

# Aplikasi *Machine Learning* Method pada Pemetaan Kerawanan Tanah Longsor di Kabupaten Karanganyar

Nada Hanifah Putri<sup>1\*</sup>, Raden Harya Dananjaya<sup>2</sup>, Niken Silmi Surjandari<sup>3</sup>

<sup>1,2,3</sup> Program Studi Teknik Sipil, Universitas Sebelas Maret, Jl. Ir. Sutami 36A, Surakarta 57126, Fax 662118; Telp. (0271) 634524.

Email: nada\_hanifah\_putri@student.uns.ac.id

## Abstrak

Indonesia berada dalam zona iklim tropis yang rawan untuk mengalami bencana hidrometeorologi. Pemetaan kerawanan longsor merupakan salah satu upaya mitigasi yang dapat dilakukan untuk mengurangi dampak dari bencana tanah longsor. Penelitian ini bertujuan untuk membuat peta kerawanan longsor wilayah Kabupaten Karanganyar menggunakan *machine learning* yang diklasifikasikan menjadi lima kelas yaitu sangat rendah, rendah, sedang, tinggi, dan sangat tinggi. Metode yang digunakan untuk pembuatan model adalah *Voting Classifier Ensemble Technique*. Sembilan faktor pengondisi yang digunakan yaitu jarak terhadap jalan sekunder dan tersier, slope, TWI, elevasi, *land use*, litologi, NDVI, serta curah hujan. Algoritma *machine learning* didapatkan dari modul *Scikit Learn*. Kombinasi parameter yang digunakan yaitu pada metode *Random Forest* menggunakan parameter *random\_state* = 0, *n\_estimators* = 750, *criterion* = 'entropy', metode *Support Vector Machine* menggunakan parameter *random\_state* = 0, *Probability* = True, *gamma* = 0.005, *C* = 1, metode *K-Nearest Neighbors* menggunakan parameter *n\_neighbors* = 11, *weights* = 'distance', *leaf\_size* = 20, dan metode *Voting Classifier* menggunakan parameter *voting* = 'soft', *weights* = [1,1,1] untuk parameter lain yang digunakan diatur sesuai dengan default modul. Model yang didapatkan memiliki AUC sebesar 0,9563 yang mendekati 1 sehingga dapat dikatakan bahwa model yang dimiliki performa yang baik untuk melakukan prediksi probabilitas longsor.

**Kata kunci:** longsor, *machine learning*, *Voting Classifier*.

## Abstract

Indonesia is in a tropical climate zone which is prone to hydrometeorological disasters. Mapping landslide susceptibility is one of the countermeasures that can reduce the impact of landslides. This study aims to make Karanganyar Regency landslides susceptibility map using machine learning which is classified into five classes, namely very low, low, medium, high, and very high. The method used for making model is *Voting Classifier Ensemble Technique*. Nine conditioning factors used are distance to secondary and tertiary roads, slope, TWI, elevation, *land use*, lithology, NDVI, and rain. The machine learning algorithm is taken from *Scikit Learn* module. The combination of parameters used is the *Random Forest* method using parameters *random\_state* = 0, *n\_estimators* = 750, *criteria* = 'entropy', *Support Vector Machine* method using parameters *random\_state* = 0, *Probability* = True, *gamma* = 0.005, *C* = 1, *K-method Nearest Neighbors* uses parameters *n\_neighbors* = 11, *weights* = 'distance', *leaf\_size* = 20, and *Voting Classifier* method uses parameters *voting* = 'soft', *weights* = [1,1,1] for the other parameters used are set according to the default module. The model obtained has an AUC of 0.9563 which is close to 1 so can be said that model has good performance for predicting landslide stability.

**Keywords:** landslide, machine learning, *Voting Classifier*.



Copyright © 2022 The Author(s)

This is an open access article under the [CC-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license.

## 1. PENDAHULUAN

Indonesia berada dalam zona iklim tropis yang memiliki kerawanan tinggi untuk mengalami bencana hidrometeorologi. Jawa Tengah merupakan provinsi dengan kasus tanah longsor terbanyak kedua di pulau Jawa dengan total 2975 kasus dalam kurun waktu tahun 2003 sam-

pai tahun 2022 (BNPB, 2023). Pemetaan wilayah kerawanan longsor dapat menjadi salah satu upaya mitigasi yang dapat dilakukan. Bencana longsor dipengaruhi oleh banyak faktor baik faktor buatan (aktivitas manusia, jarak dengan jalan, dll) ataupun faktor alam non geoteknis (kerapatan vegetasi, perubahan curah hujan, dsb).

Pada penelitian ini dilakukan ensambel tiga metode *machine learning* yang sudah sering digunakan untuk membuat peta kerawanan longsor, ketiga metode itu di ensambel menggunakan metode *Voting Classifier*. Pemilihan metode yang di ensambel diambil dari hasil penelitian sebelumnya, pemilihan metode *Random Forest* berdasarkan pada penelitian yang dilakukan oleh Achour & Pourghasemi (2020), Youssef & Pourghasemi (2021), serta Merghadi dkk. (2020), pemilihan metode *Support Vector Machine* berdasarkan hasil terbaik dari penelitian yang dilakukan oleh Pham dkk. (2017), pemilihan metode *K-Nearest Neighbors* berdasarkan dari penelitian yang dilakukan oleh Merghadi dkk. (2020). Penelitian ini memiliki kesamaan dengan penelitian yang dilakukan oleh Marzuki (2021) dalam hal jenis faktor pengondisi dan metode validasi model yang digunakan, perbedaan yang dimiliki adalah pada penelitian yang dilakukan oleh Marzuki (2021) wilayah studi hanya lima kecamatan (Kecamatan Ngargoyoso, Karangpandan, Jenawi, Tawangmangu, dan Jatiyoso) di Kabupaten Karanganyar sedangkan pada penelitian ini wilayah studi mencakup seluruh wilayah Kabupaten Karanganyar.

Penelitian ini bertujuan untuk membuat pemetaan kerawanan longsor menggunakan algoritma *ensemble machine learning* yang basis pada kecerdasan buatan. Metode *machine learning* yang digunakan dalam penelitian ini adalah *Voting Classifier Ensemble Technique* yang mengklasifikasikan wilayah longsor dan tidak longsor dengan kombinasi faktor penyebab longsor yang dianggap paling berpengaruh dengan mempertimbangkan kondisi disekitar titik-titik yang pernah terjadi longsor dengan menggabungkan tiga metode *machine learning*, kemudian akan dipresentasikan untuk mendapatkan lima tingkat kerawanan (sangat rendah, rendah, sedang, tinggi, dan sangat tinggi) yang kemudian digunakan sebagai dasar untuk membuat pemetaan kerawanan longsor. Sembilan faktor pengondisi yang digunakan dalam penelitian ini merujuk pada hasil penelitian uji tingkat pengaruh faktor pengondisi yang telah dilakukan oleh Fitria (2021) yang meliputi jarak terhadap jalan sekunder dan tersier, slope, *Topographic Wetness Index* (TWI), elevasi, tata guna lahan (*land use*), litologi, *Normalized Difference Vegetation Index* (NDVI), dan curah hujan. Penelitian ini diharapkan dapat membantu mitigasi bencana tanah longsor dan dapat

membantu dalam perencanaan di wilayah Kabupaten Karanganyar.

### **Tanah Longsor**

Tanah longsor merupakan salah satu jenis gerakan massa tanah atau batuan, ataupun percampuran keduanya, menuruni atau keluar lereng akibat terganggunya kestabilan tanah atau batuan penyusun lereng (BPBD, 2022). Tanah longsor terjadi ketika air yang meresap ke dalam tanah menambah bobot tanah dan jika air terus meresap sampai pada tanah kedap air yang merupakan bidang gelincir maka tanah menjadi licin dan tanah yang ada di atasnya akan bergerak mengikuti lereng dan keluar lereng. Pusat Vulkanologi dan Mitigasi Bencana Geologi menyatakan ada enam jenis tanah longsor yaitu longsor translasi, longsor rotasi, pergerakan blok, runtuh batu, rayapan tanah dan aliran bahan rombakan.

### **Faktor Pengondisi Jarak Terhadap Jalan**

Penelitian sebelumnya menunjukkan bahwa jarak terhadap jalan menjadi faktor pengondisian yang berpengaruh dalam terjadinya tanah longsor. Penggalan jalan dan jalan raya, terutama di daerah pegunungan yang dibangun di samping lereng, dapat menyebabkan perubahan keseimbangan alami tanah yang menyebabkan potensi ketidakstabilan lereng. Jarak yang lebih jauh dari jalan menyebabkan penurunan beban baik di topografi maupun di kaki jalan lereng (Benbouras, 2022).

### **Faktor Pengondisi Slope**

Lereng/*slope* merupakan faktor pengondisi yang paling penting dalam tanah longsor. Pada skala lokal, *slope* mempengaruhi konsentrasi kelembaban dan tingkat tekanan pori. Dalam skala besar *slope* mengontrol kesinambungan hidraulik regional, dan dianggap sebagai faktor utama untuk pemetaan berbasis GIS. Sudut kemiringan umumnya berkorelasi dengan *slope movement* karena sangat terkait dengan kekuatan yang terlibat. Semakin tinggi sudut kemiringan, semakin rentan terhadap tanah longsor (Benbouras, 2022)

### **Faktor Pengondisi Topographic Wetness Index (TWI)**

*Topographic Wetness Index* (TWI) adalah angka kadar air tanah yang dipengaruhi oleh topografi, yang mengontrol arah aliran air tanah (Meinhardt dkk., 2015). Nilai TWI yang tinggi

mengindikasikan tingginya kandungan air tanah, dimana tanah dengan kandungan air yang tinggi lebih rentan terhadap longsor (Zhou dkk., 2021). Nilai TWI dinyatakan dalam Persamaan (1).

$$TWI = \ln \frac{\alpha}{\tan \beta} \quad (1)$$

dengan:

$\alpha$  : daerah penyumbang kemiringan lereng yang mengindikasikan arah aliran air ke suatu wilayah (derajat)

$\beta$  : nilai kemiringan lokal (derajat)

### Faktor Pengondisi Elevasi

Elevasi juga dianggap sebagai faktor pengondisi penting lain yang mempengaruhi manifestasi tanah longsor. Secara tidak langsung ketinggian terkait dengan faktor-faktor lain seperti kemiringan, erosi, curah hujan, ketebalan tanah, dan penggunaan lahan (Benbouras, 2022). Angka variasi elevasi suatu daerah mengindikasikan perbedaan elevasi permukaan suatu daerah, dimana longsor sering terjadi pada daerah perbukitan dan pegunungan yang terdapat perbedaan elevasi cenderung besar.

### Faktor Pengondisi Tata Guna Lahan

Aktivitas manusia berupa perluasan wilayah perkotaan dan pembangunan infrastruktur berperan penting dalam memicu tanah longsor. Kegiatan rekayasa ini memerlukan pemotongan atau penggalian lereng untuk jalan raya dan konstruksi bangunan, hal ini akan mengubah kondisi geologi asli dan kemudian mengganggu stabilitas lereng (Youssef & Pourghasemi, 2021). Parameter seperti kohesi, sudut gesek, berat satuan tanah, dan tekanan pori cenderung mengalami perubahan ketika mengalami berbagai jenis penutupan (Bragagnolo dkk., 2020).

### Faktor Pengondisi Litologi

Litologi dianggap menjadi salah satu faktor pemicu longsor yang paling mendasar (Benbouras, 2022). Variasi litologi memiliki dampak yang signifikan pada berbagai jenis geo-hazard misalnya, tanah longsor dan penurunan tanah (Youssef & Pourghasemi, 2021). Setiap jenis litologi bervariasi dalam karakteristik fisik dan mekanik termasuk jenis, kekuatan, tingkat pelapukan, daya tahan, kepadatan, dan permeabilitas (Youssef & Pourghasemi, 2021).

### Faktor Pengondisi *Normalized Difference Vegetation Index* (NDVI)

NDVI berperan penting dalam penyerapan air dalam jumlah besar dan meningkatkan geser resistensi dan kohesi tanah dari massa litologi (Youssef & Pourghasemi, 2021). NDVI digunakan dalam penelitian ini untuk mencerminkan kerapatan vegetasi. Nilai NDVI berkisar antara -1 hingga 1 (Youssef & Pourghasemi, 2021). Nilai NDVI dinyatakan dalam Persamaan (2).

$$NDVI = \frac{(NIR-R)}{(NIR+R)} \quad (2)$$

dengan:

*NIR* : *near infrared* (%)

*R* : *red bands of the electromagnetic spectrum* (%)

### Faktor Pengondisi Hujan

Kasus longsor yang pernah terjadi beberapa diantaranya disebabkan oleh curah hujan sehingga bisa dikatakan bahwa hujan menjadi salah satu penyebab utama kejadian longsor. Curah hujan dianggap sebagai faktor pemicu longsor karena dapat mengubah sifat kekuatan geser lereng (Youssef & Pourghasemi, 2021) Proses infiltrasi air hujan ke dalam tanah juga dipengaruhi oleh intensitas hujan suatu daerah, proses infiltrasi mengubah tekanan air pori dan melemahkan tegangan tanah.

### *Random Forest*

*Random Forest* (RF) adalah algoritma *supervised classification* dan metode ansambel menggunakan model *Decision Tree* sedemikian rupa sehingga setiap pohon sesuai dengan subset data yang disampel secara independen menggunakan *bootstrapping* dan metode ansambel yang telah berhasil diterapkan di studi penilaian kerawanan longsor. RF dapat menangani kumpulan data besar untuk menyelesaikan masalah dengan hasil yang akurat dan andal tanpa atau sedikit penyempurnaan, serta menghasilkan resistensi *overfitting* model (Hammad Khaliq dkk., 2023). *Bagging* adalah langkah pertama metode *Random Forest*, yang kemudian dilanjutkan prediksi sejumlah besar *tree* (Hoarau dkk., 2023).

RF terdiri dari banyak model *Decision Tree* individual, oleh karena itu digunakan istilah '*forest*'. RF diketahui memberikan tingkat akurasi yang tinggi sehubungan dengan *outlier*

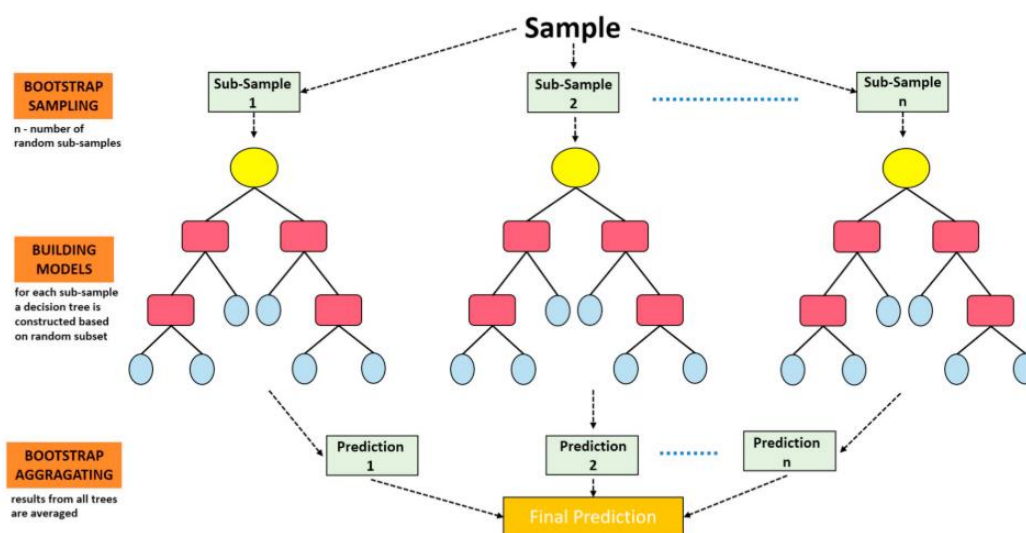
dalam prediktor karena penggunaan pemilihan acak pada setiap node yang terpisah tergantung pada dua objek data, yaitu *Out-Of-Bag* (OOB) dan *proximities*. Secara keseluruhan, keacakan dalam algoritma RF dapat mengurangi *overfitting* dengan membangun beberapa pohon, menggambarkan pengamatan dengan penggantian yaitu, *bootstrap*, dan dalam subset acak, membagi node pada pemisahan terbaik. Keuntungan utama RF adalah menghindari dengan pemilihan fitur secara acak. Prinsipnya adalah untuk mengatasi kelemahan *Decision Tree*, *high variance*, dengan menggabungkan

### Support Vector Machine (SVM)

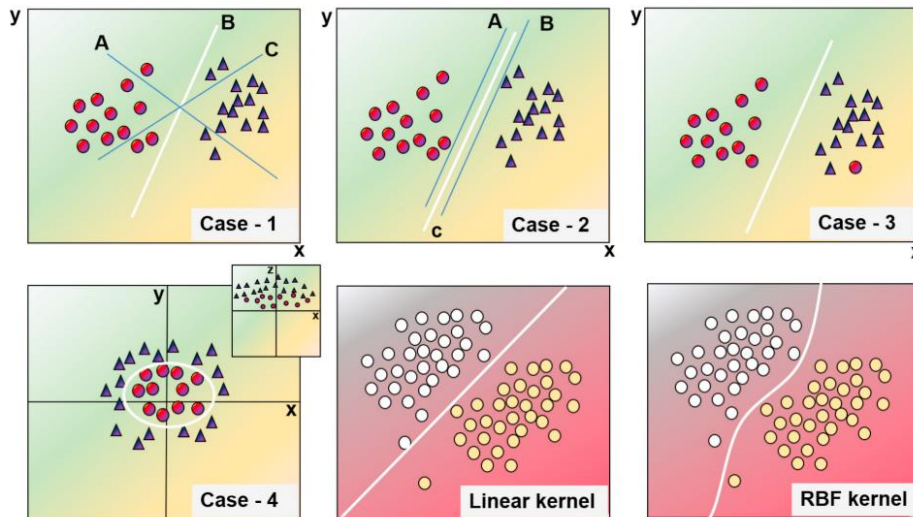
*Support Vector Machines* (SVM) adalah *binary classifier* berdasarkan teori pembelajaran statistik yang diterapkan untuk mengidentifikasi yang terbaik pemisahan *hyperplane* yang dapat membagi dua zona (Achour & Pourghasemi, 2020). *Support Vector Machine* adalah algoritma *Machine Learning* yang diawasi non-parametrik yang menggunakan *mathematical tools* untuk memecahkan solusi non-linear untuk masalah klasifikasi dan regresi (Merghadi dkk., 2020). SVM adalah algoritma pembelajaran mesin yang diawasi berdasarkan teori pembelajaran statistik dan prinsip minimalisasi risiko struktural (Yang dkk., 2022).

*overfitting*, masalah yang umum terjadi pada model *Decision Tree* generik. Namun, RF dianggap sebagai *black-box* model karena interpretasi keputusan yang terbatas. Namun demikian, RF atau variannya (misalnya, *Extremely Randomized Trees*) dianggap sebagai model yang tangguh (Merghadi dkk., 2020). Langkah-langkah utama RF secara skematis ditunjukkan pada Gambar 1.

SVM cocok untuk penanganan data korelasi non-linear oleh kernel fungsi. Empat jenis kernel bisa dibedakan untuk pemetaan SVM yaitu linier, polinomial, *Radial Basis Function* (RBF), dan sigmoid (Achour & Pourghasemi, 2020). *Output* yang diinginkan dalam model SVM dicapai dengan mengubah data input menjadi ruang berdimensi tinggi, menggunakan trik kernel (*function*), kemudian dilatih dan dibawa kembali ke ruang dua dimensi. SVM cocok untuk masalah dua kelas di mana ada sampel positif dan negatif. Untuk masalah klasifikasi dua kelas, SVM berupaya menemukan rencana pemisahan yang diperlukan, yaitu *hyper-plane* di ruang fitur yang secara maksimal membagi dua kelas sampel seperti pada Gambar 2 (Merghadi dkk., 2020).



Gambar 1. Prinsip kerja RF, Merghadi dkk. (2020)



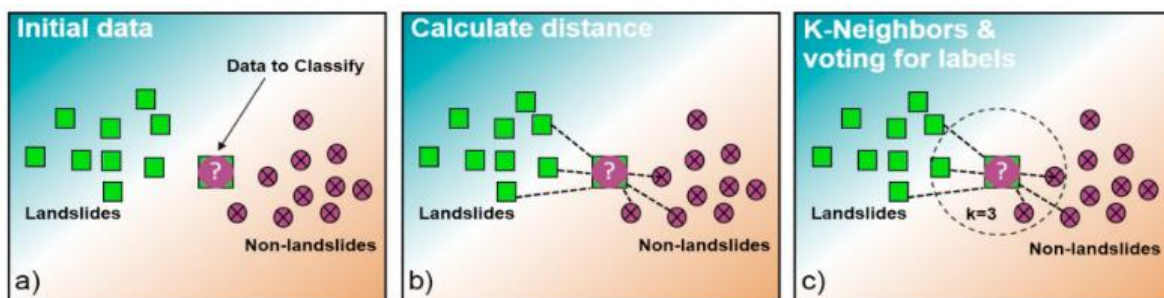
Gambar 2. Skema pemilihan hyper-plane dan tipe kernel yang optimal dalam *Support Vector Machine*, Merghadi dkk. (2020)

### ***K-Nearest neighbors***

*K-Nearest neighbors* banyak digunakan sebagai model klasifikasi yang efektif untuk menentukan label kelas dari sampel yang dipilih sebelumnya dengan mendeteksi kategori sampel pelatihan terdekat (Zhao dkk., 2023). KNN merupakan metode yang sederhana dari semua algoritma *machine learning* lainnya dan beberapa waktu belakangan ini digunakan di berbagai bidang termasuk pemetaan kerawanan longsor (Merghadi dkk., 2020). Algoritma KNN mengklasifikasikan *instance* dengan *majority vote* (*k neighbor*), parameter *k* adalah bilangan bulat positif dan sering berupa bilangan kecil, untuk klasifikasi biner, jumlah *k* yang ganjil dapat membantu untuk menghindari kesulitan dengan *tied vote* (Sameen dkk., 2020).

KNN biasanya disebut sebagai *supervised algorithm nonparametrik*, berbasis *instance*, dan '*lazy*'. Dalam konteks ini, '*lazy*' atau '*lazy*

*learning*' mengacu pada model di mana pembelajaran ditunda hingga kueri dibuat ke sistem, sebagai lawan dari '*eager learning*', di mana algoritma mencoba menggeneralisasi data pelatihan sebelum menerima kueri. Istilah '*non-parametrik*', mengacu pada fakta bahwa algoritme tidak bergantung pada distribusi data untuk perhitungannya, dan model menyesuaikan dengan data secara otomatis. Hal ini sangat berguna untuk *Landslide Susceptibility Map* di mana sebaran tanah longsor biasanya tidak sesuai dengan sebaran yang rapi. *Lazy learning* memastikan bahwa fungsi hanya didekati secara lokal dan semua perhitungan ditunda hingga klasifikasi. Akibatnya, KNN tidak mempelajari fungsi diskriminatif dari data pelatihan, melainkan menghafal dataset pelatihan. Langkah-langkah dalam implementasi model KNN dapat dilihat pada Gambar 3 (Merghadi dkk., 2020).



Gambar 3. Langkah-langkah dalam model *k-nearest neighbor*: a) menerima data yang tidak terklasifikasi; b) menghitung jarak; dan c) menentukan tetangga terdekat dan memilih label, Merghadi dkk. (2020)



### Voting classifier ensemble technique

Teknik *Voting Classifier Ensemble* adalah contoh umum dari pendekatan *multi-expert*, yang membantu menggabungkan pengklasifikasi secara paralel. Selanjutnya, setiap pengklasifikasi melatih semua data dan berkontribusi pada keputusan. Akhirnya, teknik pemungutan suara membantu menghasilkan solusi akhir (Naji dkk., 2021). Dalam metode *Voting Classifier*, label kelas yang diprediksi untuk sampel tertentu adalah label kelas yang mewakili mayoritas (*mode*) dari label kelas yang diprediksi oleh masing-masing pengklasifikasi individu (Scikit-learn, 2023).

*Voting Classifier* mendukung dua jenis *voting*, *hard voting* dan *soft voting*. *Output class* yang diproyeksikan dalam *hard voting* adalah yang memiliki suara terbanyak, yaitu yang memiliki kemungkinan tertinggi untuk diprediksi oleh masing-masing pengklasifikasi. *Hard voting* mungkin berbobot sama atau berbobot tidak sama yang memungkinkan penetapan prioritas yang lebih tinggi salah satu pengklasifikasi. *Output class* diprediksi dalam *soft voting* berdasarkan probabilitas yang ditetapkan pada setiap kelas oleh pengklasifikasi. *Soft voting* dengan bobot yang sama semua pengklasifikasi memiliki pengaruh yang sama sedangkan, dalam *soft voting* dengan bobot yang tidak sama, pengklasifikasi yang berbeda diberi bobot yang berbeda tergantung pada pengaruhnya dalam klasifikasi. Probabilitas yang diberikan oleh pengklasifikasi dikalikan dengan bobot untuk menetapkan probabilitas ke kelas dan kelas dengan nilai tertinggi dipilih

(Rai dkk., 2022). Alur kerja metode *Voting Classifier* dapat dilihat pada Gambar 4.

### Natural Breaks (Jenks)

Klasifikasi metode *Natural Break* (Jenks) membagi kelas berdasarkan nilai, kelas dikelompokkan berdasarkan kemiripan nilai dan memaksimalkan perbedaan antar kelas. Batas kelas ditetapkan dimana terdapat perbedaan nilai data yang relatif besar (esri, 2023). Metode *Natural Breaks* mengelompokkan data ke dalam kelompok yang meminimalkan varian dalam kelompok dan memaksimalkan varian antar kelompok (CDC, 2023).

### Landslide Density

*Landslide density* merupakan rasio dari persentase piksel longsor dan persentase kelas piksel pada masing-masing kelas rentan, seperti ditunjukkan pada Persamaan (3). *Landslide density* dihitung untuk mengevaluasi performa hasil klasifikasi kerawanan longsor yang dilakukan. Jika *landslide density* meningkat secara bertahap dari yang sangat rendah kelas sampai kelas sangat tinggi, maka hasil visualisasi peta kerawanan longsor dapat dikatakan baik (Pham dkk., 2017).

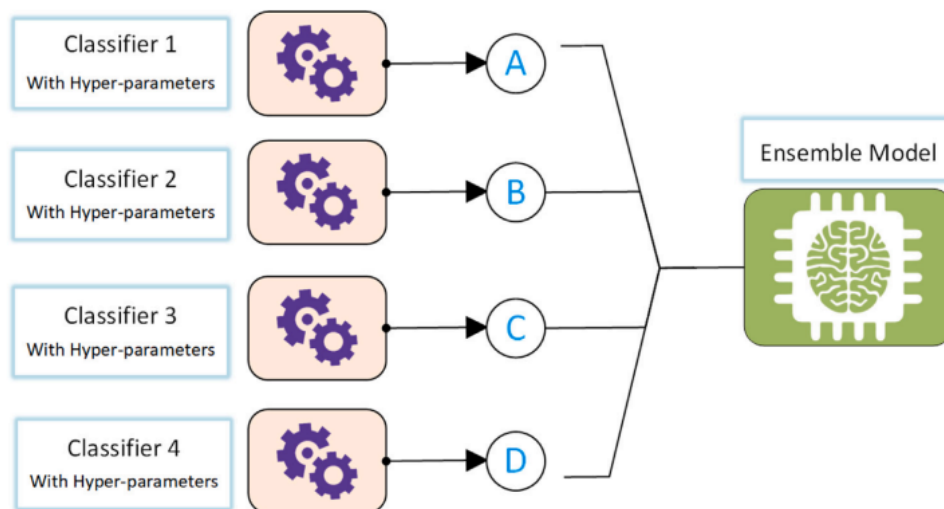
$$LD = \frac{T_{PL}}{T_P} \quad (3)$$

keterangan :

*LD* : landslide density

*T<sub>PL</sub>* : total piksel longsor dalam satu kelas

*T<sub>P</sub>* : total piksel dalam satu kelas



Gambar 4. Alur dasar dari *Voting Classifier Ensemble*, Khan dkk. (2023)

## Receiver Operating Characteristics Curve (ROC Curve) dan Area Under the Curve (AUC)

Receiver Operating Characteristics Curve (ROC Curve) adalah metode yang berguna untuk memverifikasi peta kerawanan longsor dengan menunjukkan sensitivitas pada sumbu vertikal dan 1-spesifisitas pada sumbu horizontal (Akinci, 2022). Kesesuaian seluruh data pada model yang dibuat direpresentasikan oleh kurva ROC. Suatu kurva dihasilkan dari kurva ROC, dimana area di bawah kurva tersebut (AUC) mengindikasikan nilai akurasi suatu model. Keberhasilan prediksi suatu model ditunjukkan oleh nilai AUC data training pada variabel yang diteliti, sedangkan nilai AUC pada data *testing* menunjukkan kemampuan prediksi suatu model. Nilai AUC berada di antara angka 0 dan 1, semakin tinggi nilai AUC maka semakin baik juga prediksi yang dapat dilakukan suatu model.

Tabel 2. Kombinasi parameter machine learning

Nilai AUC	Kategori Performa Model
0,90 – 1,00	Unggul ( <i>Excellent</i> )
0,80 – 0,90	Bagus ( <i>Good</i> )
0,70 – 0,80	Cukup ( <i>Fair</i> )
0,60 – 0,70	Buruk ( <i>Poor</i> )
0,50 – 0,60	Gagal ( <i>Fail</i> )

## 2. METODE

### Pengumpulan Data

Penelitian ini menggunakan data sekunder berupa peta dalam bentuk raster dan vektor. Peta yang diperlukan yaitu peta ASTER GDEM (*Advance Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation Model*), peta data TRMM (*Tropical Rainfall Measuring Mission*), dan peta data MODIS (*Moderate Resolution Imaging Spectroradiometer*) dalam bentuk raster yang didapatkan dari *website* yang disediakan oleh NASA. Peta *landslide inventory*, *land use*, dan jalan indonesia diperoleh dari penelitian yang dilakukan oleh Priyono dkk. (2020), serta peta geologi diperoleh dari PT. Indonesia Geospasial Tech (2020) yang disediakan dalam format data vektor. Rekapitulasi data sekunder yang digunakan dapat dilihat pada Tabel 1.

Tabel 1. Nama dan format peta

Nama Peta	Format Peta	Sumber
Peta <i>Landslide Inventory</i>	Vektor	Priyono <i>et al.</i> (2020)
Peta ASTER GDEM	Raster	NASA (2013)
Peta Data TRMM	Raster	NASA (2020)
Peta Data MODIS	Raster	NASA (2022)
Peta Geologi	Vektor	PT. Indonesia Geospasial Tech (2020)
Peta <i>Land use</i>	Vektor	Priyono <i>et al.</i> (2020)
Peta Jalan Indonesia	Vektor	Priyono <i>et al.</i> (2020)

### Analisis Data Penelitian

Peta bahan penelitian diolah menggunakan aplikasi GRASS GIS dan QGIS. Data peta yang diolah perlu untuk dipotong terlebih dahulu sesuai dengan batasan wilayah studi yang digunakan, hal ini dilakukan agar pada saat komputasi dilakukan hanya peta dengan batasan wilayah studi saja yang diolah sehingga proses kalkulasi tidak memakan waktu yang lama. Peta bahan penelitian harus dirubah dalam format peta raster dengan resolusi satu detik dan dengan proyeksi peta longlat atau sama dengan format peta ASTER GDEM. Setelah peta data penelitian selesai dibuat, semua peta kemudian di simpan kembali dalam format ASCII, hal ini dilakukan untuk memperingan proses pengolahan data dalam *machine learning*.

### Penyusunan Script Model

Bahasa pemrograman *python* digunakan dalam menyusun *script* model pada penelitian ini. *Script* disusun berdasarkan hasil input features dan target. Data yang digunakan terdiri dari beberapa faktor pengondisi serta histori kejadian longsor. Beberapa sampel data digunakan dalam penelitian ini, sample tersebut pada tahap selanjutnya dibagi menjadi dua yaitu data *training* dan data *testing* yang dibagi secara acak. Pada tahap pengklasifikasian menggunakan *machine learning* setiap pengklasifikasi individu melatih semua data dan berkontribusi pada keputusan dan pada akhirnya *Voting Classifier Technique* menghasilkan solusi akhir.

## Validasi Model

Validasi merupakan tahapan yang penting karena data training yang diambil secara acak berkemungkinan terdapat kesalahan saat mengalami perubahan skenario pengacakan. Penelitian ini menggunakan *ten-folds cross validation* sebagai metode proses validasi. Metode ini membagi data *training* menjadi 10 bagian data (*folds*), dan melakukan *training* pada sembilan bagian data pertama dan satu bagian data sisanya dilakukan *testing*. Proses ini dilakukan bergiliran dengan setiap skenario *training* dan *testing* diambil satu bagian data secara bergantian.

## Pembuatan Peta Kerawanan Longsor

Peta kerawanan longsor dibentuk setelah proses pelatihan model longsor selesai, proses pembentukan peta kerawanan longsor dibuat dalam dua langkah utama langkah pertama, melakukan prediksi probabilitas longsor menggunakan model yang sudah disimpan. Langkah kedua, pengklasifikasian menggunakan metode *Natural Breaks* (Jenks) di dalam pemrograman *python* dan ditampilkan dalam aplikasi QGIS. Dalam penelitian ini dibuat lima kelas kerawanan longsor, yaitu sangat rendah, rendah, sedang, tinggi, dan sangat tinggi.

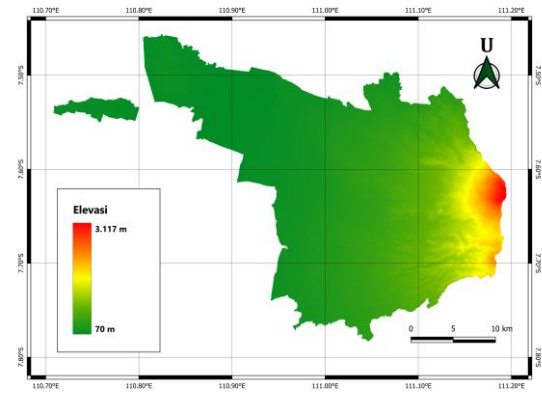
## Perhitungan *Landslide Density* Peta

*Landslide density* dihitung untuk mengevaluasi performa hasil klasifikasi kerawanan longsor yang dilakukan. *Landslide density* merupakan rasio dari persentase piksel longsor dan persentase kelas piksel pada masing-masing kelas rentan, seperti ditunjukkan pada Persamaan (3). Jika *landslide density* meningkat secara bertahap dari yang sangat rendah kelas sampai kelas sangat tinggi, maka hasil visualisasi peta kerawanan longsor dapat dikatakan baik.

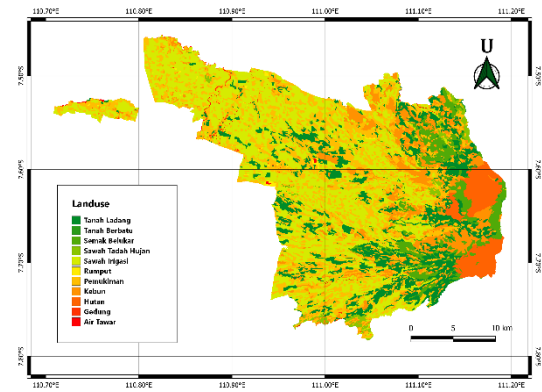
## 3. HASIL DAN PEMBAHASAN

### Hasil Analisis Data Penelitian

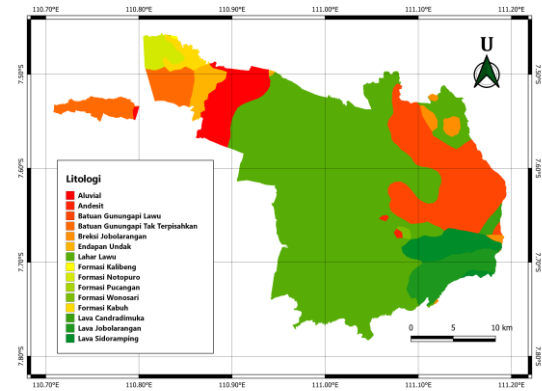
Hasil akhir dari pengolahan data sekunder terdiri dari sembilan peta faktor pengondisi dan peta *inventory* longsor. Peta yang digunakan dalam format raster dan ditampilkan seperti pada Gambar 5. Peta faktor pengondisi dan *inventory* longsor kemudian dikonversikan ke format ASCII Grid untuk mempermudah pengolahan *script python*.



(a)



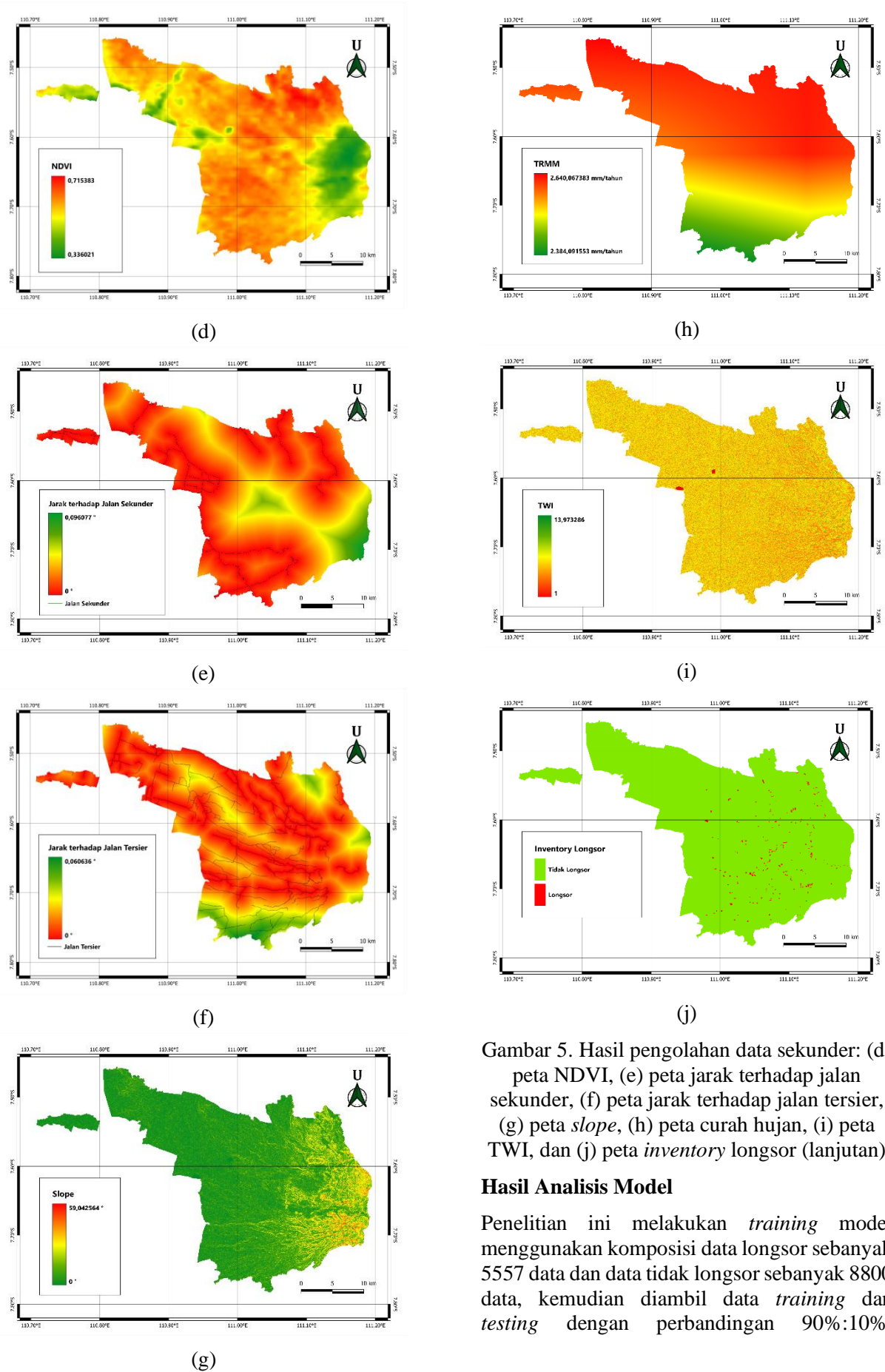
(b)



(c)

Gambar 5. Hasil pengolahan data sekunder: (a) peta elevasi, (b) peta *land use*, dan (c) peta litologi





Gambar 5. Hasil pengolahan data sekunder: (d) peta NDVI, (e) peta jarak terhadap jalan sekunder, (f) peta jarak terhadap jalan tersier, (g) peta *slope*, (h) peta curah hujan, (i) peta TWI, dan (j) peta *inventory* longsor (lanjutan)

### Hasil Analisis Model

Penelitian ini melakukan *training* model menggunakan komposisi data longsor sebanyak 5557 data dan data tidak longsor sebanyak 8800 data, kemudian diambil data *training* dan *testing* dengan perbandingan 90%:10%.

Kombinasi parameter yang digunakan pada penelitian ini dapat dilihat pada Tabel 2.

Tabel 2. Kombinasi parameter *machine learning*

Nama Metode	Parameter
Random Forest	<i>random_state</i> = 0, <i>n_estimators</i> = 750, <i>criterion</i> = 'entropy'
Support Vector Machine	<i>random_state</i> = 0, <i>Probability</i> = True, <i>gamma</i> = 0.005, <i>C</i> = 1
K-Nearest Neighbors	<i>n_neighbors</i> = 11, <i>weights</i> = 'distance', <i>leaf_size</i> = 20
Voting Classifier	<i>voting</i> = 'soft', <i>weights</i> = [1,1,1]

Model yang dimiliki memiliki nilai *AUC testing* sebesar 0,94702 dan memiliki nilai *AUC cross validation* sebesar 0,95816 Model tergolong dalam kategori model *Excellent*. Model yang sudah dibuat kemudian disimpan dalam format *joblib* agar bisa dipanggil kembali untuk melakukan prediksi.

### Hasil Klasifikasi Kerawanan Longsor

Metode *Natural Breaks* digunakan pada penelitian ini untuk mengklasifikasikan hasil

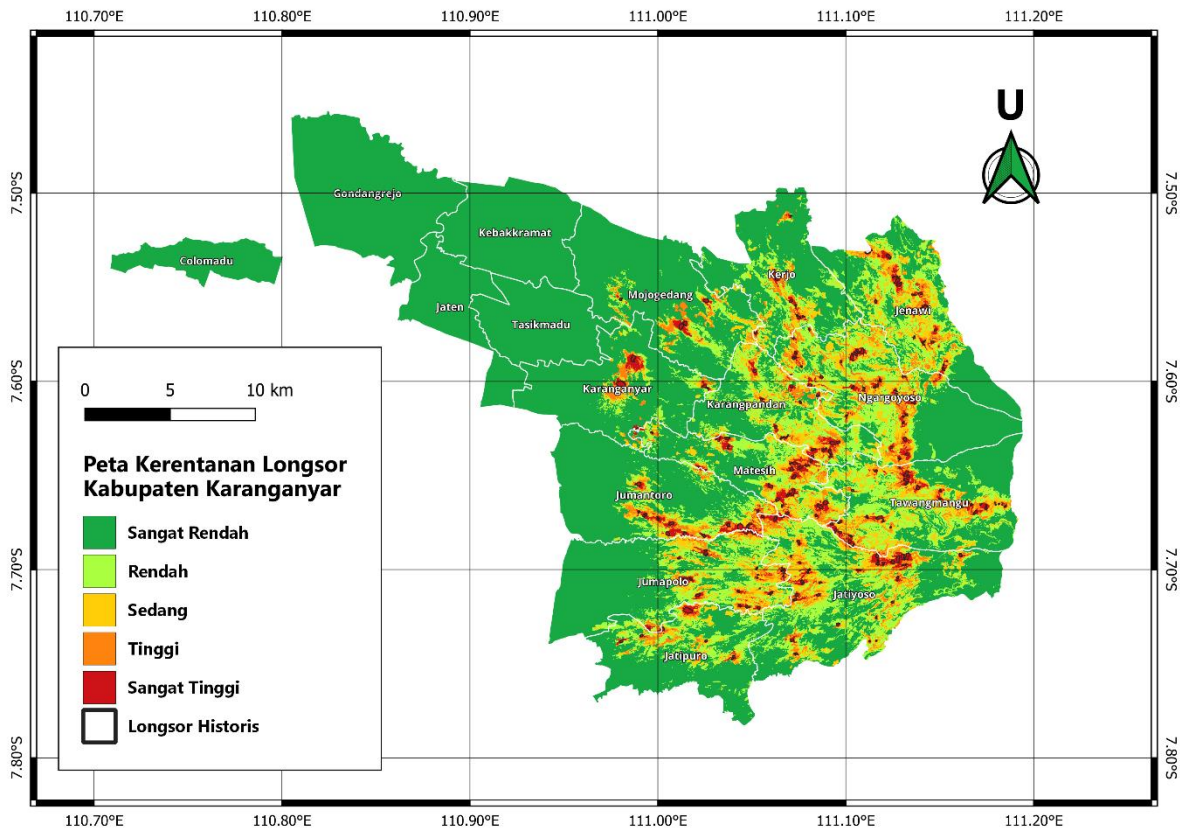
prediksi. Klasifikasi dibagi menjadi lima kelas, klasifikasi metode *Natural Breaks* dilakukan dengan menggunakan *script* di dalam pemrograman *python*. Hasil klasifikasi prediksi longsor yang diperoleh pada penelitian ini dapat dilihat pada Tabel 4.

Tabel 4. Batas-batas kelas

Keterangan	Probabilitas
Batas bawah kelas 1	0.000072
Batas atas kelas 1	0.087483
Batas atas kelas 2	0.244555
Batas atas kelas 3	0.46151
Batas atas kelas 4	0.722971
Batas atas kelas 5	0.996967

### Hasil Pemetaan Kerawanan Longsor

Hasil pengklasifikasian prediksi yang telah disimpan dipanggil menggunakan aplikasi GRASS GIS yang kemudian disimpan kembali dalam format raster (.tif). Peta kerawanan longsor kemudian dapat dibuka menggunakan aplikasi QGIS. Peta kerawanan longsor yang diperoleh dapat dilihat pada Gambar 6.



Gambar 6. Peta kerawanan longsor Kabupaten Karanganyar dengan metode *Voting Classifier Ensemble Technique*

### Hasil Perhitungan *Landslide Density*

Peta kerawanan longsor yang diperoleh perlu untuk dihitung *landslide density* yang dimiliki untuk melihat apakah memiliki visualisasi yang baik. Perhitungan *landslide density* dapat dilakukan menggunakan Persamaan (3). Hasil perhitungan *landslide density* dari peta kerawanan longsor yang telah dibuat dapat dilihat pada Tabel 5.

Tabel 5. Hasil perhitungan *landslide density*

Kelas kerawanan	<i>Landslide density</i>
Sangat Rendah	0
Rendah	0.00002
Sedang	0.00020
Tinggi	0.00261
Sangat Tinggi	0.19375

Pada Tabel 5 dapat dilihat bahwa peta kerawanan tanah longsor yang telah dibuat memiliki visualisasi prediksi spasial tanah longsor yang baik karena *landslide density* meningkat secara bertahap dari kelas sangat rendah ke kelas sangat tinggi.

### 4. SIMPULAN

Penelitian ini menghasilkan beberapa kesimpulan diantaranya yaitu:

- Peforma model terbaik memiliki nilai AUC sebesar 0,94702 dan *cross validation* sebesar 0,95816, hasil AUC yang diperoleh termasuk dalam kategori *Excellent* sehingga dapat dikatakan bahwa model yang dimiliki peforma yang baik untuk melakukan prediksi probabilitas longsor.
- Nilai *landslide density* peta yang dibuat meningkat secara bertahap dari kelas sangat rendah ke kelas sangat tinggi menunjukkan bahwa pemetaan yang dibuat memiliki visualisasi yang baik.

### 5. UCAPAN TERIMAKASIH

Terimakasih penulis ucapkan kepada civitas akademika UNS yang telah mendukung penulis hingga mencapai titik ini. Terimakasih penulis ucapkan kepada teman-teman sipil UNS yang sudah mendukung serta membantu dalam bentuk support materiil maupun mental. Terima kasih kepada semua pihak yang sudah mendukung dalam segala bentuk seperti dukungan dalam proses pengerjaan penulisan maupun dukungan berupa doa.

### 6. DAFTAR PUSTAKA

Achour, Y., & Pourghasemi, H. R. (2020). How do machine learning techniques help in increasing accuracy of landslide susceptibility maps? *Geoscience Frontiers*, 11(3), 871–883. <https://doi.org/10.1016/j.gsf.2019.10.001>

Akinci, H. (2022). Assessment of rainfall-induced landslide susceptibility in Artvin, Turkey using machine learning techniques. *Journal of African Earth Sciences*, 191(April), 104535. <https://doi.org/10.1016/j.jafrearsci.2022.104535>

Benbouras, M. A. (2022). Hybrid meta-heuristic machine learning methods applied to landslide susceptibility mapping in the Sahel-Algiers. *International Journal of Sediment Research*, 37(5), 601–618. <https://doi.org/10.1016/j.ijsrc.2022.04.003>

BNPB. (n.d.). *Geoportal Data Bencana Indonesia*. Diambil 9 Maret 2023, dari <https://gis.bnpb.go.id/>

Bragagnolo, L., Silva, R. V. d., & Grzybowski, J. M. V. (2020). Artificial neural network ensembles applied to the mapping of landslide susceptibility. *Catena*, 184(September 2019), 104240. <https://doi.org/10.1016/j.catena.2019.104240>

CDC. (2023). *Jenks natural breaks classification method - Health, United States*. <https://www.cdc.gov/nchs/hus/sources-definitions/jenks-natural-breaks.htm>

esri. (2023). *Data classification methods—ArcGIS Pro | Documentation*. <https://pro.arcgis.com/en/pro-app/latest/help/mapping/layer-properties/data-classification-methods.htm>

Fitria, I. (2021). *Elimination Of Factors In Landslide Vulnerability Risk Mapping Model Based On Artificial Intelligence System*.

Hammad Khaliq, A., Basharat, M., Talha Riaz, M., Tayyib Riaz, M., Wani, S., Al-Ansari, N., Ba Le, L., & Thi Thuy Linh, N. (2023). Spatiotemporal landslide susceptibility mapping using machine learning models: A case study from district Hattian Bala, NW Himalaya, Pakistan. *Ain Shams Engineering Journal*, 14(3), 101907. <https://doi.org/10.1016/j.asej.2022.101907>

Hoarau, A., Martin, A., Dubois, J. C., & Le Gall, Y. (2023). Evidential Random Forests. *Expert Systems with Applications*, 230(May), 120652. <https://doi.org/10.1016/j.eswa.2023.120652>

Khan, M. A., Iqbal, N., Imran, Jamil, H., & Kim, D. H. (2023). An optimized ensemble prediction model using AutoML based on soft voting classifier for network intrusion detection. *Journal of Network and Computer Applications*, 212(November 2022), 103560. <https://doi.org/10.1016/j.jnca.2022.103560>

Marzuki, T. (2021). *REVISI Skripsi Thoriq Marzuki*

- (30092021). Universitas Sebelas Maret.
- Meinhardt, M., Fink, M., & Tünschel, H. (2015). Landslide susceptibility analysis in central Vietnam based on an incomplete landslide inventory: Comparison of a new method to calculate weighting factors by means of bivariate statistics. *Geomorphology*, 234, 80–97.  
<https://doi.org/10.1016/j.geomorph.2014.12.042>
- Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., Avtar, R., & Abderrahmane, B. (2020). Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Science Reviews*, 207(September 2019), 103225.  
<https://doi.org/10.1016/j.earscirev.2020.103225>
- Naji, M. A., Filali, S. El, Bouhhal, M., Benlahmar, E. H., Abdelouahid, R. A., & Debauche, O. (2021). Breast Cancer Prediction and Diagnosis through a New Approach based on Majority Voting Ensemble Classifier. *Procedia Computer Science*, 191, 481–486.  
<https://doi.org/10.1016/j.procs.2021.07.061>
- NASA. (2013). *Earthdata Search | Earthdata Search*.  
<https://search.earthdata.nasa.gov/search>
- NASA. (2020). *GES DISC*.  
<https://disc.gsfc.nasa.gov/>
- NASA. (2022). *MODIS Web*.  
<https://modis.gsfc.nasa.gov/>
- Pham, B. T., Tien Bui, D., Prakash, I., & Dholakia, M. B. (2017). Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *Catena*, 149, 52–63.  
<https://doi.org/10.1016/j.catena.2016.09.007>
- Priyono, K. D., Jumadi, Saputra, A., & Fikriyah, V. N. (2020). Risk analysis of landslide impacts on settlements in Karanganyar, Central Java, Indonesia. *International Journal of GEOMATE*, 19(73), 100–107.  
<https://doi.org/10.21660/2020.73.34128>
- PT. Indonesia Geospasial Tech. (2020). *√ Download Shapefile Geologi Seluruh Indonesia - INDONESIA GEOSPASIAL*.  
<https://www.indonesia-geospasial.com/2020/03/download-data-shapefile-shp-geologi-se.html>
- Rai, N., Kaushik, N., Kumar, D., Raj, C., & Ali, A. (2022). Mortality prediction of COVID-19 patients using soft voting classifier. *International Journal of Cognitive Computing in Engineering*, 3(September), 172–179.  
<https://doi.org/10.1016/j.ijcce.2022.09.001>
- Sameen, M. I., Pradhan, B., Bui, D. T., & Alamri, A. M. (2020). Systematic sample subdividing strategy for training landslide susceptibility models. *Catena*, 187(October 2019), 104358.  
<https://doi.org/10.1016/j.catena.2019.104358>
- Scikit-learn. (2023). *1.11. Ensemble methods — scikit-learn 1.2.2 documentation*.  
<https://scikit-learn.org/stable/modules/ensemble.html>
- Yang, C., Liu, L. L., Huang, F., Huang, L., & Wang, X. M. (2022). Machine learning-based landslide susceptibility assessment with optimized ratio of landslide to non-landslide samples. *Gondwana Research*, xxx.  
<https://doi.org/10.1016/j.gr.2022.05.012>
- Youssef, A. M., & Pourghasemi, H. R. (2021). Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia. *Geoscience Frontiers*, 12(2), 639–655.  
<https://doi.org/10.1016/j.gsf.2020.05.010>
- Zhao, S., Dai, F., Deng, J., Wen, H., Li, H., & Chen, F. (2023). Insights into landslide development and susceptibility in extremely complex alpine geoenvironments along the western Sichuan–Tibet Engineering Corridor, China. *Catena*, 227(24), 107105.  
<https://doi.org/10.1016/j.catena.2023.107105>
- Zhou, X., Wen, H., Zhang, Y., Xu, J., & Zhang, W. (2021). Landslide susceptibility mapping using hybrid random forest with GeoDetector and RFE for factor optimization. *Geoscience Frontiers*, 12(5), 101211.  
<https://doi.org/10.1016/j.gsf.2021.101211>