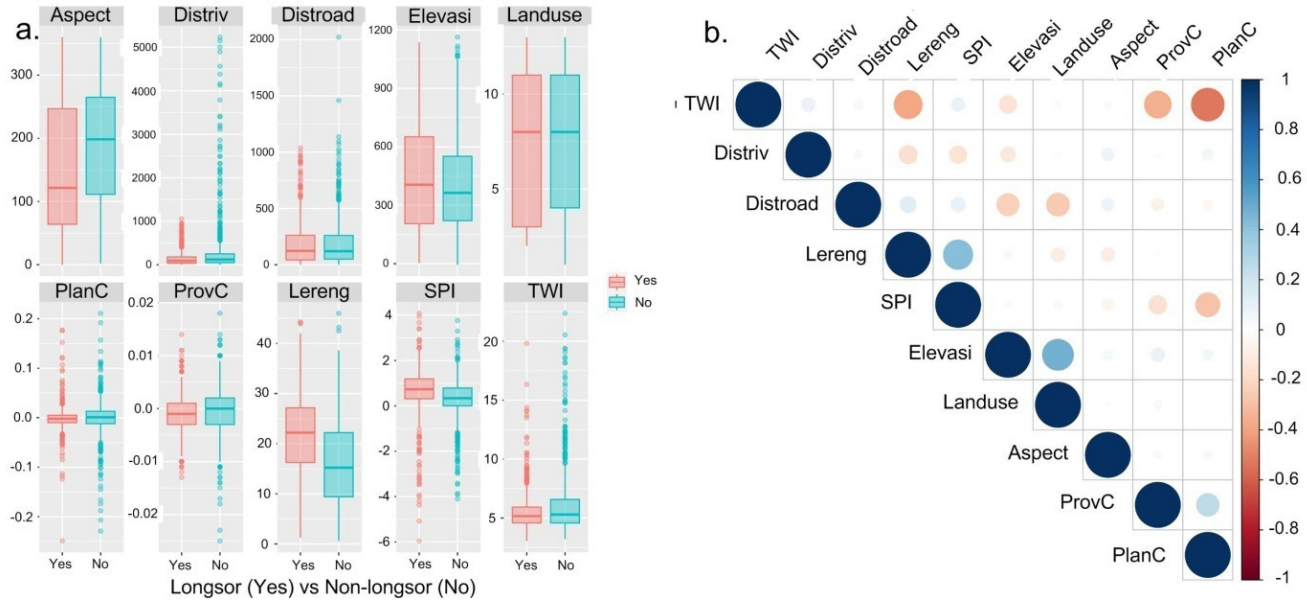
## HASIL DAN PEMBAHASAN

### Pra pemrosesan data

Pemilihan variabel pengontrol longsor ditentukan secara subyektif berdasarkan pengalaman peneliti. Variabel-variabel tersebut sebisa mungkin tidak berkorelasi satu sama lain sehingga dapat memprediksi dengan baik kejadian longsor. Box-plot faktor pengontrol longsor yang telah dibedakan berdasarkan data longsor dan non-longsor (Gambar 5a) menunjukkan bahwa sebaran masing-masing faktor pengontrol longsor baik longsor dan non-longsor cukup berbeda signifikan. Korelasi antar faktor pengontrol longsor juga cukup kecil (Gambar 5b) sehingga faktor-




**Gambar 4.** Diagram cara kerja Random Forest




**Gambar 5.** Box-plot data eksplorasi (a), plot korelasi (b)

**Tabel 1.** Nilai konstanta dan koefisien regresi logistik, *Tolerance* dan *Variance Inflation Factor* (VIF)

| Faktor Pengontrol | Koefisien | Tolerance | VIF |
|---|---|---|---|
| Intercept | -0,3477 | - | - |
| Elevasi | -0,0003 | 0,73 | 1,38 |
| Lereng | -0,0519 | 0.60 | 1,66 |
| Aspect | 0,0035 | 0,97 | 1,03 |
| Profile curvature | 10,8331 | 0,85 | 1,18 |
| Stream Power Index | -0,0906 | 0,71 | 1,4 |
| Terrain Wetness Index | 0,0763 | 0,50 | 2,01 |
| Penggunaan lahan | 0,0255 | 0,73 | 1,36 |
| Jarak terhadap jalan | 0,0004 | 0,88 | 1,14 |
| Plan curvature | 1,2571 | 0,64 | 1,56 |
| Jarak terhadap sungai | -0,3477 | 0,93 | 1,08 |

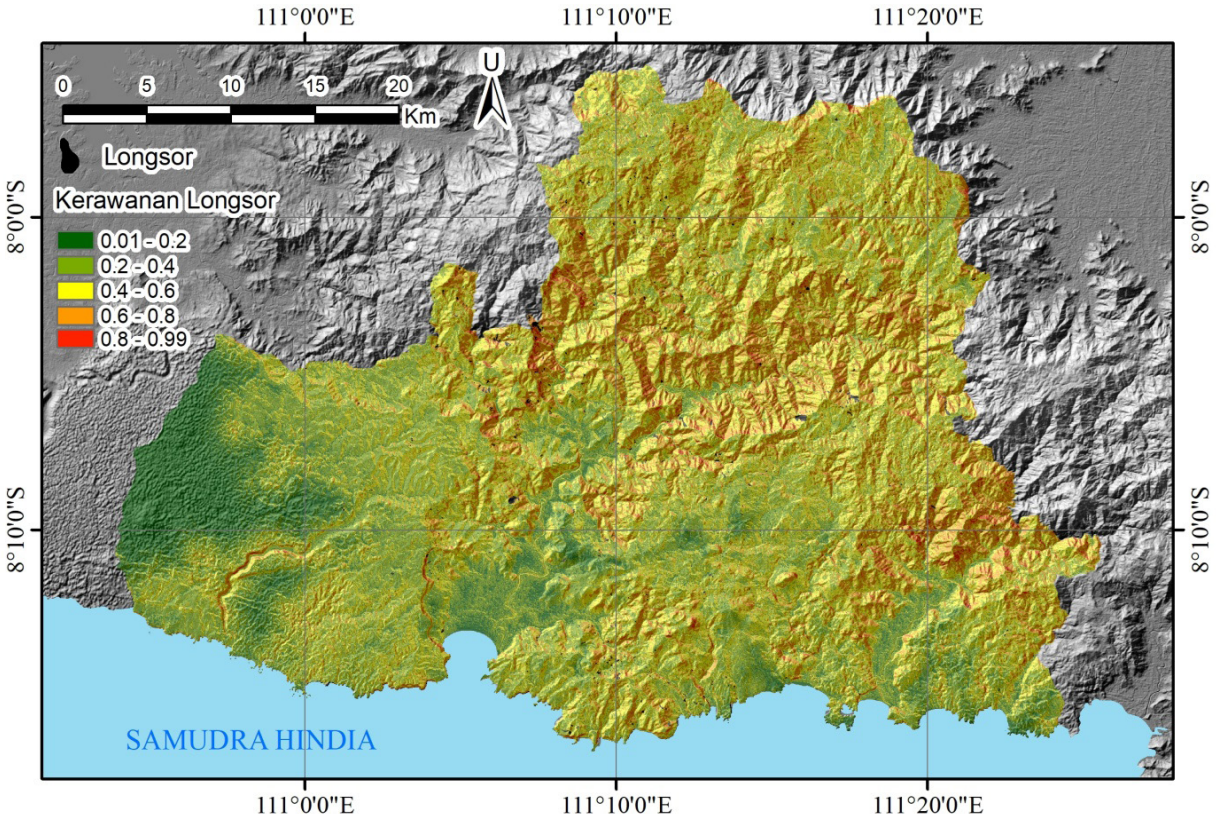
faktor pengontrol yang dipilih terindikasi cukup handal untuk memprediksi sebaran spasial potensi longsor. Analisis korelasi dapat digunakan untuk memastikan hanya variabel yang signifikan yang digunakan dalam model regresi logistik (Guyon & Elisseeff, 2003). Hal ini memungkinkan seleksi fitur dapat dilakukan secara langsung.

## Alur kerja regresi logistik

Alur kerja model regresi logistik lebih sederhana daripada model pembelajaran mesin karena tidak memerlukan penyetelan hiper-parameter. Model regresi logistik dapat secara langsung menentukan konstanta *intercept* dan koefisien regresi untuk masing-masing faktor pengontrol longsor (Tabel 1.). Uji *Tolerance* dan *Variance Inflation Factor* (VIF) (uji multikolinearitas) dapat digunakan untuk menguji kolinearitas antara faktor pengontrol longsor. Faktor pengontrol dengan VIF < 4 (O'Brien, 2007) dan *Tolerance* > 0,2 (Menard, 2002) dapat dianggap bahwa variabel dependen tersebut memiliki korelasi yang rendah satu sama lain. Kesepuluh faktor pengontrol (Tabel 1) dapat digunakan untuk memodelkan kerawanan longsor.

Persamaan (1) dapat digunakan secara langsung untuk memodelkan spasial kerawanan longsor. Substitusi nilai-nilai faktor pengontrol yang diwujudkan dalam peta raster ke dalam persamaan (1) sebagai variabel dependen akan menghasilkan peta kerawanan longsor (Gambar 6). Nilai mendekati 0 mengindikasikan bahwa lokasi tersebut memiliki peluang kecil untuk terjadi longsor, sedangkan nilai mendekati 0 mengindikasikan bahwa lokasi tersebut memiliki peluang besar untuk terjadi longsor. Lima klasifikasi nilai kerawanan yang dibagi setara digunakan untuk membandingkan hasil akhir peta kerawanan longsor (Gambar 6). Persentase luas area kerawanan rendah (0,01-0,2) hingga tinggi (0,8-0,99) secara berurutan adalah 9,7%, 31,4%, 38,5%, 19,2%, dan 1,2%. Wilayah dengan nilai 0,4-0,6 adalah 38,5%. Semakin besar persentase luas wilayah dalam kelas 0,4-0,5 berarti bahwa model memiliki *discriminative power* yang rendah. *Discriminative power* merupakan kekuatan model untuk membedakan area yang memiliki tingkat kerawanan rendah dan kerawanan tinggi. Peta kerawanan longsor harus secara efektif dapat menyoroti area yang memiliki kerawanan tinggi sekaligus secara akurat menunjukkan area yang memiliki kerawanan rendah atau stabil.

Model regresi logistik menggunakan fungsi logit untuk menghubungkan variabel pengontrol longsor sebagai variabel independen dengan probabilitas kejadian (Menard, 2002). Dampak individual dari variabel longsor dapat langsung diinterpretasi berdasarkan nilai koefisien. Berdasarkan model




**Gambar 6.** kerawanan longsor Kab. Pacitan menggunakan teknik regresi logistik

regresi logistik di Pacitan, faktor yang berpengaruh dominan terhadap longsor adalah bentuk lereng (kecembungan dan kecekungan). Hal ini membuat model regresi logistik memiliki keunggulan dalam memahami faktor alamiah yang paling berpengaruh terhadap suatu peristiwa (Hastie et al., 2009). Interpretasi langsung keterkaitan variabel pengontrol dengan peristiwa longsor memudahkan pembuat kebijakan dan pengelola lahan mengambil keputusan dalam mengevaluasi kondisi lahan eksisting berkaitan dengan frekuensi kejadian longsor (Agresti, 2013). Meski demikian, ada beberapa keterbatasan model regresi logistik seperti adanya asumsi linearitas dalam logit. Variabel geospasial yang dipilih berdasarkan intuisi peneliti mungkin tidak berkaitan langsung dengan kejadian longsor (Goetz et al., 2015). Kinerja model juga mungkin berbeda antara satu tempat dengan tempat yang lain, sehingga keterbukaan alur kerja model sangat diperlukan untuk *benchmarking* di berbagai tempat dengan kondisi lingkungan yang berbeda-beda.

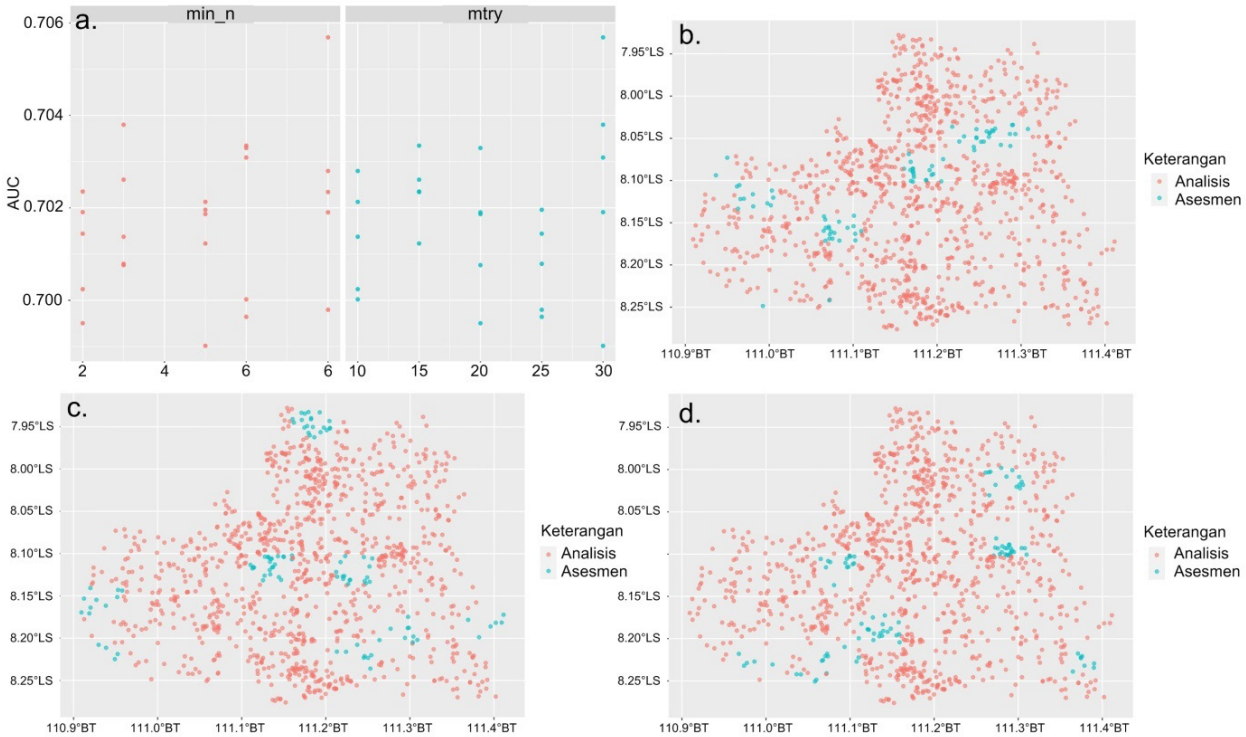
## Alur kerja randomforest

Alur kerja random forest lebih kompleks daripada regresi logistik. Penyetelan hiperparameter dan validasi silang dilakukan saat pembuatan model. Empat pilihan nilai hiperparameter min_n dan lima pilihan nilai hiperparameter mtry digunakan dalam penelitian ini (Gambar 7a). Validasi silang dilakukan menggunakan 10 fold *spatial block cross validation* yang berarti bahwa dalam setiap *running* satu *fold* data latih akan dibagi menjadi 90% analisis dan 10% asesmen (Gambar 7b, c, dan d).

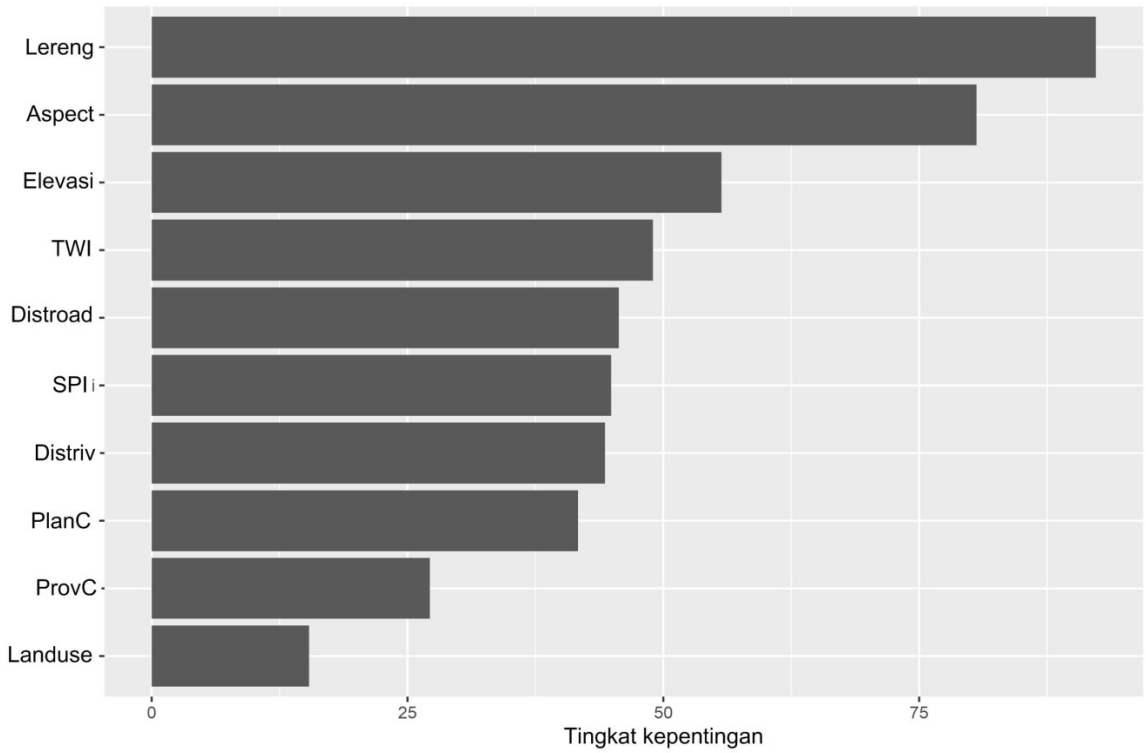
Pembagian data analisis dan asesmen pada validasi silang selain secara proporsi juga dapat dibagi berdasarkan blok secara spasial. Penyetelan 4 dan 5 nilai hiperparameter dikombinasikan dengan 10 validasi silang menghasilkan 200 model random forest dan 200 nilai AUC (*Area Under Curve*). Model terbaik dengan nilai AUC paling tinggi dipilih menjadi model random forest yang akan digunakan dalam pemodelan spasial. Pada penelitian ini, nilai min_n = 8 dan mtry = 30 merupakan nilai hiperparameter yang menghasilkan model terbaik. Konsekuensi dari proses penyetelan hiperparameter ini adalah beban komputer yang lebih berat dan proses perhitungan yang lebih lama dibandingkan dengan model regresi logistik.

Alur kerja random forest memungkinkan peneliti untuk menginterpretasi seberapa penting masing-masing faktor pengontrol dalam memprediksi secara spasial kemungkinan terjadinya longsor. *Tidymodel* menggunakan metode *Permutation Importance* dilakukan untuk mengukur seberapa penting masing-masing faktor pengontrol. *Permutation Importance* bekerja dengan cara mengevaluasi penurunan akurasi model yang disebabkan oleh pengurangan dan penggantian secara acak satu faktor pengontrol longsor, sementara yang lainnya dibiarkan. Penggantian faktor pengontrol yang menyebabkan turunnya akurasi model secara signifikan dianggap sebagai faktor pengontrol yang lebih penting. Pada penelitian ini faktor pengontrol yang paling penting adalah lereng dan yang paling tidak mempengaruhi berubahnya akurasi model adalah penggunaan lahan (Gambar 8)

Persentase luas area kerawanan rendah (0,01-0,2) hingga tinggi (0,8-0,99) pada model Random Forest secara berurutan adalah 17,9%, 30,7%, 28,4%, 19,5%, dan 3,4%. Wilayah dengan nilai 0,4-0,6 adalah 28,4%. Hal ini menunjukkan bahwa peta yang dihasilkan oleh model Random Forest memiliki *discriminative power* yang lebih baik dibandingkan dengan peta yang dihasilkan oleh model Regresi Logistik. Selain itu peta yang dihasilkan oleh model Random Forest memiliki pola spasial yang lebih baik dibandingkan dengan peta yang dihasilkan oleh model Regresi Logistik (Gambar 9).




**Gambar 7.** Penyetelan hiperparameter (a), validasi silang *fold* ke 1 (b), validasi silang fold ke 5 (c), dan validasi silang fold ke 9 (d)

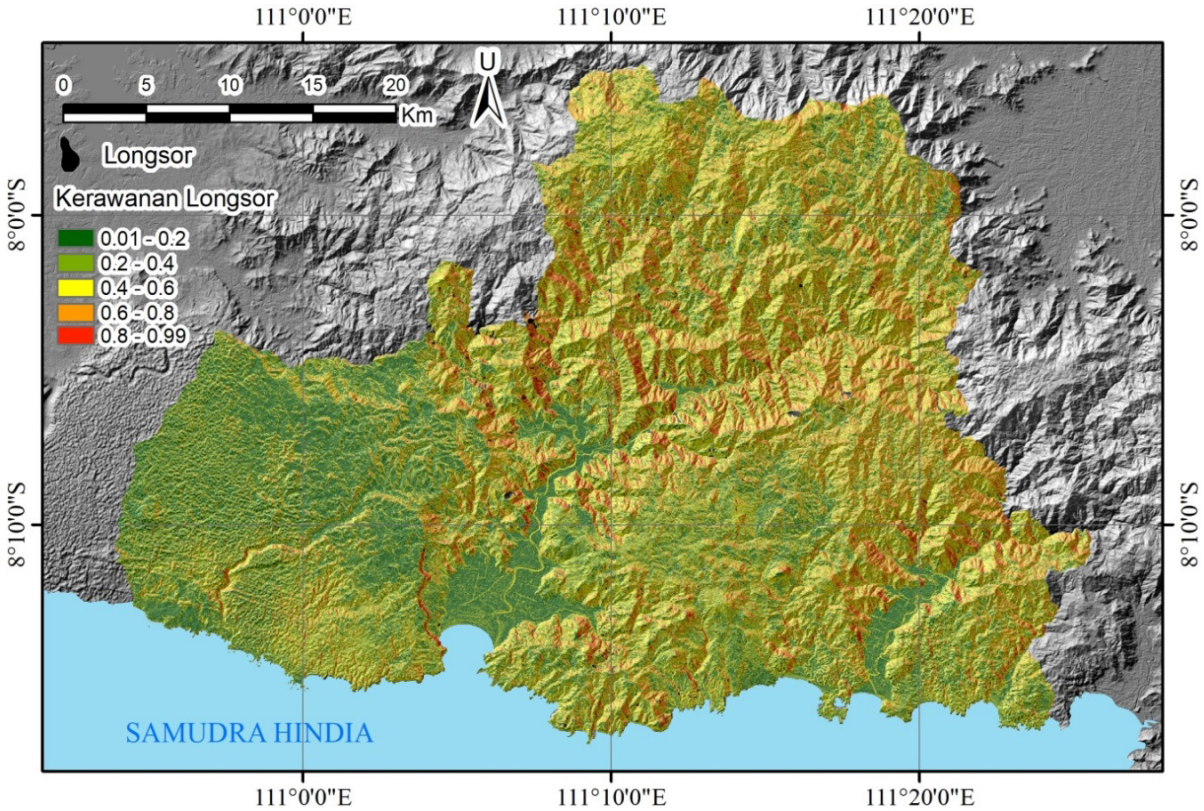



**Gambar 8.** *Variable importance* model Random Forest untuk pemodelan spasial kerawanan longsor

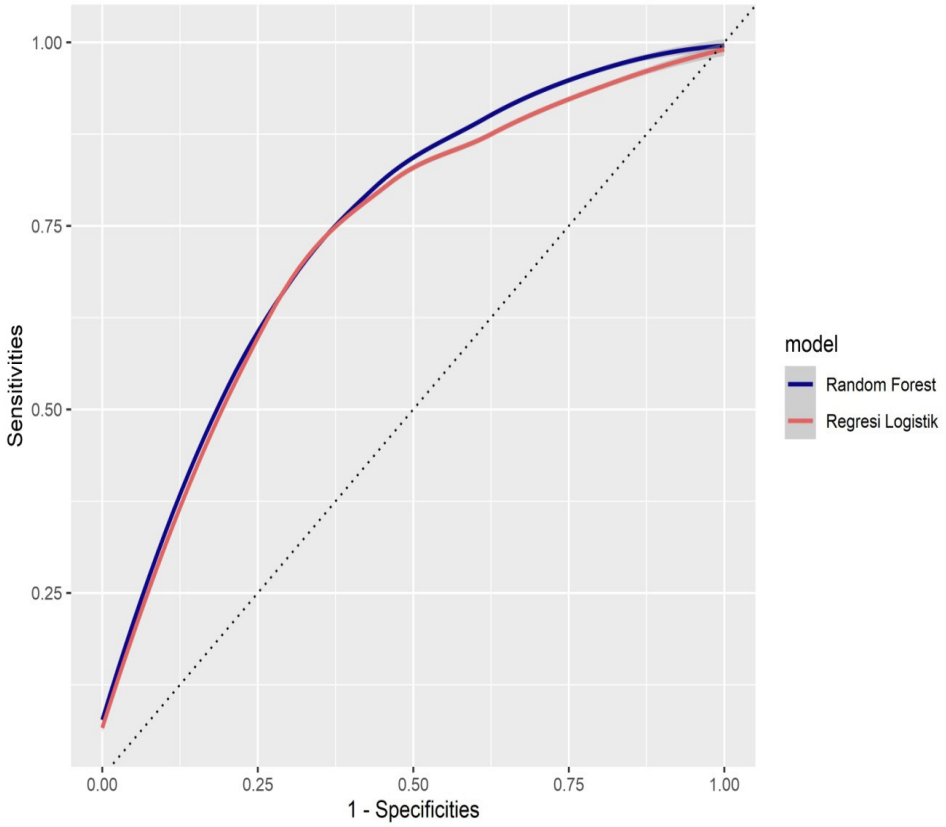
Secara visual, peta faktor pengontrol jarak dari sungai sangat mendominasi pola spasial peta kerawanan longsor Regresi Logistik. Area yang jauh dari sungai diklasifikasikan oleh model regresi logistik menjadi zona kerawanan 0,01-0,2. Area ini secara mengelompok berada di sisi barat Kabupaten Pacitan dengan morfologi berbukit dan didominasi oleh batuan gamping (Gambar 2b dan Gambar 8). Wilayah ini sering disebut sebagai bagian dari zona karst Gunung Sewu (Haryono & Day, 2004).

## Uji akurasi

Uji akurasi wajib dilakukan dalam pembuatan peta kerawanan longsor. Uji akurasi bermanfaat untuk mengevaluasi keakuratan model, membandingkan kehandalan




**Gambar 9.** kerawanan longsor Kab. Pacitan menggunakan teknik random forest




**Gambar 10.** Perbandingan uji akurasi regresi logistik dan random forest

model, mendeteksi *overfitting*, dan membantu pengambilan keputusan dalam tindakan pengurangan risiko bencana. Uji akurasi peta kerawanan longsor dapat menggunakan kurva *Receiver Operating Characteristic* (ROC) yang mencerminkan kemampuan prediksi peta kerawanan longsor dalam memperkirakan di mana akan terjadi longsor di masa depan.

Akurasi kurva ROC dapat dibaca dengan cara menghitung area di bawah kurva. Perhitungan akurasi diperoleh dengan cara membandingkan hasil pemodelan dengan data tes. Meski mirip, uji akurasi sering disalah artikan dengan proses validasi. Perbedaannya adalah proses validasi hanya menggunakan data latih dan tujuannya adalah untuk

memperoleh nilai hiperparamater yang paling baik (sub bab alur kerja RF). Gambar 10 menunjukkan akurasi model RF dan LR masing-masing sebesar 0,75 dan 0,73. Model RF lebih superior daripada LR dalam memprediksi kerawanan longsor. Model RF mengkombinasikan banyak model *decision trees* sehingga memiliki kemampuan prediksi yang lebih baik. Meski demikian, model RF memerlukan waktu yang lebih lama untuk melakukan perhitungan daripada model LR. Interpretasi keterkaitan faktor pengontrol pada model LR juga lebih mudah dilakukan daripada model RF. Koefisien LR mengindikasikan seberapa kuat faktor pengontrol dalam mempengaruhi kejadian longsor di masa lalu. Faktor pengontrol yang paling berpengaruh berdasarkan model LR adalah kecembungan baik itu *profile curvature* maupun *plan curvature*.

Penelitian ini menunjukkan bahwa kinerja RF lebih unggul deibandingkan LR sejalan dengan penelitian yang dilakukan oleh Catani et al. (2013) yang menunjukkan keunggulan RF disbanding model lainnya. Keunggulan model RF dapat disebabkan oleh kemampuan model dalam menangkap hubungan non-linier antar variabel pengontrol longsor (Liaw & Wiener, 2002). Model RF juga lebih tanah terhadap *overfitting* (akurat untuk data latih namun tidak untuk data tes) (Breiman, 2001). Meski demikian, RF memiliki model yang lebih kompleks dibanding LR (Cutler et al., 2007) sehingga memerlukan sumberdaya komputasi yang lebih besar dan waktu lebih lama. Selain itu interpretasi variabel pengontrol pada model RF tidak dapat dilakukan secara langsung. Meskipun terdapat nilai variabel importance, namun hasilnya tidak sejelas model LR yang menghasilkan koefisien yang langsung dapat diinterpretasi (Biau & Scornet, 2016). Akurasi model RF yang diterapkan di Pacitan hanya sedikit lebih baik daripada model LR karena data yang relatif kecil. Penggunaan jumlah data latih dan data tes yang lebih banyak akan meningkatkan akurasi model RF karena memberikan informasi yang lebih kaya dan bervariasi (Amato et al., 2013; Breiman, 2001; Chen et al., 2018).

## KESIMPULAN

Alur kerja pemodelan spasial untuk pemetaan kerawanan longsor sangat bervariasi. Penelitian ini membandingkan alur kerja pemodelan kerawanan longsor menggunakan model LR dan RF. Alur kerja LR dapat berupa penyiapan data, pra pemrosesan data, melatih model, pemodelan spasial, dan uji akurasi. Alur kerja RF dapat berupa penyiapan data, pra-pemrosesan data , melatih dan menyetel model, memvalidasi model, pemodelan spasial, dan uji akurasi. Hasil uji akurasi menunjukkan bahwa model RF lebih superior daripada LR dalam memprediksi kerawanan longsor. Meski demikian model RF memerlukan interpretasi yang lebih kompleks dan memerlukan waktu pemodelan yang lebih lama dibandingkan LR. Penjelasan alur kerja pembelajaran mesin perlu dideskripsikan dalam setiap pembuatan peta kerawanan longsor. Penjelasan tersebut dapat menjamin reproduktifitas, transparansi, interpretabilitas, kejelasan metode yang dipakai, kemudahan mengidentifikasi kesalahan atau bias, dan dapat memfasilitasi kolaborasi dan berbagi pengetahuan.

## UCAPAN TERIMA KASIH

Penelitian ini didanai oleh Hibah Penelitian Penelitian mandiri Dosen Tahun 2024 *Batch* 1 nomor SK/kontrak 130/UN1/GE/KPT/2024.

## KONTRIBUSI PENULIS

**Guruh Samodra** mendisain metode penelitian, analisis data, dan membuat naskah publikasi; melakukan analisis data dan interpretasi hasil; *coding* program dan review naskah publikasi.

## DAFTAR PUSTAKA

Abraham, M. T., Satyam, N., Lokesh, R., Pradhan, B., & Alamri, A. (2021). Factors Affecting Landslide Susceptibility Mapping: Assessing the Influence of Different Machine Learning Approaches, Sampling Strategies and Data Splitting. *Land 2021, Vol. 10, Page 989*, *10*(9), 989. https://doi.org/10.3390/LAND10090989

Agresti, A. (2013). *Categorical Data Analysis*. John Wiley & Sons, Ltd.

Amato, F., López, A., Peña-Méndez, E. M., Vaňhara, P., Hampl, A., & Havel, J. (2013). Artificial neural networks in medical diagnosis. *Journal of Applied Biomedicine*, *11*(2), 47–58. https://doi.org/10.2478/V10136-012-0031-X

Badan Standardisasi Nasional. (2016). *Standar nasional Indonesia 8291:2016 Penyusunan dan Penentuan Zona Kerentanan Gerakan Tanah*.

Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, *24*(1), 43–69. https://doi.org/10.1080/02626667909491834

Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, *25*(2), 197–227. https://doi.org/10.1007/S11749-016-0481-7/METRICS

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Cascini, L. (2008). Applicability of landslide susceptibility and hazard zoning at different scales. *Engineering Geology*, *102*(3–4), 164–177. https://doi.org/10.1016/j.enggeo.2008.03.016

Catani, F., Lagomarsino, D., Segoni, S., & Tofani, V. (2013). Landslide susceptibility estimation by random forests technique: Sensitivity and scaling issues. *Natural Hazards and Earth System Sciences*, *13*(11), 2815–2831. https://doi.org/10.5194/NHESS-13-2815-2013

Chen, W., Xie, X., Peng, J., Shahabi, H., Hong, H., Bui, D. T., Duan, Z., Li, S., & Zhu, A. X. (2018). GIS-based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based random forest method. *CATENA*, *164*, 135–149. https://doi.org/10.1016/J.CATENA.2018.01.012

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., & Böhner, J. (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, *8*(7), 1991–2007. https://doi.org/10.5194/GMD-8-1991-2015

Corominas, J., & Moya, J. (2008). A review of assessing landslide frequency for hazard zoning purposes. *Engineering Geology*, *102*(3–4), 193–213.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*(11), 2783–2792. https://doi.org/10.1890/07-0539.1

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., & Savage, W. Z. (2008). Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Engineering Geology*, *102*(3–4), 85–98.

Goetz, J. N., Brenning, A., Petschko, H., & Leopold, P. (2015). Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Computers & Geosciences*, *81*, 1–11. https://doi.org/10.1016/J.CAGEO.2015.04.007

Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, *3*, 1157–1182.

Haryono, E., & Day, M. (2004). Landform Differentiation within The Gunung Kidul Kegelkarst, Java, Indonesia. *Journal of Cave and Karst Studies*, *66*(2), 62–69. https://caves.org/pub/journal/PDF/V66/v66n2-Haryono.pdf

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. https://doi.org/10.1007/978-0-387-84858-7

Hussin, H. Y., Zumpano, V., Reichenbach, P., Sterlacchini, S., Micu, M., van Westen, C., & Bălteanu, D. (2016). Different landslide sampling strategies in a grid-based bi-variate statistical susceptibility model. *Geomorphology*, *253*, 508–523. https://doi.org/10.1016/J.GEOMORPH.2015.10.030

Kuhn, M., & Wickham, H. (2020). *Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles*. https://www.tidymodels.org

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R. News*, *2*(3), 18–22.

Lovelace, R., Nowosad, J., & Münchow, J. (2020). *Geocomputation with R* (1st ed.). Chapman and Hall/CRC Press.

Menard, S. (2002). Applied Logistic Regression Analysis. *Applied Logistic Regression Analysis*. https://doi.org/10.4135/9781412983433

Moore, I. D., Grayson, R. B., & Ladson, A. R. (1991). Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrological Processes*, *5*(1), 3–30. https://doi.org/10.1002/hyp.3360050103

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, *41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6

Probst, P., Wright, M. N., & Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(3), e1301. https://doi.org/10.1002/WIDM.1301

Quinn, P., Beven, K., Chevallier, P., & Planchon, O. (1991). The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrological Processes*, *5*(1), 59–79. https://doi.org/10.1002/hyp.3360050106

R Core Team. (2021). *R: a Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/

Samodra, G. (2023). *Teknologi Geospasial Inventarisasi Longsor* (Eds Pertam). Gadjah Mada University Press.

Samodra, G., Chen, G., Sartohadi, J., & Kasama, K. (2017). Comparing data-driven landslide susceptibility models based on participatory landslide inventory mapping in Purwosari area, Yogyakarta, Java. *Environmental Earth Sciences*, *76*(4), 1–19. https://doi.org/10.1007/s12665-017-6475-2

Samodra, G., Ngadisih, N., Malawani, M. N., Mardiatno, D., Cahyadi, A., & Nugroho, F. S. (2020). Frequency–magnitude of landslides affected by the 27–29 November 2017 Tropical Cyclone Cempaka in Pacitan, East Java. *Journal of Mountain Science*, *17*(4), 773–786. https://doi.org/10.1007/s11629-019-5734-y

Schuster, R. L., & Kockelman, W. J. (1996). Principles of landslide hazard reduction. In A. K. Turner & R. L. Schuster (Eds.), *Landslides: Investigation and Mitigation* (pp. 91–105). National Academy Press. National Research Council, Transportation Research Safety Board Special Report.

Smyth, H. R., Hall, R., & Nichols, G. J. (2008). Cenozoic volcanic arc history of East Java, Indonesia: The stratigraphic record of eruptions on an active continental margin. *Special Paper of the Geological Society of America*, *436*, 199–222. https://doi.org/10.1130/2008.2436(10)

Soeters, R., & Westen, C. J. van. (1996). Slope instability recognition, analysis, and zonation. In A. K. Turner & R.L. Schuster (Eds.), *Landslides, investigation and mitigation ( Transportation Research Board, National Research Council, Special Report ; 247)* (pp. 129–177). National Academy Press.

Steger, S., Brenning, A., Bell, R., Petschko, H., & Glade, T. (2016). Exploring discrepancies between quantitative validation results and the geomorphic plausibility of statistical landslide susceptibility maps. *Geomorphology*, *262*, 8–23. https://doi.org/10.1016/j.geomorph.2016.03.015

van Westen, C. J., Castellanos, E., & Kuriakose, S. L. (2008). Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. *Engineering Geology*, *102*(3–4), 112–131. https://doi.org/10.1016/J.ENGGEO.2008.03.010