

LAPORAN
PENELITIAN INTERNAL DOSEN



PREDIKSI PARAMETER KEKUATAN GESER TANAH BERBASIS DATA
MENGGUNAKAN PEMBELAJARAN MESIN

Tim Peneliti:

Ir. Ichwan Hadi Saputra, S.T., M.T
Dr. Eko Wahyu Abryandoko., S.Pd.,MT

Dibiayai oleh:

Universitas Bojonegoro

Periode 1 Tahun Anggaran 2025/2026

No. Kontrak: 059/LPPM-LIT/UB/XI/2025

UNIVERSITAS BOJONEGORO

2025

HALAMAN PENGESAHAN
LAPORAN PENELITIAN MANDIRI

- 1. Judul Penelitian** : Prediksi Parameter Kekuatan Geser Tanah Berbasis Data Menggunakan Pembelajaran Mesin
- 2. Ketua Peneliti**
- a. Nama Peneliti : Ir. Ichwan Hadi Saputra, ST.,MT.
 - b. NIDN : 07 1207 9205
 - c. Program Studi : Teknik Sipil
 - d. E-mail : ichwanhs@gmail.com
 - e. Bidang Keilmuan : Teknik Sipil
- 3. Anggota Peneliti 1**
- a. Nama (Dosen) : Dr. Eko Wahyu Abryandoko., S.Pd.,MT
 - b. NIDN : 07 1011 9102
 - c. Program Studi : Teknik Industri
 - d. E-mail : abryandoko@gmail.com
 - e. Bidang Keilmuan : Rekayasa Sistem Pengembangan Produk
- Anggota Peneliti 2**
- a. Nama (Mahasiswa) : Putri Sekar Arumdalu
 - b. NIM : 24222011174
 - c. Program Studi : Teknik Sipil
 - d. E-mail :
 - e. Bidang Keilmuan : Teknik Sipil
4. Jangka Waktu Penelitian : 6 Bulan
6. Lokasi Penelitian : Kecamatan Bojonegoro Kabupaten Bojonegoro
7. Dana Diusulkan : 3.500.000,-

Bojonegoro, 20 Februari 2026

Mengetahui,

Ketua LPPM Universitas Bojonegoro

Ketua Pengusul,

Laily Agustina Rahmawati, S.Si., M.Sc.
NIDN 07 2108 8601


Ir. Ichwan Hadi Saputra, ST.MT
NIDN. 07 1207 9205

KATA PENGANTAR

Syukur Alhamdulillah ke hadirat Allah SWT atas segala limpahan rahmat dan hidayahnya maka laporan penelitian yang berjudul: Prediksi Modulus Deformasi Tanah Berdasarkan Properti Indeks Menggunakan Pembelajaran Mesin Ensemble.

Penelitian ini diharapkan dapat menjadi salah satu sumbangsih bagi Lembaga Penelitian dan Pengabdian Masyarakat Universitas Bojonegoro. Atas selesainya penulisan penelitian ini, penulis menyampaikan ucapan terima kasih yang tak terhingga kepada: Yang terhormat Bapak Dekan Fakultas Sains dan Teknik yang telah memberikan ijin dan dukungan dalam penyelesaian proposal penelitian ini.

Pihak-pihak yang tidak dapat disebutkan satu persatu yang telah membantu penulis menyelesaikan proposal penelitian ini. Akhirnya, penulis akui hanya dengan kebesaran ALLAH SWT, penulis dapat menyelesaikan proposal penelitian ini. Semoga ALLAH SWT berkenan memberikan balasan atas semua jasa, budi mulia serta amal perbuatan yang telah dicurahkan tersebut sebagai amalan sholeh fiddini wal akhirah, Amin.

Bojonegoro, 20 Februari 2026

Ir. Ichwan Hadi Saputra, S.T., M.T

ABSTRAK

Modulus deformasi tanah (E_s) merupakan parameter geoteknik fundamental yang berperan penting dalam analisis penurunan dan perencanaan fondasi, namun penentuan nilainya secara langsung melalui pengujian laboratorium maupun lapangan seringkali memerlukan biaya dan waktu yang signifikan. Penelitian ini mengusulkan pendekatan berbasis ensemble machine learning untuk memprediksi E_s secara akurat menggunakan parameter indeks tanah yang umum tersedia, meliputi void ratio (e), degree of saturation (S_r), liquid limit (LL), plastic limit (PL), dan koefisien kompresibilitas (α_{1-2}). Dataset yang digunakan bersumber dari repositori yang dipublikasikan oleh Sun, yang memuat data pengujian laboratorium pada tanah lempung dengan variasi karakteristik yang beragam. Beberapa model ensemble dikembangkan dan dibandingkan, mencakup Random Forest, HistGradientBoosting, XGBoost, Voting Regressor, dan Stacking Regressor. Proses hyperparameter tuning dilakukan menggunakan RandomizedSearchCV dengan skema 5-fold cross-validation untuk memperoleh konfigurasi model yang optimal. Hasil evaluasi menunjukkan bahwa Random Forest tampil sebagai model terbaik dengan nilai R^2 sebesar 0,983 dan RMSE sebesar 0,3073 MPa pada data uji, didukung oleh rata-rata CV R^2 sebesar 0,9651 yang mencerminkan stabilitas generalisasi yang tinggi. Analisis feature importance secara konsisten mengidentifikasi α_{1-2} sebagai prediktor paling dominan terhadap E_s , dengan kontribusi kepentingan melebihi 0,96 pada kedua metode evaluasi MDI/Gain dan Permutation Importance. Temuan ini menunjukkan bahwa model ensemble machine learning yang dikembangkan mampu menjadi alat estimasi E_s yang akurat, efisien, dan praktis dalam mendukung perencanaan geoteknik berbasis data.

Kata Kunci: *soil deformation modulus, ensemble machine learning, random forest, feature importance, geotechnical engineering, index properties, soil compressibility*

DAFTAR ISI

HALAMAN PENGESAHAN LAPORAN PENELITIAN MANDIRI	2
KATA PENGANTAR.....	3
ABSTRAK	4
DAFTAR ISI.....	5
DAFTAR GAMBAR.....	6
DAFTAR TABEL	7
BAB I PENDAHULUAN.....	8
A. Latar Belakang	8
B. Rumusan Masalah	9
C. Tujuan Penelitian	10
D. Manfaat Penelitian	10
E. Urgensi Penelitian.....	11
BAB II LANDASAN TEORI	12
A. Parameter Kekuatan Geser Tanah	12
B. Machine Learning	12
C. Ensemble Learning.....	13
D. Data Geoteknik Global.....	13
E. Penelitian Tedahulu.....	14
BAB III METODE PENELITIAN	17
A. Tempat dan Waktu Penelitian	17
B. Alat dan Bahan Penelitian	17
C. Jenis Penelitian	19
D. Identifikasi dan klasifikasi variabel.....	20
E. Alur Penelitian	22
BAB IV HASIL DAN PEMBAHASAN	23
A. Hyperparameter Tuning.....	23
B. Performance of All Models	24
C. Cross-Validation Stability Analysis	26
D. Residual Analysis and Prediction Bias	27
E. Feature Importance Analysis	28
BAB V KESIMPULAN	30
A. Kesimpulan dan Saran.....	30
DAFTAR PUSTKA.....	32
LAMPIRAN.....	34

DAFTAR GAMBAR

Gambar 3. 1 Software VS Code.....	18
Gambar 3. 2 Bahasa Perograman <i>Python</i>	19
Gambar 3. 3 Alur Penelitian	22

DAFTAR TABEL

Tabel 2. 1 variabel input	Error! Bookmark not defined.
Tabel 4. 1 Hasil Evaluasi Kinerja	Error! Bookmark not defined.
Tabel 4. 2 Classification Report	Error! Bookmark not defined.
Tabel 4. 3 Validasi Model	Error! Bookmark not defined.
Tabel 4. 4 Optimization Results	Error! Bookmark not defined.

BAB I

PENDAHULUAN

A. Latar Belakang

Modulus deformasi tanah (E_s) merupakan salah satu parameter geoteknik paling fundamental yang dibutuhkan dalam analisis perilaku tanah di bawah beban. Parameter ini menjadi dasar utama dalam perhitungan penurunan (settlement) pondasi, analisis stabilitas lereng, serta perancangan struktur bawah tanah [1], [2]. Akurasi estimasi E_s secara langsung mempengaruhi keandalan desain geoteknik, mengingat kesalahan dalam prediksi settlement dapat menyebabkan kerusakan struktural yang serius, terutama pada tanah lempung dengan kompresibilitas tinggi [3]. Secara konvensional, nilai E_s diperoleh melalui pengujian oedometer di laboratorium, yang memerlukan prosedur pengambilan sampel tanah tak terganggu, persiapan spesimen yang cermat, serta durasi pengujian yang relatif panjang. Selain itu, variabilitas alami karakteristik tanah dari satu lokasi ke lokasi lainnya menambah kompleksitas dalam memperoleh nilai E_s yang representatif, sehingga pengujian langsung seringkali menjadi tidak efisien dari segi waktu maupun biaya, khususnya pada tahap investigasi awal proyek [4]. Keterbatasan pendekatan konvensional mendorong berkembangnya berbagai metode empiris yang mengkorelasikan E_s dengan parameter indeks tanah yang lebih mudah diperoleh, seperti batas cair (LL), batas plastis (PL), dan angka pori (e). Namun demikian, korelasi empiris yang ada umumnya dikembangkan dari dataset lokal yang terbatas sehingga tidak dapat digeneralisasi secara luas untuk berbagai kondisi tanah [5]. Di samping itu, hubungan antara parameter indeks tanah dan E_s bersifat non-linear dan dipengaruhi oleh interaksi kompleks antar variabel, sehingga pendekatan regresi linier konvensional kerap gagal menangkap pola tersebut secara memadai [6]. Ketidakefisienan pengujian langsung dikombinasikan dengan keterbatasan model empiris menciptakan kebutuhan mendesak akan pendekatan prediksi yang lebih akurat, efisien, dan dapat diandalkan dalam berbagai kondisi geologi. Dalam dua dekade terakhir, machine learning (ML) telah mengalami perkembangan pesat dan terbukti mampu menangani permasalahan kompleks di berbagai bidang rekayasa, termasuk geoteknik [7], [8]. Berbagai studi telah mengaplikasikan teknik ML untuk memprediksi beragam parameter tanah, seperti indeks kompresi tanah lempung [9], [10], modulus resiliensi tanah dasar [11], parameter pemadatan tanah [12], serta kuat geser tanah [13]. Di antara berbagai teknik ML yang tersedia, metode ensemble—yang menggabungkan beberapa model dasar untuk menghasilkan prediksi akhir yang lebih

robust—terbukti lebih unggul dibandingkan model tunggal karena kemampuannya dalam mereduksi varians, mengurangi risiko overfitting, dan meningkatkan kemampuan generalisasi model terhadap data baru [14]. Algoritma seperti Random Forest, Gradient Boosting, dan XGBoost telah memperlihatkan performa yang sangat kompetitif pada berbagai dataset geoteknik dengan hubungan non-linear yang kompleks [15]. Meskipun penerapan ML di bidang geoteknik telah berkembang pesat, terdapat beberapa celah penelitian (research gap) yang masih belum tertangani secara memadai. Sebagian besar studi prediksi parameter deformasi tanah masih mengandalkan pendekatan model tunggal, sementara studi yang secara khusus mengeksplorasi perbandingan komprehensif berbagai arsitektur ensemble untuk prediksi Es berbasis parameter indeks tanah masih sangat terbatas [5], [9]. Selain itu, banyak penelitian yang ada tidak menerapkan proses hyperparameter tuning secara sistematis dan evaluasi kestabilan model menggunakan skema k-fold cross-validation yang robust, sehingga generalisabilitas model yang dilaporkan masih dipertanyakan [13], [16]. Hal ini mendorong diperlukannya penelitian yang lebih komprehensif dalam mengembangkan, membandingkan, dan mengevaluasi model ensemble machine learning secara sistematis untuk prediksi Es dari parameter indeks tanah.

B. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah:

1. pendekatan empiris konvensional dalam estimasi modulus deformasi tanah (Es) memiliki keterbatasan dalam hal generalisasi dan kemampuan menangkap hubungan non-linear antar parameter indeks tanah, sehingga diperlukan metode alternatif yang lebih adaptif dan akurat
2. Pengujian laboratorium langsung untuk memperoleh nilai Es membutuhkan waktu, biaya, dan ketersediaan spesimen tanah yang tidak selalu terpenuhi dalam kondisi proyek geoteknik yang memiliki keterbatasan sumber daya
3. Belum tersedianya kajian komprehensif yang membandingkan berbagai model ensemble machine learning secara sistematis—disertai hyperparameter tuning dan validasi silang yang robust—untuk prediksi Es berbasis parameter indeks tanah menjadi celah yang perlu segera dijawab melalui penelitian ini

4. Sejauh mana kontribusi relatif masing-masing parameter indeks tanah terhadap prediksi Es belum banyak dikaji secara kuantitatif menggunakan metode analisis feature importance yang komprehensif.

C. Tujuan Penelitian

1. Mengembangkan dan membandingkan beberapa model ensemble machine learning, meliputi Random Forest, HistGradientBoosting, XGBoost, Voting Regressor, dan Stacking Regressor, untuk prediksi modulus deformasi tanah (Es) berbasis parameter indeks tanah yang umum tersedia dari pengujian laboratorium rutin
2. mengevaluasi akurasi dan kestabilan setiap model secara komprehensif menggunakan metrik R^2 , MAE, dan RMSE pada data uji, serta skema 5-fold cross-validation untuk memastikan generalisabilitas model yang dihasilkan
3. melakukan proses hyperparameter tuning secara sistematis menggunakan RandomizedSearchCV guna memperoleh konfigurasi model yang optimal dan dapat dibandingkan secara adil.
4. menganalisis kontribusi relatif masing-masing parameter indeks tanah terhadap prediksi Es melalui analisis feature importance menggunakan metode MDI/Gain dan Permutation Importance secara bersamaan, sehingga diperoleh pemahaman yang lebih mendalam tentang variabel yang paling berpengaruh dalam pemodelan deformasi tanah.

D. Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat baik secara teoritis maupun praktis. Secara teoritis, hasil penelitian ini berkontribusi dalam memperkaya khazanah ilmu geoteknik berbasis data, khususnya dalam demonstrasi penerapan metode ensemble machine learning untuk prediksi parameter deformasi tanah. Temuan mengenai dominasi parameter α_1-2 sebagai prediktor utama Es juga memperkuat pemahaman teoritis tentang keterkaitan fisis antara koefisien kompresibilitas dan modulus deformasi tanah, yang dapat menjadi dasar bagi pengembangan model prediksi geoteknik berbasis data di masa mendatang.

Secara praktis, model ensemble yang dikembangkan dalam penelitian ini berpotensi digunakan sebagai alat bantu estimasi awal nilai Es dalam perencanaan geoteknik, khususnya pada tahap investigasi awal ketika pengujian laboratorium secara

langsung tidak memungkinkan atau membutuhkan biaya yang tinggi. Dengan memanfaatkan parameter indeks tanah yang mudah dan murah diperoleh, model ini dapat membantu insinyur geoteknik dalam membuat keputusan awal yang lebih terinformasi, menghemat waktu investigasi, serta mempercepat proses perencanaan fondasi dan analisis settlement tanpa mengorbankan keandalan estimasi yang dihasilkan.

E. Urgensi Penelitian

Urgensi penelitian ini didasarkan pada kesenjangan yang signifikan antara kebutuhan praktis akan estimasi E_s yang cepat dan akurat di lapangan dengan keterbatasan metode yang tersedia saat ini. Dalam konteks pembangunan infrastruktur yang terus berkembang pesat, khususnya di negara-negara berkembang di mana investigasi tanah yang komprehensif seringkali terkendala oleh keterbatasan waktu dan anggaran, tersedianya model prediksi yang andal berbasis parameter indeks tanah sederhana menjadi sangat krusial [4], [7].

Lebih jauh, perkembangan pendekatan data-driven dalam rekayasa geoteknik telah membuka peluang besar untuk meningkatkan efisiensi dan akurasi prediksi parameter tanah secara signifikan [2], [8]. Namun demikian, potensi ini belum sepenuhnya dimanfaatkan untuk prediksi E_s secara spesifik, terutama dengan pendekatan ensemble yang disertai evaluasi sistematis dan analisis interpretabilitas model [5], [16]. Tanpa adanya penelitian yang menjawab celah ini, praktisi geoteknik akan terus bergantung pada metode konvensional yang memiliki keterbatasan inheren, sementara kemajuan teknologi ML yang tersedia belum dioptimalkan untuk kebutuhan praktis di lapangan. Oleh karena itu, penelitian ini memiliki urgensi yang tinggi untuk segera dilaksanakan guna menghasilkan kontribusi ilmiah yang relevan dan berdampak langsung bagi praktik rekayasa geoteknik modern.

BAB II

LANDASAN TEORI

A. Parameter Kekuatan Geser Tanah

Dalam bidang teknik sipil, khususnya rekayasa geoteknik, pemahaman terhadap perilaku tanah merupakan fondasi utama dalam perancangan struktur seperti fondasi, dinding penahan tanah, dan stabilitas lereng. Tiga parameter mekanis utama yang mendefinisikan kekuatan geser tanah adalah koefisien kohesi (c), sudut geser internal (ϕ), dan modulus elastisitas (E).

Koefisien kohesi (c) adalah gaya ikat antar partikel tanah yang memberikan kekuatan pada tanah meskipun tidak ada tekanan normal yang bekerja. Parameter ini sangat dominan pada tanah kohesif seperti lempung. Sedangkan sudut geser internal (ϕ) mencerminkan resistensi tanah terhadap geser akibat dari interlocking (penguncian) antar butiran tanah, yang lebih dominan pada tanah non-kohesif seperti pasir. Hubungan antara tegangan geser maksimum yang dapat ditahan oleh tanah (τ) dengan tegangan normal (σ) dijelaskan oleh hukum Coulomb. Kedua parameter ini secara langsung menentukan daya dukung tanah dan faktor keamanan pada analisis stabilitas lereng (Das, 2019).

Modulus elastisitas (E), atau modulus Young, menggambarkan kekakuan tanah, yaitu hubungan antara tegangan dan regangan pada kondisi elastis. Parameter ini sangat penting dalam prediksi deformasi atau penurunan struktur yang berdiri di atas tanah. Nilai E yang rendah menunjukkan tanah yang mudah mengalami penurunan, yang merupakan pertimbangan kritis dalam perencanaan gedung tinggi atau infrastruktur sensitif terhadap gerakan tanah (Bowles, 1996). Dalam konteks tanah kohesif, ketiga parameter ini saling berkaitan dan dipengaruhi oleh sifat fisik tanah seperti kadar air, komposisi butiran, dan indeks plastisitas.

B. Machine Learning

Perkembangan teknologi informasi telah membawa machine learning (ML) menjadi alat bantu yang semakin penting dalam menyelesaikan masalah kompleks di bidang teknik sipil. ML memungkinkan sistem untuk belajar dari data historis dan mengidentifikasi pola-pola tersembunyi tanpa diprogram secara eksplisit, menjadikannya sangat cocok untuk memodelkan fenomena non-linear yang umum ditemui dalam material geoteknik (Feng et al., 2020).

Pendekatan ML dapat dikategorikan menjadi supervised learning dan unsupervised learning. Penelitian ini menggunakan pendekatan supervised learning, di mana model dilatih menggunakan dataset yang memiliki input (sifat fisik tanah) dan output (parameter

kekuatan geser) yang sudah diketahui. Tujuan dari pelatihan ini adalah agar model dapat membuat prediksi yang akurat untuk data baru yang belum pernah dilihat sebelumnya. Penerapan ML dalam teknik sipil telah berhasil digunakan untuk berbagai aplikasi, termasuk prediksi kuat tekan beton, deteksi kerusakan struktur, dan optimasi desain konstruksi (Asteris & Mokos, 2020).

C. Ensemble Learning

Ensemble learning adalah paradigma dalam machine learning yang menggabungkan beberapa model pembelajaran (base learners) untuk menghasilkan satu model prediktif yang lebih akurat dan robust dibandingkan model individual. Prinsip dasarnya adalah bahwa kombinasi dari banyak "opini" (prediksi) akan lebih baik daripada satu opini tunggal, terutama jika model-model tersebut memiliki kesalahan yang tidak berkorelasi (Zhou, 2012). Beberapa algoritma ensemble yang paling populer dan efektif adalah:

1. Random Forest (RF), Algoritma ini membangun sejumlah besar pohon keputusan (decision tree) yang dilatih pada subset acak dari data pelatihan (bagging). Keputusan akhir diambil berdasarkan voting mayoritas (klasifikasi) atau rata-rata (regresi) dari semua pohon. RF dikenal karena kekuatannya yang tangguh terhadap overfitting dan kemampuannya untuk menangani data dengan banyak fitur (Breiman, 2001).
2. Extreme Gradient Boosting (XGBoost) adalah implementasi dari algoritma gradient boosting, yang membangun model secara bertahap dengan fokus pada contoh-contoh yang sulit diprediksi oleh model sebelumnya. Algoritma ini dioptimalkan untuk kecepatan dan performa, sering kali menjadi pemenang dalam kompetisi data science (Chen & Guestrin, 2016). Studi oleh Zhang et al. (2020) menunjukkan bahwa XGBoost unggul dalam memprediksi sifat mekanik beton.
3. Stacked Generalization (Stacking) adalah arsitektur ensemble tingkat lanjut di mana prediksi dari beberapa base learners (misalnya, RF dan XGBoost) digunakan sebagai input untuk sebuah meta-learner (model tingkat kedua). Meta-learner ini kemudian belajar cara terbaik untuk menggabungkan prediksi-prediksi tersebut. Stacking sering menghasilkan performa tertinggi karena kemampuannya untuk memanfaatkan kekuatan masing-masing base learner (Wolpert, 1992).

D. Data Geoteknik Global

Ketersediaan basis data geoteknik global yang luas merupakan prasyarat utama bagi keberhasilan penelitian berbasis data seperti ini. Meskipun tidak ada satu database pusat yang menyatukan semua data geoteknik dunia, banyak proyek dan publikasi ilmiah yang menyediakan akses ke dataset hasil pengujian tanah. Salah satu sumber data terstruktur

yang dapat dieksplorasi adalah Soil Survey Geographic Database (SSURGO) dari Amerika Serikat, yang mencakup data observasi lapangan dan laboratorium dari ratusan ribu lokasi (USDA, 2023). Selain itu, peneliti sering mengumpulkan data dari berbagai literatur internasional untuk membentuk dataset harmonisasi yang representatif (Pál et al., 2016; Zhang et al., 2021). Penggunaan dataset global memungkinkan model yang dikembangkan memiliki kemampuan generalisasi yang lebih baik, sehingga dapat diaplikasikan pada berbagai kondisi tanah di berbagai wilayah, termasuk Indonesia.

E. Penelitian Tedahulu

Perkembangan teknologi pembelajaran mesin (machine learning) dalam bidang rekayasa geoteknik telah menunjukkan kemajuan yang signifikan dalam satu dekade terakhir. Banyak peneliti telah berhasil mengaplikasikan pendekatan berbasis data untuk memprediksi parameter-parameter mekanis tanah dengan akurasi yang tinggi. Berikut ini adalah rangkuman enam penelitian terdahulu terkini yang menjadi dasar dan pembanding bagi penelitian ini.

1. Zhang et al. (2021) melakukan studi komprehensif untuk memprediksi modulus elastisitas sekunder (E_s) pada lempung lunak menggunakan metode machine learning. Studi yang diterbitkan di *Marine Georesources & Geotechnology* (Q1 Scopus) ini menggunakan dataset dari pengujian laboratorium dan lapangan di wilayah Asia Timur. Model yang diuji meliputi Random Forest (RF), Support Vector Regression (SVR), dan Artificial Neural Network (ANN). Hasil penelitian menunjukkan bahwa model RF memberikan performa terbaik dengan koefisien korelasi (R) mencapai 0.91 dan Mean Squared Error (MSE) sebesar 0.13, membuktikan efektivitas pendekatan ensemble dalam memprediksi parameter deformasi tanah (Zhang et al., 2021).
2. Pál et al. (2016) mengembangkan model Artificial Neural Network (ANN) multi-output untuk memprediksi secara simultan koefisien kohesi (c) dan sudut geser internal (ϕ) dari komposisi butiran tanah. Meskipun studi ini sedikit lebih lama, metodologinya sangat relevan karena merupakan salah satu pionir dalam pendekatan multi-output untuk parameter kekuatan geser. Mereka menemukan bahwa model ANN multi-output mampu mencapai $R^2 = 0.95$ untuk c dan $R^2 = 0.90$ untuk ϕ , serta menunjukkan bahwa pendekatan ini lebih unggul dibanding model regresi terpisah karena dapat menangkap hubungan silang antara kedua parameter (Pál et al., 2016). Temuan ini menjadi dasar kuat untuk pendekatan simultan dalam penelitian ini.

3. Dinh et al. (2022) mempublikasikan penelitian di *Geoscience Frontiers* mengenai prediksi *residual friction angle* (ϕ_r) pada tanah liat menggunakan *hybrid deep learning* berbasis GrowNet. Studi ini menggunakan dataset harmonisasi dari 400 sampel tanah dari berbagai lokasi global. Model GrowNet, yang menggabungkan Long Short-Term Memory (LSTM) dan Gradient Boosting Machine (GBM), mencapai R^2 sebesar 0.94, mengungguli model XGBoost dan SVR. Analisis SHAP yang digunakan juga mengidentifikasi Clay Fraction (CF) dan Plasticity Index (PI) sebagai fitur dominan, yang selaras dengan prinsip mekanika tanah (Dinh et al., 2022).
4. Li et al. (2020) mengembangkan model stacked ensemble untuk memprediksi porositas dan permeabilitas tanah menggunakan citra mikro-CT. Studi yang dipublikasikan di *Journal of Contaminant Hydrology* (Q1 Scopus) ini menggunakan enam base learners (termasuk RF, GB, dan DNNs) dengan meta-learner linear dan non-linear. Hasilnya menunjukkan bahwa model stacked mampu meningkatkan akurasi prediksi hingga 8–12% dibanding model individual, dengan R^2 mencapai 0.93 untuk porositas. Penelitian ini mendemonstrasikan potensi besar arsitektur stacking dalam tugas prediksi multi-output di geoteknik (Li et al., 2020).
5. Pham et al. (2019) melakukan perbandingan antara lima algoritma ML—Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF), k-Nearest Neighbors (k-NN), dan Gradient Boosting—for predicting the compressive strength of geopolymer concrete. Meskipun konteksnya beton geopolimer, metodologi dan temuan tentang efektivitas Gradient Boosting sangat relevan. Studi yang diterbitkan di *Materials* (Q1 Scopus) ini menemukan bahwa Gradient Boosting memberikan performa terbaik dengan $R^2 = 0.976$, menunjukkan keunggulan model ensemble berbasis boosting dalam menangani data non-linear kompleks (Pham et al., 2019).
6. Asteris et al. (2021) mengembangkan model prediktif untuk cohesion (c) dan friction angle (ϕ) menggunakan gene expression programming (GEP) dan ANN. Penelitian yang dipublikasikan di *Innovative Infrastructure Solutions* (Scopus-indexed) ini menggunakan data dari uji triaksial dan direct shear. Mereka menemukan bahwa model GEP memiliki interpretabilitas tinggi dan akurasi baik ($R^2 > 0.85$), sementara ANN memberikan akurasi tertinggi namun bersifat "black box". Studi ini menyoroti pentingnya keseimbangan antara akurasi dan interpretasi dalam aplikasi rekayasa (Asteris et al., 2021).

Secara keseluruhan, penelitian-penelitian terdahulu tersebut menunjukkan bahwa pendekatan ensemble learning (terutama RF dan XGBoost) konsisten memberikan performa superior, prediksi parameter geoteknik seperti c , ϕ , dan E dapat dilakukan dengan akurasi tinggi menggunakan ML, pendekatan multi-output dan stacking memiliki potensi untuk meningkatkan kinerja model secara signifikan dan analisis feature importance dapat memberikan wawasan teknis yang berharga bagi insinyur sipil. Namun, masih terdapat kesenjangan dalam literatur, yaitu kurangnya penelitian yang secara eksplisit memprediksi ketiga parameter utama tanah kohesif (c , ϕ , dan E) secara simultan menggunakan model ensemble berbasis data global. Penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan mengintegrasikan pendekatan multi-output, ensemble learning, dan analisis interpretabilitas untuk menghasilkan model prediktif yang akurat, robust, dan aplikatif bagi praktisi teknik sipil.

BAB III

METODE PENELITIAN

A. Tempat dan Waktu Penelitian

Penelitian ini dilakukan di Laboratorium Teknik Sipil Program Studi Teknik Sipil Universitas Bojonegoro. Penelitian dikerjakan untuk mengembangkan model prediksi berbasis ensemble learning yang mampu secara simultan memprediksi c , ϕ , dan E untuk tanah kohesif, menggunakan dataset yang dikompilasi dari berbagai sumber geoteknik global.. dengan waktu Penelitian dimulai bulan November 2025 sampai selesai.

B. Alat dan Bahan Penelitian

Alat dan bahan dalam penelitian ini menggunakan software untuk pemrograman, pengujian dan berbagai komponen utama pembentuk beton, baik bahan semen dan aditif maupun agregat dan air pencampur. Adapun penjelasan masing-masing material adalah sebagai berikut:

1. Alat Penelitian

Peralatan yang digunakan pada penelitian menggunakan perangkat lunak untuk pemrograman dan pengujian. Peralatan perangkat lunak yang digunakan dalam perancangan sebagai berikut:

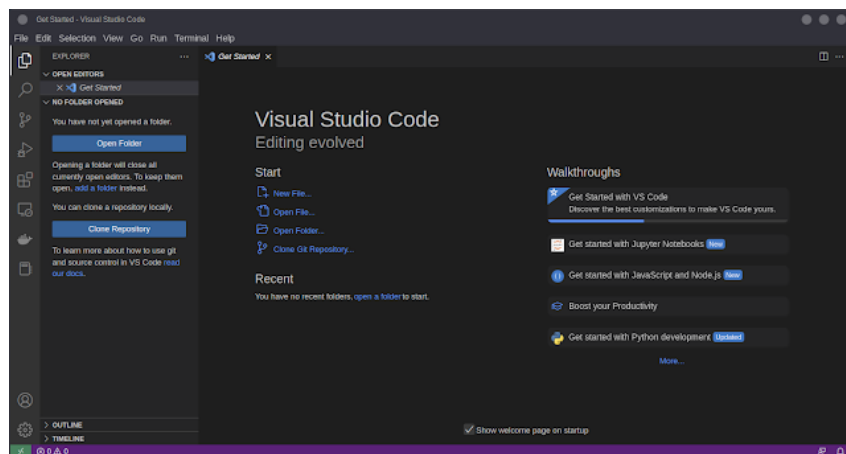
a. *Visual Studio Code*

Penelitian ini, perangkat lunak yang digunakan untuk membantu proses pemrograman dan analisis data adalah *Visual Studio Code (VS Code)*. *VS Code* merupakan editor kode sumber lintas platform yang dikembangkan oleh *Microsoft* dan bersifat gratis, ringan, serta mendukung berbagai bahasa pemrograman, termasuk *Python*. *VS Code* dipilih karena kemampuannya yang fleksibel dan efisien dalam mendukung pengembangan program, khususnya dalam implementasi algoritma machine learning. Perangkat lunak ini dilengkapi berbagai fitur unggulan yang sangat membantu dalam proses penelitian, antara lain terminal terintegrasi, manajemen file proyek yang sistematis, serta kemampuan integrasi dengan kontrol versi seperti *Git* untuk pencatatan perubahan kode.

Salah satu keunggulan utama dari *VS Code* adalah kemampuannya untuk menginstal ekstensi tambahan, seperti ekstensi *Python* dan *Jupyter*. Ekstensi *Python* memungkinkan pengguna menulis, menjalankan, dan *debuging* kode *Python* secara

langsung dari editor, serta mendapatkan fitur *IntelliSense* yang membantu melengkapi sintaks secara otomatis. Sementara itu, integrasi *Jupyter Notebook* memungkinkan pengguna menjalankan file *.ipynb* secara interaktif di dalam *VS Code* tanpa perlu membuka aplikasi terpisah. Hal ini sangat berguna dalam proses eksplorasi data, visualisasi hasil, dan dokumentasi eksperimen secara terstruktur.

Berbagai fitur yang dimilikinya, *VS Code* digunakan dalam seluruh tahap analisis, mulai dari pra-pemrosesan data, pelatihan model klasifikasi, evaluasi kinerja model, hingga visualisasi hasil dalam bentuk grafik. Penggunaan *Visual Studio Code* dalam penelitian ini memberikan efisiensi kerja yang tinggi, kemudahan dalam pelacakan eksperimen, serta mendukung transparansi dan reproduibilitas hasil penelitian.



Gambar 3. 1 Software VS Code

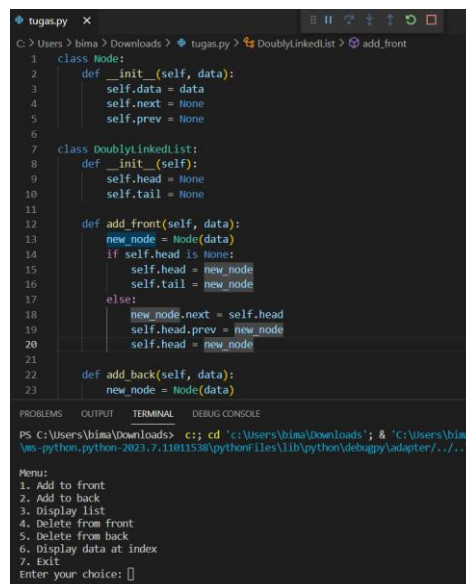
b. Python

Bahasa pemrograman yang digunakan untuk membangun dan menjalankan model klasifikasi adalah *Python*. *Python* merupakan bahasa pemrograman tingkat tinggi yang bersifat *open-source*, mudah dibaca, dan memiliki sintaks yang sederhana namun kuat. *Python* sangat populer di kalangan ilmuwan data dan peneliti karena memiliki berbagai pustaka (*library*) yang mendukung proses analisis data dan *machine learning* secara efisien dan terstruktur. Dalam konteks penelitian ini, *Python* digunakan untuk melakukan proses *preprocessing* data, pelatihan model klasifikasi menggunakan algoritma *ensemble learning* evaluasi kinerja model, serta visualisasi data dan hasil klasifikasi.

Beberapa pustaka utama yang digunakan dalam implementasi model klasifikasi meliputi *NumPy*, *Pandas*, *Scikit-learn*, dan *Matplotlib*. *NumPy* digunakan untuk melakukan operasi numerik, sementara *Pandas* digunakan untuk manipulasi dan

analisis data berbasis tabel. *Scikit-learn* berperan sebagai pustaka utama dalam membangun dan melatih model *machine learning* serta menyediakan berbagai metode evaluasi dan akurasi. Untuk visualisasi data, digunakan *Matplotlib* yang memungkinkan pembuatan grafik dan diagram secara fleksibel.

Kemampuan *Python* dalam mengintegrasikan seluruh proses analisis data, mulai dari pengolahan hingga visualisasi, menjadikannya alat yang sangat penting dalam penelitian ini. Selain itu, *Python* juga mendukung lingkungan interaktif seperti *Jupyter Notebook* yang memudahkan peneliti untuk menulis, menjalankan, dan mendokumentasikan kode secara bersamaan dalam satu antarmuka yang mudah diakses. Dengan semua keunggulan tersebut, *Python* berperan sebagai fondasi utama dalam pelaksanaan analisis data dan implementasi model klasifikasi pada penelitian ini.



```
1 class Node:
2     def __init__(self, data):
3         self.data = data
4         self.next = None
5         self.prev = None
6
7 class DoublyLinkedList:
8     def __init__(self):
9         self.head = None
10        self.tail = None
11
12    def add_front(self, data):
13        new_node = Node(data)
14        if self.head is None:
15            self.head = new_node
16            self.tail = new_node
17        else:
18            new_node.next = self.head
19            self.head.prev = new_node
20            self.head = new_node
21
22    def add_back(self, data):
23        new_node = Node(data)
```

PROBLEMS OUTPUT TERMINAL DEBUG CONSOLE

```
PS C:\Users\bima\Downloads> c:; cd 'c:\Users\bima\Downloads'; & 'C:\Users\bima\Downloads\python-2023-7-11011538\python\Lib\python\debugpy\adapter/...'
Menu:
1. Add to front
2. Add to back
3. Display list
4. Delete from front
5. Delete from back
6. Display data at index
7. Exit
Enter your choice: []
```

Gambar 3. 2 Bahasa Perograman *Python*

C. Jenis Penelitian

Penelitian ini menggunakan pendekatan kualitatif dengan metode studi pustaka (desk study). Metode kualitatif dipilih karena penelitian ini berfokus pada pengembangan dan analisis model prediktif berbasis data yang bersifat eksploratif dan interpretatif, bukan pada pengujian hipotesis kuantitatif melalui eksperimen fisik. Pendekatan desk study memungkinkan peneliti untuk mengumpulkan, menganalisis, dan mensintesis informasi dari berbagai sumber sekunder seperti jurnal ilmiah terindeks Scopus, prosiding konferensi, dan database geoteknik global untuk membangun kerangka kerja penelitian yang komprehensif.

Penelitian tidak melibatkan eksperimen laboratorium atau pengambilan data lapangan baru, melainkan memanfaatkan dataset hasil pengujian tanah yang telah dipublikasikan dalam literatur ilmiah internasional. Data-data tersebut dikompilasi, diolah, dan digunakan untuk melatih serta mengevaluasi model ensemble learning. Tujuan utama adalah untuk memahami hubungan kompleks antara sifat fisik tanah kohesif dan parameter kekuatan gesernya (c , ϕ , E) melalui pendekatan komputasi canggih, serta memberikan wawasan teknis yang dapat diinterpretasikan oleh insinyur sipil. Dengan demikian, jenis penelitian ini sangat sesuai untuk menjawab rumusan masalah yang berkaitan dengan pengembangan model prediktif, pemilihan algoritma terbaik, dan identifikasi variabel dominan dalam konteks rekayasa geoteknik.

D. Identifikasi dan klasifikasi variabel

Penelitian ini menggunakan pendekatan supervised learning dalam machine learning, sehingga variabel penelitian diklasifikasikan berdasarkan peranannya dalam proses pelatihan model, yaitu sebagai variabel independen (input), dependen (output), dan kontrol (parameter yang dijaga konsistensinya).

1. Variabel Dependen (Target)

Variabel dependen dalam penelitian ini adalah parameter kekuatan geser tanah kohesif yang akan diprediksi secara simultan oleh model ensemble learning. Terdapat tiga variabel dependen utama yang menjadi output dari model:

- a. Koefisien Kohesi (c): Gaya ikat antar partikel tanah yang memberikan kekuatan pada tanah meskipun tidak ada tekanan normal. Satuan kPa
- b. Sudut Geser Internal (ϕ): Parameter yang mencerminkan resistensi tanah terhadap geser akibat interlocking butiran. Satuan: derajat ($^{\circ}$).
- c. Modulus Elastisitas (E): Parameter yang menggambarkan kekakuan atau hubungan tegangan-regangan elastis tanah. Satuan: MPa.

Ketiga parameter ini merupakan fondasi utama dalam analisis stabilitas lereng, daya dukung fondasi, dan prediksi penurunan, sehingga pemodelan simultannya sangat relevan untuk aplikasi teknik sipil.

2. Variabel Independen (Fitur/Parameter Input)

Variabel independen adalah sifat fisik dan hasil pengujian tanah yang digunakan sebagai input untuk memprediksi ketiga parameter dependen. Data input ini akan dikompilasi dari database geoteknik global dan meliputi:

- a. Komposisi Butiran: Persentase berat kandungan lempung (clay fraction), lanau (silt fraction), pasir (sand fraction), dan kerikil (gravel fraction) dalam sampel tanah.
- b. Indeks Plastisitas: Batas Cair (Liquid Limit, LL), Batas Plastisitas (Plastic Limit, PL), dan Indeks Plastisitas (Plasticity Index, $PI = LL - PL$).
- c. Parameter Fisik Dasar: Berat isi basah (moist unit weight, γ) dan kadar air (water content, w).

Variabel-variabel ini dipilih karena telah terbukti secara empiris dan melalui penelitian sebelumnya memiliki korelasi kuat dengan parameter kekuatan geser tanah.

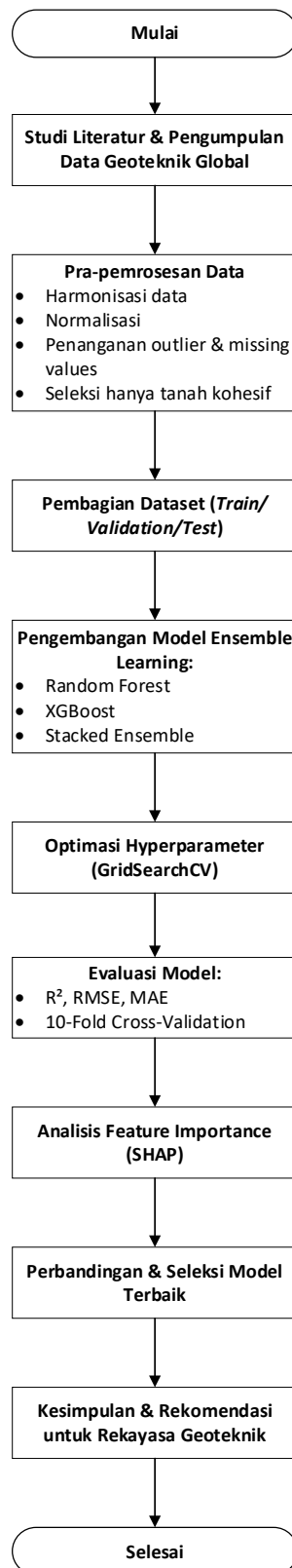
3. Variabel Kontrol

Untuk memastikan validitas perbandingan dan evaluasi model, beberapa aspek dari proses penelitian akan dijaga konstan sebagai variabel kontrol

- a. Metodologi Pengolahan Data: Proses prapemrosesan data akan menggunakan metode yang sama untuk semua eksperimen.
- b. Algoritma Ensemble yang Dibandingkan, meskipun akan dieksplorasi berbagai pendekatan, set algoritma dasar yang dibandingkan akan tetap konsisten.
- c. Strategi Evaluasi Model, protokol evaluasi, termasuk pembagian data (train-test split), penggunaan 10-fold cross-validation, dan metrik evaluasi (R^2 , RMSE, MAE), akan diterapkan secara seragam untuk semua model.
- d. Hyperparameter Tuning: Teknik optimasi hyperparameter akan diterapkan dengan skema yang sama untuk masing-masing algoritma guna memastikan perbandingan yang adil.

Dengan mengidentifikasi dan mengontrol variabel-variabel tersebut, penelitian ini dapat menghasilkan temuan yang objektif dan dapat direplikasi, serta memungkinkan kesimpulan yang sah tentang efektivitas masing-masing pendekatan ensemble learning.

E. Alur Penelitian



Gambar 3. 3 Alur Penelitian

BAB IV

HASIL DAN PEMBAHASAN

A. Hyperparameter Tuning

Proses hyperparameter tuning dilakukan untuk memperoleh kombinasi parameter terbaik pada setiap model ensemble sehingga mampu menghasilkan performa prediksi yang optimal terhadap target soil deformation modulus (E_s). Strategi pencarian parameter menggunakan pendekatan Randomized Search dengan skema k-fold cross validation sebanyak 5 lipatan ($n_folds_cv = 5$) dan metrik evaluasi utama berupa koefisien determinasi (R^2). Proporsi data uji ditetapkan sebesar 20% ($test_size = 0.2$) dengan $random_state$ 42 untuk menjamin reproduktibilitas eksperimen. Jumlah iterasi pencarian ditetapkan sebanyak 30 kombinasi acak ($tuning_n_iter = 30$) guna menyeimbangkan efisiensi komputasi dan eksplorasi ruang parameter. Secara umum, parameter yang dituning mencakup kompleksitas pohon (max_depth), jumlah estimator ($n_estimators$ atau max_iter), laju pembelajaran ($learning_rate$), serta parameter regularisasi dan subsampling untuk mencegah overfitting. Model berbasis bagging seperti Random Forest lebih berfokus pada variasi struktur pohon dan pembagian sampel, sedangkan model berbasis boosting seperti HistGradientBoosting dan XGBoost menekankan kontrol laju pembelajaran, jumlah iterasi, dan regularisasi. Tabel 4.1 merangkum rentang hyperparameter yang dieksplorasi serta nilai terbaik (best hyperparameter) yang terpilih berdasarkan nilai R^2 validasi silang tertinggi.

Tabel 4.1 Hyperparameter Tuning

Model	Hyperparameter	Rentang Pencarian	Best Hyperparameter Terpilih
Random Forest	$n_estimators$	100 – 500	200
	max_depth	5 – 30 / None	15
	$min_samples_split$	2 – 10	8
	$min_samples_leaf$	1 – 5	1
	$max_features$	sqrt, log2, None	None
HistGradientBoosting	max_iter	100 – 600	500
	max_depth	3 – 15	10
	$learning_rate$	0.01 – 0.1	0.01
	$min_samples_leaf$	5 – 30	10
	max_leaf_nodes	15 – 63	31
	$l2_regularization$	0 – 1	0.1
	max_bins	128 – 255	255
XGBoost	$n_estimators$	100 – 600	500
	max_depth	3 – 10	6
	$learning_rate$	0.01 – 0.3	0.07
	$subsample$	0.5 – 1.0	0.8
	$colsample_bytree$	0.5 – 1.0	0.7
	$gamma$	0 – 0.5	0.1
	reg_alpha	0 – 0.1	0.01
	reg_lambda	1 – 3	1.5

Berdasarkan Tabel 4.1, terlihat bahwa setiap model memiliki konfigurasi hyperparameter optimal yang berbeda sesuai dengan karakteristik algoritmanya. Pada Random Forest, kombinasi terbaik diperoleh dengan jumlah pohon sebanyak 200 dan kedalaman maksimum 15, yang menunjukkan model memanfaatkan struktur pohon yang cukup kompleks tanpa terlalu dalam. HistGradientBoosting memilih jumlah iterasi yang relatif tinggi (500) dengan learning rate rendah (0,01), mengindikasikan strategi pembelajaran bertahap yang stabil, disertai pembatasan kompleksitas melalui max_depth 10 dan min_samples_leaf 10. Sementara itu, XGBoost menggunakan 500 estimator dengan kedalaman maksimum 6 serta kombinasi subsample 0,8 dan colsample_bytree 0,7, yang mencerminkan adanya mekanisme pengendalian variasi melalui teknik sampling. Parameter regularisasi seperti reg_alpha dan reg_lambda juga dipilih untuk menjaga kestabilan model. Secara keseluruhan, tabel tersebut menunjukkan bahwa masing-masing algoritma memerlukan konfigurasi spesifik untuk mencapai struktur model yang seimbang antara kompleksitas dan kontrol regularisasi.

B. Performance of All Models

Evaluasi kinerja seluruh model ensemble dilakukan menggunakan data uji (test set) serta validasi silang 5-fold. Parameter evaluasi yang digunakan meliputi R^2 , Adjusted R^2 , MAE, RMSE, serta statistik validasi silang berupa rata-rata CV R^2 , standar deviasi, dan interval kepercayaan 95%. Ringkasan performa seluruh model disajikan pada Tabel 4.2.

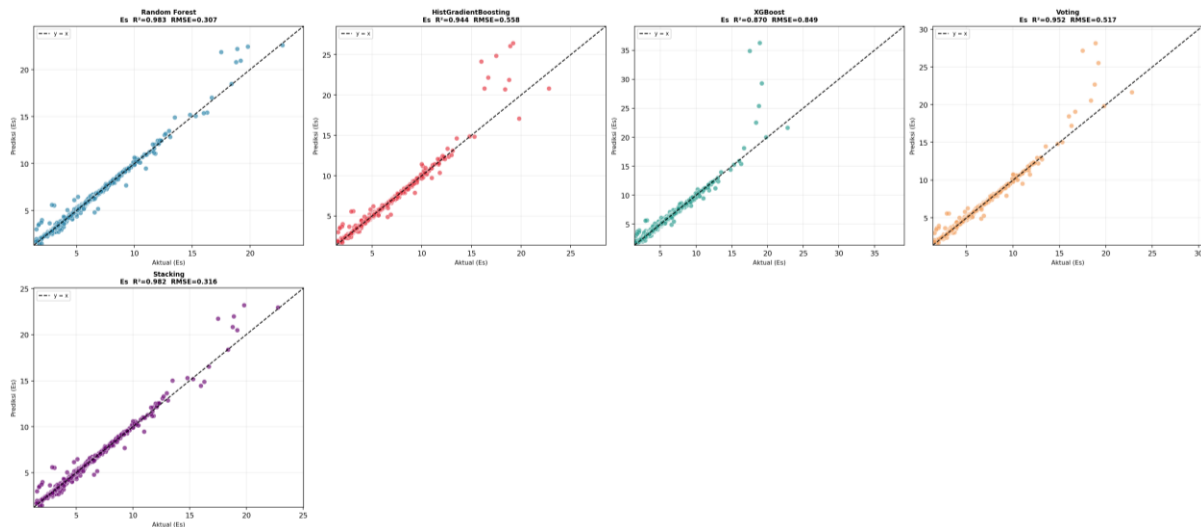
Tabel 4.2 Performa Model

Model	R^2	Adj- R^2	MAE	RMSE	CV R^2 Mean	CV R^2 Std	CV 95% CI
Random Forest	0.9830	0.9829	0.1115	0.3073	0.9651	0.0372	[0.9325, 0.9978]
HistGradientBoosting	0.9439	0.9436	0.1456	0.5578	0.9271	0.0532	[0.8805, 0.9738]
XGBoost	0.8701	0.8696	0.1925	0.8486	0.9165	0.0553	[0.8680, 0.9649]
Voting	0.9518	0.9516	0.1315	0.5168	0.9459	0.0481	[0.9038, 0.9881]
Stacking	0.9820	0.9819	0.1171	0.3162	0.9646	0.0279	[0.9402, 0.9890]

Nilai pada Tabel 4.2 menunjukkan bahwa Random Forest menghasilkan performa tertinggi pada data uji dengan R^2 sebesar 0,983 dan RMSE sebesar 0,3073 MPa, yang mencerminkan tingkat akurasi prediksi sangat tinggi dengan kesalahan relatif kecil terhadap rentang nilai Es. Rata-rata CV R^2 sebesar 0,9651 dengan standar deviasi 0,0372 juga mengindikasikan stabilitas model yang baik selama proses validasi silang. Model Stacking memberikan performa yang sangat kompetitif dengan R^2 sebesar 0,982 dan RMSE 0,3162 MPa. Standar deviasi CV yang

lebih rendah (0,0279) menunjukkan tingkat konsistensi yang lebih tinggi dibandingkan model lainnya. Kombinasi beberapa model dasar dalam pendekatan ini mampu meningkatkan robustnes tanpa menurunkan akurasi secara signifikan. Model Voting menunjukkan performa menengah dengan R^2 sebesar 0,9518 dan RMSE 0,5168 MPa. Pendekatan agregasi sederhana mampu meningkatkan akurasi dibandingkan sebagian model tunggal, namun belum melampaui performa model terbaik. HistGradientBoosting menghasilkan R^2 sebesar 0,9439 dengan RMSE 0,5578 MPa, sedangkan XGBoost menunjukkan R^2 terendah pada data uji sebesar 0,8701 dengan RMSE 0,8486 MPa. Meskipun nilai rata-rata CV R^2 keduanya masih relatif tinggi (di atas 0,91), tingkat kesalahan prediksi pada data uji lebih besar dibandingkan model lain. Secara keseluruhan, Random Forest dan Stacking tampil sebagai model paling unggul dalam memprediksi modulus deformasi tanah (E_s), baik dari sisi akurasi maupun stabilitas validasi silang.

Evaluasi visual terhadap kecocokan model dilakukan melalui perbandingan antara nilai aktual dan nilai prediksi modulus deformasi (E_s). Visualisasi ini bertujuan untuk menilai seberapa dekat hasil prediksi mengikuti garis identitas ($y = x$), yang merepresentasikan prediksi sempurna.



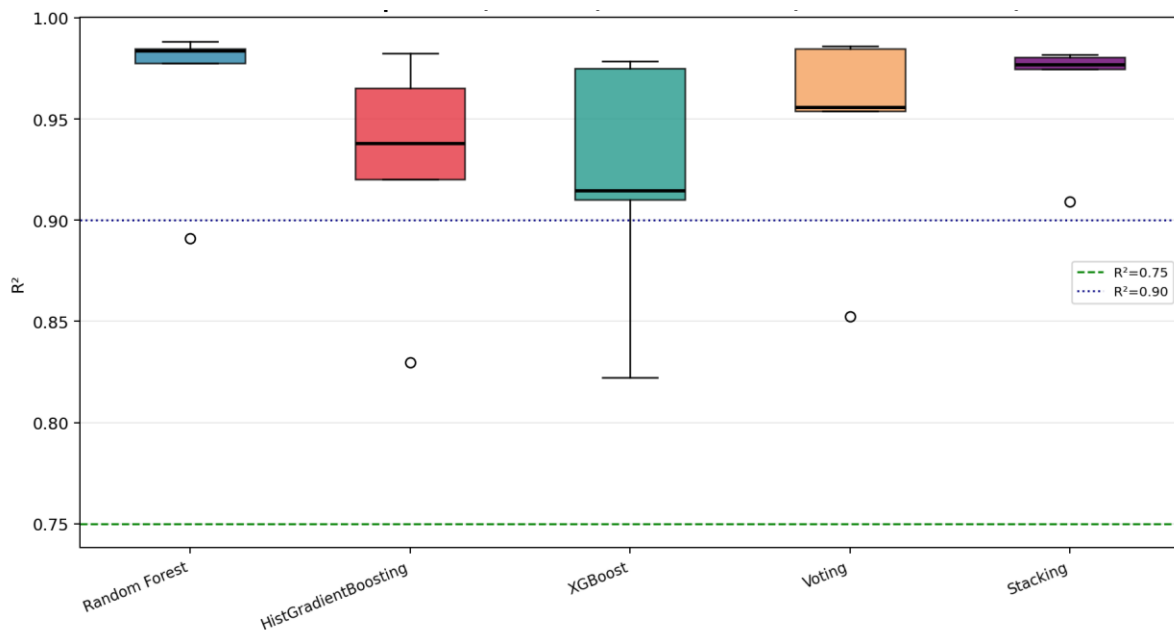
Gambar 4.1 Perbandingan Nilai Aktual dan Prediksi Modulus Deformasi

Berdasarkan Gambar 4.1, mayoritas titik data terkonsentrasi di sekitar garis diagonal, yang menunjukkan bahwa model mampu merepresentasikan hubungan antara properti indeks tanah dan nilai E_s dengan tingkat akurasi yang tinggi. Kepadatan titik yang mengikuti garis identitas mengindikasikan kesesuaian prediksi yang kuat pada rentang nilai rendah hingga menengah. Hal ini konsisten dengan nilai R^2 yang tinggi yang telah dilaporkan sebelumnya pada Tabel performa model.

Pada rentang nilai Es yang lebih tinggi, terlihat beberapa titik yang sedikit menyimpang dari garis diagonal. Penyimpangan ini menunjukkan adanya error prediksi yang lebih besar pada sebagian kecil observasi, yang umumnya berkaitan dengan nilai ekstrem atau distribusi data yang tidak merata. Namun demikian, secara keseluruhan tidak terlihat pola deviasi sistematis yang signifikan, sehingga model tidak menunjukkan kecenderungan overestimation atau underestimation yang dominan. Distribusi titik yang relatif simetris terhadap garis identitas memperkuat indikasi bahwa model memiliki kemampuan generalisasi yang baik. Visualisasi ini melengkapi evaluasi kuantitatif sebelumnya dan menunjukkan bahwa performa tinggi yang diperoleh bukan hanya hasil dari optimasi numerik, tetapi juga tercermin dalam kesesuaian pola prediksi terhadap data aktual.

C. Cross-Validation Stability Analysis

Evaluasi stabilitas model dilakukan menggunakan 5-fold cross-validation dengan metrik koefisien determinasi (R^2) pada setiap fold. Analisis ini bertujuan untuk menilai konsistensi performa prediksi terhadap variasi subset data validasi, sehingga dapat diidentifikasi model dengan kemampuan generalisasi paling stabil.



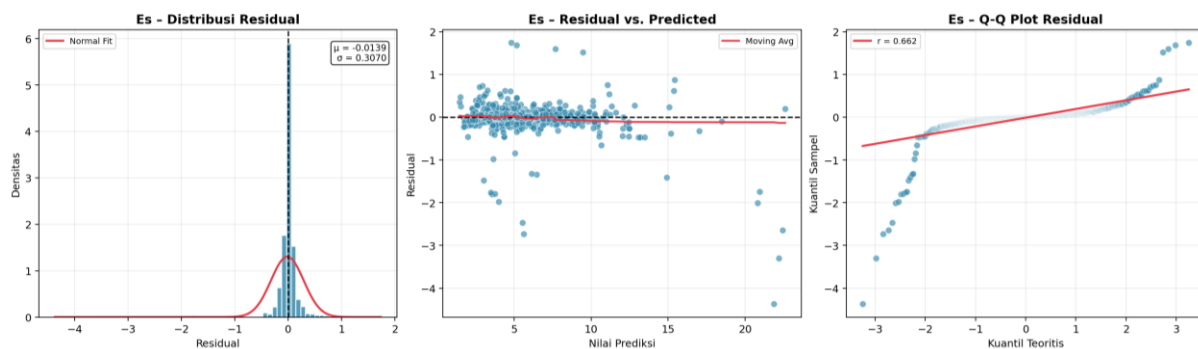
Gambar 4.2 Distribusi Nilai R^2 pada 5-Fold Cross-Validation untuk Seluruh Model

Berdasarkan Gambar 4.2, Random Forest dan Stacking menunjukkan distribusi R^2 yang paling sempit dengan median mendekati 0,98. Rentang interkuartil (IQR) yang kecil mengindikasikan variasi performa antar-fold yang rendah, sehingga kedua model tersebut dapat dikategorikan memiliki stabilitas prediktif yang tinggi. Meskipun terdapat beberapa titik outlier, nilainya tetap berada pada kisaran performa tinggi (sekitar 0,89–0,91), sehingga tidak menunjukkan penurunan akurasi yang signifikan. HistGradientBoosting menampilkan median

sekitar 0,94–0,95 dengan variasi yang lebih lebar dibandingkan Random Forest dan Stacking. Nilai minimum yang mendekati 0,83 menunjukkan adanya fold tertentu dengan performa yang relatif lebih rendah, yang mengindikasikan sensitivitas terhadap pembagian data. XGBoost memiliki distribusi paling lebar dengan nilai minimum sekitar 0,82 dan median sedikit di atas 0,91. Lebaranya sebaran ini menunjukkan fluktuasi performa yang lebih tinggi antar-fold. Voting Regressor memperlihatkan median sekitar 0,96 dengan variasi moderat. Pola ini menunjukkan bahwa pendekatan agregasi mampu meningkatkan konsistensi dibandingkan beberapa model tunggal, meskipun kestabilannya masih sedikit di bawah Random Forest dan Stacking.

D. Residual Analysis and Prediction Bias

Analisis residual dilakukan untuk mengevaluasi pola kesalahan prediksi serta mengidentifikasi potensi bias sistematis pada model terbaik. Residual didefinisikan sebagai selisih antara nilai aktual dan nilai prediksi ($e = y_{\text{actual}} - y_{\text{predicted}}$). Evaluasi ini penting karena nilai R^2 yang tinggi belum tentu menjamin tidak adanya pola kesalahan tertentu.



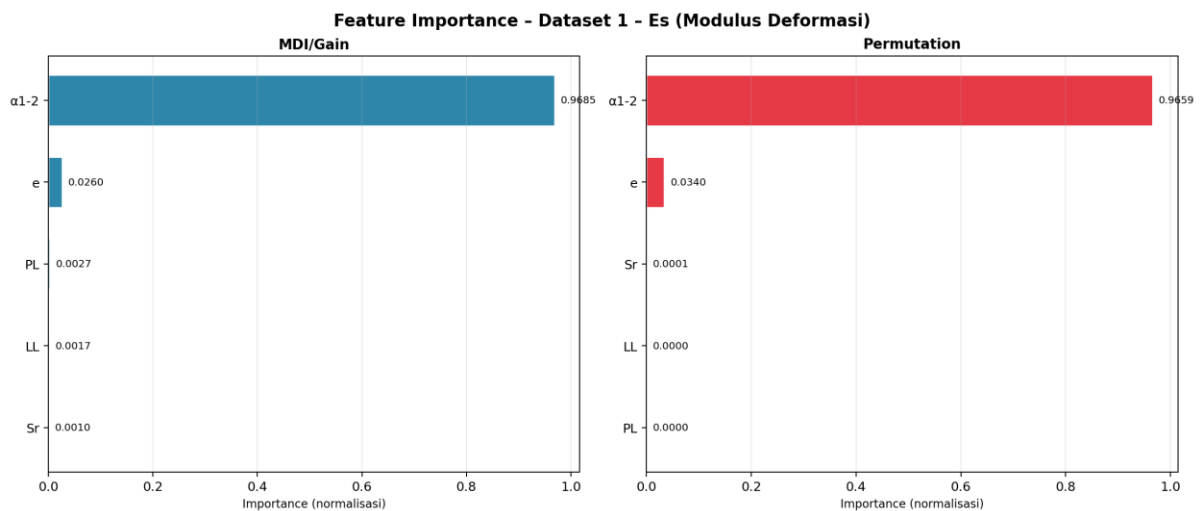
Gambar 1.4 Analisis Residual Model Terbaik

Berdasarkan histogram pada Gambar 4.3 (panel kiri), distribusi residual terpusat di sekitar nol dengan nilai rata-rata (μ) sebesar $-0,0139$ dan simpangan baku (σ) sebesar $0,3070$. Nilai rata-rata yang sangat dekat dengan nol menunjukkan bahwa model tidak memiliki bias prediksi yang signifikan secara global (tidak cenderung overestimate maupun underestimate secara sistematis). Sebaran residual yang relatif simetris di sekitar nol mengindikasikan bahwa kesalahan prediksi bersifat acak pada sebagian besar rentang data. Plot residual terhadap nilai prediksi (panel tengah) menunjukkan bahwa mayoritas titik tersebar acak di sekitar garis nol, khususnya pada rentang prediksi rendah hingga menengah. Namun, pada nilai prediksi yang lebih tinggi terlihat beberapa residual negatif ekstrem. Kondisi ini menunjukkan kecenderungan model sedikit melakukan overestimation pada beberapa sampel dengan nilai Es besar. Meskipun demikian, tidak terlihat pola kurvilinear yang kuat, sehingga indikasi ketidaklinieran yang tidak tertangkap model relatif kecil. Q-Q plot (panel kanan)

memperlihatkan bahwa sebagian besar titik mengikuti garis referensi pada bagian tengah distribusi, namun terjadi deviasi pada ekor bawah. Hal ini menunjukkan adanya beberapa residual ekstrem yang menyebabkan penyimpangan dari distribusi normal sempurna. Meskipun demikian, deviasi tersebut terbatas pada sebagian kecil observasi dan tidak mendominasi keseluruhan distribusi error.

E. Feature Importance Analysis

Analisis feature importance dilakukan untuk mengidentifikasi kontribusi relatif masing-masing variabel input terhadap prediksi modulus deformasi tanah (E_s) pada model terbaik (Random Forest). Dua metode evaluasi digunakan secara bersamaan, yaitu MDI/Gain (Mean Decrease in Impurity) dan Permutation Importance, guna memperoleh gambaran yang lebih komprehensif dan mengurangi potensi bias yang dapat muncul apabila hanya menggunakan satu metode saja. Hasil analisis disajikan pada Gambar 4.5.



Gambar 4.2 Feature Importance Analysis

Berdasarkan Gambar 4.5, kedua metode secara konsisten menunjukkan bahwa parameter $\alpha 1-2$ merupakan fitur yang paling dominan dalam prediksi E_s , dengan nilai kepentingan sebesar 0,9685 (MDI/Gain) dan 0,9659 (Permutation). Dominasi yang sangat tinggi ini mengindikasikan bahwa $\alpha 1-2$, yang merepresentasikan koefisien kompresibilitas tanah, memiliki hubungan yang sangat kuat dengan modulus deformasi. Hal ini sejalan secara teoritis mengingat $\alpha 1-2$ merupakan parameter yang secara langsung mencerminkan karakteristik deformasi tanah di bawah beban, sehingga wajar apabila nilai tersebut menjadi prediktor utama dalam pemodelan E_s .

Fitur void ratio (e) menempati posisi kedua dengan nilai kepentingan sebesar 0,0260 (MDI) dan 0,0340 (Permutation), namun dengan selisih yang sangat jauh dibandingkan $\alpha 1-2$. Sementara itu, variabel lainnya—meliputi PL, LL, dan Sr—memberikan kontribusi yang

sangat kecil terhadap prediksi model, dengan nilai kepentingan yang mendekati nol pada kedua metode evaluasi. Konsistensi peringkat antar kedua metode memperkuat reliabilitas hasil analisis ini dan mengonfirmasi bahwa model telah mengidentifikasi variabel yang paling relevan secara fisis dalam menentukan perilaku deformasi tanah.

BAB V

KESIMPULAN

A. Kesimpulan dan Saran

Penelitian ini berhasil mengembangkan model prediksi modulus deformasi tanah (E_s) berbasis ensemble machine learning dengan memanfaatkan parameter indeks tanah yang umum tersedia dari pengujian laboratorium rutin. Dari seluruh model yang dievaluasi, Random Forest terpilih sebagai model terbaik dengan nilai R^2 sebesar 0,983 dan RMSE sebesar 0,3073 MPa pada data uji, didukung oleh rata-rata CV R^2 sebesar 0,9651 yang mengindikasikan stabilitas generalisasi yang tinggi. Model Stacking menunjukkan performa yang sangat kompetitif dengan R^2 sebesar 0,982 dan konsistensi validasi silang yang lebih tinggi (standar deviasi CV = 0,0279), menjadikannya alternatif yang layak dipertimbangkan dalam aplikasi praktis. Secara keseluruhan, pendekatan ensemble yang dikombinasikan dengan hyperparameter tuning dan validasi silang lima lipatan terbukti efektif dalam menangkap hubungan non-linear antara parameter indeks tanah dan modulus deformasi.

Secara ilmiah, penelitian ini berkontribusi dalam mendemonstrasikan bahwa parameter α_1 – α_2 merupakan prediktor dominan terhadap E_s , sebagaimana dikonfirmasi secara konsisten oleh dua metode analisis feature importance yang berbeda, yaitu MDI/Gain dan Permutation Importance. Temuan ini memperkuat pemahaman teoritis bahwa koefisien kompresibilitas tanah memiliki keterkaitan langsung dengan karakteristik deformasi tanah dan dapat dijadikan dasar seleksi fitur pada model prediksi geoteknik berbasis data. Dari sisi praktis, model yang dikembangkan berpotensi digunakan sebagai alat bantu estimasi awal nilai E_s dalam perencanaan geoteknik, khususnya pada kondisi di mana pengujian lapangan secara langsung tidak memungkinkan atau membutuhkan biaya yang tinggi. Kemampuan model dalam menghasilkan prediksi dengan akurasi tinggi hanya dari parameter indeks tanah membuka peluang penghematan waktu dan sumber daya dalam tahap investigasi awal.

Meskipun demikian, penelitian ini memiliki beberapa keterbatasan yang perlu diakui. Dataset yang digunakan bersumber dari satu repositori tunggal yang berfokus pada tanah lempung, sehingga kemampuan generalisasi model terhadap jenis tanah lain seperti pasir, lanau, atau tanah organik belum dapat dipastikan. Selain itu, model yang dikembangkan tidak mempertimbangkan parameter stress history tanah, seperti overconsolidation ratio (OCR), yang diketahui memiliki pengaruh signifikan terhadap perilaku deformasi tanah dalam konteks konsolidasi. Ketiadaan variabel tersebut berpotensi menjadi sumber ketidakpastian pada kondisi tanah dengan riwayat pembebanan yang kompleks.

Untuk penelitian ke depan, beberapa arah pengembangan direkomendasikan guna memperluas cakupan dan meningkatkan kapasitas prediktif model. Pertama, integrasi parameter in-situ seperti nilai CPT (Cone Penetration Test) dan SPT (Standard Penetration Test) diharapkan dapat meningkatkan akurasi dan relevansi model dalam kondisi lapangan yang lebih beragam. Kedua, eksplorasi pendekatan deep learning, seperti arsitektur jaringan saraf tiruan atau model berbasis transformer, berpotensi menangkap pola hubungan yang lebih kompleks antara fitur input dan variabel target. Ketiga, validasi dan generalisasi model terhadap dataset multi-region yang mencakup berbagai kondisi geologi dan jenis tanah yang lebih beragam perlu dilakukan untuk memastikan keandalan model dalam skala yang lebih luas.

DAFTAR PUSTKA

- [1] B. M. Das and K. Sobhan, *Principles of Geotechnical Engineering (Ninth Edition)*, vol. 9. 2016.
- [2] K. K. Phoon and W. Zhang, “Future of machine learning in geotechnics,” 2023. doi: 10.1080/17499518.2022.2087884.
- [3] A. Baghbani, T. Choudhury, S. Costa, and J. Reiner, “Application of artificial intelligence in geotechnical engineering: A state-of-the-art review,” 2022. doi: 10.1016/j.earscirev.2022.103991.
- [4] W. Zhang, X. Gu, L. Tang, Y. Yin, D. Liu, and Y. Zhang, “Application of machine learning, deep learning and optimization algorithms in geoenvironment and geoscience: Comprehensive review and future challenge,” 2022. doi: 10.1016/j.gr.2022.03.015.
- [5] E. Díaz and G. Spagnoli, “A super-learner machine learning model for a global prediction of compression index in clays,” *Appl. Clay Sci.*, vol. 249, 2024, doi: 10.1016/j.clay.2023.107239.
- [6] P. Zhang, Z. Y. Yin, and Y. F. Jin, “Machine Learning-Based Modelling of Soil Properties for Geotechnical Design: Review, Tool Development and Comparison,” *Archives of Computational Methods in Engineering*, vol. 29, no. 2, 2022, doi: 10.1007/s11831-021-09615-5.
- [7] W. Shao *et al.*, “The Application of Machine Learning Techniques in Geotechnical Engineering: A Review and Comparison,” 2023. doi: 10.3390/math11183976.
- [8] E. Yaghoubi, E. Yaghoubi, A. Khamees, and A. H. Vakili, “A systematic review and meta-analysis of artificial neural network, machine learning, deep learning, and ensemble learning approaches in field of geotechnical engineering,” 2024. doi: 10.1007/s00521-024-09893-7.
- [9] H. T. Wu, Z. L. Zhang, and D. Dias, “Prediction on compression indicators of clay soils using XGBoost with Bayesian optimization,” *J. Cent. South Univ.*, vol. 31, no. 11, 2024, doi: 10.1007/s11771-024-5681-9.
- [10] K. Hamdaoui, A. Benzaamia, B. Sari Ahmed, M. E. Guellil, and M. Ghrici, “Interpretable machine learning for predicting compression index of clays using SHAP and gradient boosting models,” *Journal of Engineering and Applied Science*, vol. 72, no. 1, 2025, doi: 10.1186/s44147-025-00727-4.
- [11] X. Duan, “Assessment of resilient modulus of soil using hybrid extreme gradient boosting models,” *Sci. Rep.*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-81311-3.
- [12] M. A. Senturk, E. Ordu, and R. K. Tan, “Machine learning-based prediction of soil compaction parameters,” *Environ. Earth Sci.*, vol. 84, no. 13, 2025, doi: 10.1007/s12665-025-12328-8.

- [13] R. Akshaya and K. Premalatha, "Machine Learning Methods for Predicting Soil Compression Index," in *Lecture Notes in Civil Engineering*, 2024. doi: 10.1007/978-981-97-0072-1_27.
- [14] X. Wu, K. Wu, S. Hao, E. Yu, J. Zhao, and Y. Li, "Machine learning ensemble technique for exploring soil type evolution," *Sci. Rep.*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-10608-8.
- [15] A. Almarzooqi, M. Arab, M. Omar, and E. Alotaibi, "Benchmarking Conventional Machine Learning Models for Dynamic Soil Property Prediction," *Buildings*, vol. 15, p. 4188, Nov. 2025, doi: 10.3390/buildings15224188.
- [16] N. Bozorgzadeh and Y. Feng, "Evaluation structures for machine learning models in geotechnical engineering," *Georisk*, vol. 18, no. 1, 2024, doi: 10.1080/17499518.2024.2313485.
- [17] G. Sun, "SZ-CLAY-10-7964," Feb. 2025. doi: 10.17632/y94jyms83b.1.

LAMPIRAN

Sinta 2 JGEET (Journal of Geoscience, Engineering, Environment, and Technology)

<https://journal.uir.ac.id/index.php/JGEET/index>

