

LAPORAN
PENELITIAN INTERNAL DOSEN



**PREDIKSI KINERJA MEKANIS BETON BERBASIS AGREGAT DAUR ULANG
MENGUNAKAN MACHINE LEARNING: SEBUAH STUDI KOMPARATIF
DENGAN BETON KONVENSIONAL**

Tim Peneliti:

Sujat, S.T., M.T.

Dr. Eko Wahyu Abryandoko., S.Pd.,MT

Achmad Safi'i

Dibiayai oleh:

Universitas Bojonegoro

Periode 1 Tahun Anggaran 2025/2026

No. Kontrak: 064/LPPM-LIT/UB/XI/2025

UNIVERSITAS BOJONEGORO

2026

HALAMAN PENGESAHAN
LAPORAN PENELITIAN MANDIRI

- 1. Judul Penelitian** : Prediksi Kinerja Mekanis Beton Berbasis Agregat Daur Ulang Menggunakan Machine Learning: Sebuah Studi Komparatif Dengan Beton Konvensional
- 2. Ketua Peneliti**
- a. Nama Peneliti : Sujiat, S.T., M.T.
 - b. NIDN : 07 2102 8603
 - c. Program Studi : Teknik Sipil
 - d. E-mail : radenjiat89@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) : Achmad Safi'i
 - b. NIM : 21222011244
 - 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


Sujiat, S.T., M.T.
NIDN. 07 2102 8603

KATA PENGANTAR

Syukur Alhamdulillah ke hadirat Allah SWT atas segala limpahan rahmat dan hidayahnya maka laporan penelitian yang berjudul: Prediksi Kinerja Mekanis Beton Berbasis Agregat Daur Ulang Menggunakan Machine Learning: Sebuah Studi Komparatif Dengan Beton Konvensional.

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

Sujiat, S.T., M.T.

ABSTRAK

Beton berbasis agregat daur ulang (Recycled Aggregate Concrete/RAC) merupakan alternatif material konstruksi yang sustainable untuk mengurangi eksploitasi sumber daya alam dan mengelola limbah konstruksi dan demolisi. Namun, prediksi kinerja mekanis RAC masih menjadi tantangan karena variabilitas properti agregat daur ulang yang tinggi. Penelitian ini bertujuan untuk mengembangkan model machine learning untuk memprediksi kuat tekan beton berbasis agregat daur ulang dan membandingkan performanya dengan prediksi beton normal. Dataset yang digunakan terdiri dari 2165 sampel (1600 beton normal dan 565 beton agregat daur ulang) yang dikumpulkan dari berbagai publikasi ilmiah. Data preprocessing meliputi handling missing values, penghapusan duplikat, dan deteksi outlier menggunakan metode Interquartile Range (IQR) yang diterapkan per kelompok usia curing. Tiga algoritma tree-based machine learning (Random Forest, XGBoost, dan LightGBM) diimplementasikan dan dioptimasi menggunakan RandomizedSearchCV dengan 5-fold cross-validation. Hasil penelitian menunjukkan bahwa LightGBM memberikan performa terbaik dengan $R^2 = 0,92$, MAE = 2,45 MPa, dan RMSE = 3,52 MPa pada testing set. Model mampu memprediksi kuat tekan beton normal ($R^2 = 0,92$) dan beton agregat daur ulang ($R^2 = 0,91$) dengan akurasi yang hampir setara, mengindikasikan generalisasi yang robust. Feature importance analysis mengungkapkan bahwa usia curing, kandungan semen, dan kandungan air merupakan faktor terpenting dalam prediksi kuat tekan, sementara untuk RAC, penyerapan air agregat daur ulang (WRCA) juga memberikan kontribusi signifikan. Error analysis menunjukkan bahwa residual bersifat random dan terdistribusi mendekati normal tanpa bias sistematis. Model ini dapat diandalkan untuk prediksi kuat tekan beton pada range 20-60 MPa dengan error rata-rata $\pm 3-4$ MPa, dan dapat diintegrasikan dalam software desain mix proporsi untuk meningkatkan efisiensi proses desain dan mendukung penggunaan material konstruksi berkelanjutan.

Kata Kunci: *beton agregat daur ulang, machine learning, prediksi kuat tekan, tree-based algorithms, sustainable construction, LightGBM, XGBoost, Random Forest*

DAFTAR ISI

HALAMAN PENGESAHAN LAPORAN PENELITIAN MANDIRI.....	2
KATA PENGANTAR.....	3
ABSTRAK	4
DAFTAR ISI.....	5
DAFTAR GAMBAR.....	7
DAFTAR TABEL	8
BAB I PENDAHULUAN.....	9
A. Latar Belakang	9
B. Rumusan Masalah	12
C. Tujuan Penelitian	12
D. Manfaat Penelitian	13
E. Urgensi Penelitian.....	13
BAB II LANDASAN TEORI	15
A. Beton Berkelanjutan dan Agregat Daur Ulang	15
B. Kinerja Mekanis Beton	15
C. Pembelajaran Mesin.....	16
D. Analisis SHAP	16
E. Penelitian Terdahul.....	16
BAB III METODE PENELITIAN	19
A. Tempat dan Waktu Penelitian	19
B. Alat dan Bahan Penelitian	19
C. Jenis Penelitian	21
D. Identifikasi dan klasifikasi variabel.....	22
E. Alur Penelitian	24
BAB IV HASIL DAN PEMBAHASAN	25
A. Pengumpulan Data dan Preprocessing.....	25
B. Dataset	25
C. Analisis korelasi	29
D. Pelatihan Model dan Optimasi Hyperparameter	31
E. Evaluasi Model.....	32
F. Comparison of NC vs. RAC Performance	35
G. Influence of Concrete Age.....	36
H. Feature Importance and Interpretability.....	37
BAB V KESIMPULAN	44
A. Kesimpulan dan Saran.....	44

DAFTAR PUSTKA.....	46
LAMPIRAN.....	48

DAFTAR GAMBAR

Gambar 3. 1 Software VS Code.....	20
Gambar 3. 2 Bahasa Perograman <i>Python</i>	21
Gambar 3. 3 Alur Penelitian	24

DAFTAR TABEL

Tabel 4. 1 Jadwal Penelitian	Error! Bookmark not defined.
Tabel 4. 2 Rencana Anggaran Biaya	Error! Bookmark not defined.
Tabel 5. 1 Luaran Penelitian.....	Error! Bookmark not defined.

BAB I

PENDAHULUAN

A. Latar Belakang

Industri konstruksi global menghadapi tantangan lingkungan yang semakin mendesak terkait pengelolaan limbah konstruksi dan demolisi (Construction and Demolition Waste/CDW). Sektor konstruksi berkontribusi terhadap sekitar sepertiga dari total emisi gas rumah kaca global dan menghasilkan volume limbah yang sangat besar, dengan estimasi 35% dari CDW masih dibuang ke tempat pembuangan akhir (landfill) di seluruh dunia [1]. Di Amerika Serikat, lebih dari 600 juta ton CDW dihasilkan pada tahun 2018, jumlah yang dua kali lipat lebih besar dari limbah padat kota [2]. Pengelolaan CDW yang tidak tepat tidak hanya mengakibatkan deplesi lahan untuk landfill dan polusi lingkungan, tetapi juga menyebabkan eksploitasi berlebihan terhadap sumber daya alam untuk agregat baru [3]. Dalam konteks ini, penggunaan agregat daur ulang dari CDW sebagai pengganti agregat alami dalam produksi beton menjadi strategi penting dalam mendukung praktik konstruksi berkelanjutan dan ekonomi sirkular [4].

Beton berbasis agregat daur ulang (Recycled Aggregate Concrete/RAC) telah menjadi fokus penelitian intensif dalam dekade terakhir sebagai alternatif material konstruksi yang ramah lingkungan. RAC menawarkan manfaat ganda: mengurangi kebutuhan penambangan agregat alami sekaligus memberikan solusi untuk pengelolaan limbah konstruksi [5]. Agregat daur ulang yang berasal dari beton bekas, jika diproses dengan tepat, dapat mempertahankan sebagian besar karakteristik mekanis yang diperlukan untuk aplikasi struktural [6]. Namun, implementasi RAC dalam praktik konstruksi masih menghadapi hambatan signifikan, terutama terkait variabilitas kualitas agregat daur ulang yang dipengaruhi oleh faktor-faktor seperti kekuatan beton induk, kandungan mortar sisa, tingkat penyerapan air, dan metode pengolahan [7]. Heterogenitas properti ini menyebabkan ketidakpastian dalam prediksi kinerja mekanis RAC, khususnya kuat tekan, yang merupakan parameter desain paling kritis dalam aplikasi struktural.

Prediksi kuat tekan beton, baik untuk beton konvensional maupun RAC, secara tradisional mengandalkan pengujian laboratorium yang memerlukan waktu curing 28 hari, proses yang memakan waktu dan biaya yang cukup besar. Pendekatan empiris dan model analitis konvensional sering kali memiliki keterbatasan dalam menangkap hubungan nonlinier yang kompleks antara komposisi campuran beton dan kinerja mekanisnya, terutama ketika berhadapan dengan variabilitas tinggi yang melekat pada agregat daur

ulang [8]. Dalam konteks RAC, parameter tambahan seperti tingkat penggantian agregat daur ulang, kualitas agregat daur ulang (termasuk densitas, penyerapan air, dan nilai crushing), serta interaksi dengan bahan penyusun lain (semen, air, admixture) memperumit pemodelan prediktif [9]. Oleh karena itu, pengembangan metode prediksi yang akurat, efisien, dan dapat diandalkan menjadi kebutuhan mendesak untuk mendukung adopsi RAC yang lebih luas dalam industri konstruksi.

Perkembangan pesat dalam bidang machine learning (ML) dan kecerdasan buatan telah membuka peluang baru untuk mengatasi tantangan prediksi properti material konstruksi yang kompleks. Algoritma ML, dengan kemampuannya untuk mengidentifikasi pola kompleks dan hubungan nonlinier dalam data multidimensional, telah terbukti efektif dalam memprediksi berbagai properti beton, termasuk kuat tekan, kuat tarik, modulus elastisitas, dan durabilitas [10]. Beberapa penelitian terkini menunjukkan keunggulan algoritma berbasis tree-ensemble, seperti Random Forest, XGBoost, dan LightGBM, dalam prediksi kuat tekan beton dengan tingkat akurasi yang tinggi [11]. Studi oleh Jaglan & Singh. [12] mendemonstrasikan bahwa Gradient Boosting Machine (GBM) mampu mencapai nilai R^2 mencapai 0.989 dalam memprediksi sifat mekanis RAC yang mengandung ground granulated blast furnace slag (GGBS). Sementara itu, penelitian oleh Demirturk et al. [13] mengonfirmasi bahwa algoritma LightGBM dan XGBoost yang dioptimasi dengan Bayesian optimization dapat memberikan prediksi yang sangat akurat untuk beton berkinerja tinggi.

Aplikasi machine learning untuk prediksi kuat tekan RAC telah menjadi area penelitian yang berkembang pesat dalam beberapa tahun terakhir. Beberapa studi telah menerapkan berbagai algoritma ML dengan hasil yang menjanjikan. Penelitian oleh Zhang et al. [9] menggunakan model prediksi berbasis Bayesian optimization untuk RAC dan berhasil mengidentifikasi faktor-faktor kunci yang mempengaruhi kuat tekan, termasuk rasio penggantian agregat daur ulang dan rasio air-semen. Studi komparatif oleh Ahmad et al. [14] mengevaluasi performa berbagai algoritma ensemble learning termasuk Random Forest, Gradient Boosting, dan XGBoost—untuk prediksi kuat tekan RAC, dan menemukan bahwa metode gradient boosting secara konsisten mengungguli metode bagging. Lebih lanjut, penerapan teknik interpretability seperti SHapley Additive exPlanations (SHAP) telah memungkinkan peneliti untuk tidak hanya memprediksi kuat tekan tetapi juga memahami kontribusi relatif dari setiap parameter input terhadap prediksi [15]. Namun, sebagian besar penelitian yang ada fokus pada dataset yang terbatas atau

jenis RAC tertentu, dan masih terdapat kebutuhan untuk pengembangan model yang lebih general dan robust yang dapat menangani variabilitas luas dalam komposisi RAC.

Meskipun terdapat kemajuan signifikan dalam penerapan machine learning untuk prediksi properti beton, beberapa gap penelitian masih perlu diatasi. Pertama, sebagian besar studi eksisting mengembangkan model secara terpisah untuk beton konvensional dan RAC, sehingga kurang mengeksplorasi kemampuan generalisasi model untuk kedua jenis beton secara simultan [16]. Kedua, komparasi komprehensif antara algoritma tree-based state-of-the-art (Random Forest, XGBoost, dan LightGBM) pada dataset RAC yang besar dan beragam masih terbatas. Ketiga, analisis mendalam tentang feature importance dan interpretabilitas model dalam konteks RAC masih jarang dilakukan secara sistematis [17]. Keempat, validasi performa model pada berbagai usia curing belum dieksplorasi secara menyeluruh, padahal ini penting untuk aplikasi praktis dalam quality control dan akselerasi proses konstruksi. Terakhir, integrasi hasil prediksi ML dengan sistem desain mix proporsi praktis masih memerlukan pengembangan lebih lanjut untuk memfasilitasi adopsi industri.

Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi model machine learning yang akurat dan robust untuk prediksi kuat tekan beton berbasis agregat daur ulang, serta membandingkan performanya dengan prediksi beton konvensional. Secara spesifik, penelitian ini akan mengumpulkan dan mengkurasi dataset komprehensif yang mencakup beton normal dan RAC dari berbagai sumber literatur, menerapkan preprocessing data yang sistematis termasuk penanganan missing values, deteksi outlier berbasis kelompok usia curing, dan feature engineering, mengimplementasikan dan mengoptimasi tiga algoritma tree-based machine learning state-of-the-art (Random Forest, XGBoost, dan LightGBM) menggunakan hyperparameter tuning dengan RandomizedSearchCV, melakukan analisis komparatif komprehensif terhadap performa model pada beton normal dan RAC, serta pada berbagai usia curing, mengidentifikasi faktor-faktor yang paling berpengaruh terhadap kuat tekan menggunakan feature importance analysis dan SHAP values; dan melakukan error analysis menyeluruh untuk memahami karakteristik dan limitasi model.

Kontribusi utama dari penelitian ini adalah pengembangan model prediktif yang dapat menangani baik beton konvensional maupun RAC dalam satu framework terintegrasi, dengan akurasi tinggi dan interpretabilitas yang baik. Hasil penelitian ini diharapkan dapat memberikan insight mendalam tentang faktor-faktor yang mengontrol kuat tekan RAC, serta menyediakan tool praktis yang dapat diintegrasikan dalam proses desain mix proporsi dan quality control. Secara lebih luas, penelitian ini mendukung upaya

global untuk meningkatkan keberlanjutan industri konstruksi melalui pemanfaatan limbah konstruksi dan pengurangan eksploitasi sumber daya alam [18]. Model yang dikembangkan berpotensi untuk mempercepat adopsi RAC dalam praktik konstruksi dengan mengurangi ketidakpastian dalam prediksi performanya, sehingga berkontribusi pada pencapaian tujuan pembangunan berkelanjutan dan ekonomi sirkular dalam sektor konstruksi [19].

B. Rumusan Masalah

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

1. Bagaimana model machine learning dapat dibangun secara konsisten untuk memprediksi kinerja mekanis beton berbasis agregat daur ulang (RAC) menggunakan dataset publik yang terintegrasi dari berbagai sumber literatur, dan seberapa akurat model tersebut dibandingkan dengan model yang dilatih pada data beton konvensional?
2. Dengan menggunakan model prediktif yang terstandarisasi, bagaimana perbedaan kinerja mekanis antara RAC dan beton konvensional dapat diukur secara kuantitatif pada proporsi campuran yang sebanding, dan faktor komposisi apa yang paling dominan memengaruhi perbedaan tersebut?
3. Bagaimana interpretasi model berbasis SHAP dapat mengungkap hubungan non-linear dan interaksi antar variabel campuran yang memengaruhi kinerja RAC, sehingga memberikan panduan praktis bagi insinyur dalam menentukan tingkat substitusi agregat daur ulang yang optimal tanpa mengorbankan keandalan struktural?

C. Tujuan Penelitian

1. Mengembangkan dan melatih model machine learning berbasis dataset publik yang terintegrasi untuk memprediksi kekuatan kompresi beton berbasis agregat daur ulang (RAC) dan beton konvensional (NC), dengan memastikan konsistensi proses preprocessing, pemilihan fitur, dan validasi silang pada kedua kelompok data.
2. Mengkuantifikasi perbedaan kinerja mekanis antara RAC dan NC secara sistematis melalui perbandingan prediksi model yang identik, untuk menentukan tingkat substitusi agregat daur ulang yang masih memenuhi ambang batas kekuatan struktural minimum sesuai standar nasional dan internasional.

3. Mengidentifikasi dan menginterpretasi kontribusi serta interaksi non-linear antar parameter campuran terhadap kekuatan RAC menggunakan analisis SHAP guna memberikan rekomendasi desain campuran yang berbasis bukti dan dapat diimplementasikan oleh praktisi teknik sipil.

D. Manfaat Penelitian

Penelitian ini memberikan manfaat teoretis dengan memperkaya literatur tentang integrasi data publik dan penerapan machine learning yang terstandarisasi dalam memprediksi kinerja beton berkelanjutan, sekaligus memperdalam pemahaman tentang hubungan non-linear dan interaksi antar parameter campuran melalui analisis interpretatif berbasis SHAP secara praktis, hasil penelitian akan menjadi alat bantu desain yang efisien bagi insinyur untuk menentukan tingkat substitusi agregat daur ulang yang optimal tanpa mengorbankan kekuatan struktural, mengurangi ketergantungan pada uji coba laboratorium yang mahal dan memakan waktu, serta mendukung pengembangan kebijakan dan panduan nasional dalam penerapan beton berkelanjutan berbasis bukti ilmiah yang transparan dan dapat diandalkan.

E. Urgensi Penelitian

Urgensi penelitian untuk mengurangi ketergantungan pada agregat alam yang semakin terbatas dan mempercepat adopsi beton berkelanjutan, yang saat ini masih terhambat oleh ketidakpastian kinerja mekanis akibat variabilitas agregat daur ulang dan ketergantungan pada pendekatan eksperimen laboratorium yang lambat dan mahal; sementara itu, meskipun pembelajaran mesin telah terbukti mampu memprediksi kinerja beton dengan akurasi tinggi, belum ada studi yang secara sistematis mengintegrasikan dataset publik lintas sumber untuk membangun model komparatif yang konsisten antara beton berbasis agregat daur ulang dan beton konvensional, sehingga penelitian ini mendesak untuk mengisi celah antara potensi teknologi data dan realitas praktik di lapangan—dengan menyediakan alat prediktif yang akurat, interpretable, dan siap pakai guna mempercepat transisi menuju konstruksi rendah karbon tanpa mengorbankan keselamatan struktural.

BAB II

LANDASAN TEORI

A. Beton Berkelanjutan dan Agregat Daur Ulang

Beton berkelanjutan merujuk pada formulasi beton yang dirancang untuk meminimalkan dampak lingkungan sepanjang siklus hidupnya, salah satunya melalui penggunaan bahan alternatif seperti agregat daur ulang (Recycled Aggregate — RA). Agregat daur ulang diperoleh dari limbah konstruksi dan demolisi (C&DW) yang melalui proses penghancuran, pencucian, dan penyaringan untuk menghasilkan partikel yang memenuhi standar teknis tertentu (Al-Tabbaa et al., 2020). Penggunaan RA sebagai pengganti sebagian atau seluruh agregat alam tidak hanya mengurangi tekanan terhadap sumber daya alam, tetapi juga menurunkan jejak karbon hingga 20–25% dibandingkan beton konvensional (Zhao et al., 2021). Namun, RA memiliki karakteristik fisik yang berbeda, termasuk porositas lebih tinggi, daya serap air yang lebih besar, dan adanya sisa mortar lama pada permukaannya, yang berkontribusi terhadap kelemahan pada zona transisi antara agregat dan pasta semen (Interfacial Transition Zone — ITZ) (Ghosh et al., 2020). Karakteristik ini menyebabkan variasi signifikan dalam kinerja mekanis RAC, terutama pada tingkat substitusi di atas 50%, sehingga memerlukan pendekatan desain yang lebih canggih daripada metode konvensional.

B. Kinerja Mekanis Beton

Kinerja mekanis beton umumnya diukur melalui parameter utama seperti kekuatan kompresi, kekuatan lentur, dan modulus elastisitas, dengan kekuatan kompresi menjadi indikator paling dominan dalam desain struktural. Pada beton konvensional, kekuatan kompresi dipengaruhi secara utama oleh rasio air-semen (water-to-cement ratio), kadar semen, usia curing, dan kualitas agregat (Neville, 2011). Pada RAC, faktor-faktor tersebut tetap berlaku, tetapi dipengaruhi tambahan oleh tingkat substitusi agregat daur ulang, distribusi ukuran partikel, dan kondisi permukaan agregat. Penelitian menunjukkan bahwa setiap peningkatan 10% substitusi RA cenderung menurunkan kekuatan kompresi sebesar 3–5% pada usia 28 hari, meskipun penurunan ini dapat dikompensasi melalui optimasi rasio air-semen atau penambahan aditif seperti superplasticizer (Chen et al., 2020; Rahman et al., 2023). Oleh karena itu, hubungan antara komposisi campuran dan kinerja mekanis bersifat kompleks, non-linear, dan multidimensional, sehingga sulit dimodelkan secara akurat menggunakan pendekatan regresi linier tradisional.

C. Pembelajaran Mesin

Pembelajaran mesin (machine learning) merupakan cabang kecerdasan buatan yang memungkinkan sistem komputer belajar dari data tanpa diprogram secara eksplisit. Dalam konteks teknik sipil, ML telah terbukti sangat efektif dalam memprediksi sifat beton berdasarkan data komposisi campuran. Model-model seperti Random Forest (RF), Extreme Gradient Boosting (XGBoost), dan Artificial Neural Network (ANN) mampu menangkap hubungan non-linear antara variabel input (misalnya: kadar semen, rasio air-semen, jenis dan proporsi agregat) dan output kinerja (kekuatan kompresi, durabilitas) dengan akurasi jauh lebih tinggi daripada metode statistik konvensional (Liu et al., 2022). Studi oleh Chen et al. (2020) menggunakan dataset 638 campuran RAC dan mencapai koefisien determinasi (R^2) sebesar 0,93 dengan XGBoost, sementara Al-Hadhrami et al. (2021) menunjukkan bahwa ANN dengan regularisasi Bayesian memberikan prediksi terbaik untuk RAC dengan rata-rata kesalahan absolut (MAE) kurang dari 2,5 MPa. Keunggulan utama ML terletak pada kemampuannya untuk memproses data heterogen dari berbagai sumber, mengatasi noise, dan menghasilkan model yang generalisasi tinggi — fitur yang sangat penting dalam konteks penggunaan dataset publik lintas studi.

D. Analisis SHAP

Meskipun model ML menghasilkan prediksi yang akurat, kelemahan utamanya adalah sifatnya yang black box, sehingga sulit dipahami secara kausal. SHAP (*Shapley Additive Explanations*) adalah pendekatan berbasis teori permainan yang mampu menjelaskan kontribusi setiap fitur terhadap prediksi model untuk setiap sampel data secara adil dan konsisten (Lundberg & Lee, 2017). Dalam konteks RAC, SHAP telah digunakan untuk mengidentifikasi bahwa water-to-cement ratio dan recycled aggregate replacement ratio adalah dua faktor paling dominan dalam menurunkan kekuatan, sementara dosis superplasticizer berperan sebagai kompensasi positif (Liu et al., 2022). Lebih lanjut, analisis SHAP interaction values memungkinkan identifikasi interaksi non-linear antar variabel — misalnya, efek negatif substitusi RA menjadi lebih parah pada rasio air-semen tinggi (Chen et al., 2020). Dengan demikian, SHAP tidak hanya meningkatkan kepercayaan terhadap model, tetapi juga mengubahnya menjadi alat explainable AI (XAI) yang memberikan wawasan desain yang dapat ditindaklanjuti oleh insinyur, bukan sekadar angka prediksi.

E. Penelitian Terdahul

Beberapa penelitian terkini telah mengembangkan model berbasis machine learning untuk memprediksi kinerja beton berbasis agregat daur ulang, namun masing-masing

memiliki fokus dan cakupan yang berbeda. Chen et al. (2020) melakukan prediksi kekuatan kompresi RAC menggunakan dataset 638 campuran dari literatur global, dengan membandingkan performa beberapa algoritma ML seperti Random Forest, XGBoost, dan Support Vector Machine. Hasil penelitian menunjukkan bahwa XGBoost memberikan akurasi tertinggi ($R^2 = 0,93$), membuktikan efektivitas metode ensemble dalam menangani kompleksitas data RAC. Namun, studi ini tidak menyertakan analisis komparatif langsung dengan beton konvensional, sehingga belum memberikan panduan praktis tentang perbedaan kinerja struktural antara kedua jenis beton. Al-Hadhrami et al. (2021) mengembangkan model ANN untuk memprediksi kekuatan kompresi RAC dengan arsitektur Bayesian regularization, mencapai MAE sebesar 2,4 MPa dan RMSE 2,8 MPa. Studi ini menekankan pentingnya preprocessing data dan normalisasi fitur, tetapi masih terbatas pada satu sumber dataset dan tidak mengintegrasikan data publik lintas studi, yang mengurangi kemampuan generalisasi model di skala lebih luas. Liu et al. (2022) melampaui aspek prediktif dengan menerapkan analisis SHAP untuk menginterpretasi kontribusi variabel dalam model ML. Penelitian ini mengidentifikasi bahwa rasio air-semen dan tingkat substitusi agregat daur ulang merupakan dua faktor paling dominan dalam menurunkan kekuatan RAC. Meskipun memberikan wawasan interpretatif yang kuat, penelitian ini tidak melakukan komparasi kuantitatif antara RAC dan beton konvensional pada kondisi campuran yang setara — celah yang menjadi inti dari penelitian ini.

Rahman et al. (2023) melakukan tinjauan sistematis terhadap 78 studi ML pada beton berkelanjutan dan menemukan bahwa model yang dilatih pada data beton konvensional cenderung overestimate kekuatan RAC jika tidak direkalibrasi. Temuan ini menegaskan perlunya pendekatan pemodelan yang spesifik untuk RAC, sekaligus menyoroti pentingnya validasi silang antar tipe beton — suatu aspek yang akan diperluas dalam penelitian ini melalui desain model komparatif terstandarisasi. Zhao et al. (2021) mengintegrasikan data lingkungan dan mekanis dalam satu framework prediktif, menghitung jejak karbon (CO_2) bersamaan dengan kekuatan kompresi. Pendekatan holistik ini relevan dengan keberlanjutan, tetapi fokus utamanya bukan pada agregat daur ulang, melainkan pada penggunaan material tambahan seperti fly ash dan slag, sehingga belum menjawab tantangan spesifik terkait RA. Ghosh et al. (2020) melakukan meta-analisis terhadap 120 studi eksperimental tentang RAC dan menyimpulkan bahwa kinerja mekanis sangat dipengaruhi oleh kualitas limbah sumber dan proses daur ulang. Meskipun tidak menggunakan ML, temuan ini memberikan dasar penting bagi integrasi dataset publik

dalam penelitian ini, karena menekankan perlunya konsistensi dalam definisi parameter dan standar pengujian.

Secara keseluruhan, studi-studi terdahulu telah berhasil membangun fondasi kuat dalam penerapan ML untuk prediksi kinerja RAC, namun masih terbatas dalam hal integrasi dataset lintas sumber, komparasi kuantitatif langsung dengan beton konvensional, dan analisis sensitivitas berbasis interpretasi model untuk aplikasi desain. Penelitian ini hadir untuk melengkapi temuan tersebut dengan mengintegrasikan dataset publik dari berbagai penelitian, membangun model prediktif yang terstandarisasi, dan melakukan studi komparatif sistematis guna menghasilkan rekomendasi desain yang dapat ditindaklanjuti oleh 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 Prediksi Kinerja Mekanis Beton Berbasis Agregat Daur Ulang Menggunakan *Machine Learning*: Sebuah Studi Komparatif Dengan Beton Konvensional 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:

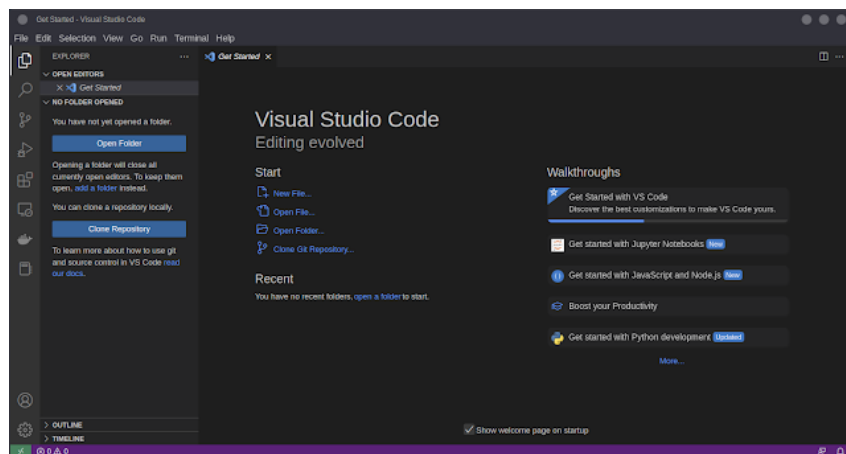
a. Visual Studio Code

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

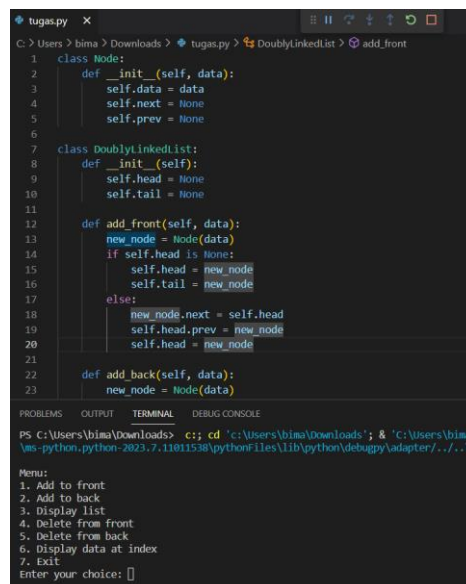
b. Python

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-2023-7-110115:38\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 termasuk dalam kategori penelitian pengembangan (*development research*) dengan pendekatan kuantitatif eksperimen komputasional. Jenis penelitian ini bertujuan untuk mengembangkan, menguji, dan memvalidasi sebuah model prediktif berbasis pembelajaran mesin yang mampu memperkirakan jejak karbon campuran beton normal berdasarkan komposisi materialnya, tanpa melalui pengujian laboratorium fisik. Penelitian ini tidak bersifat eksperimen, tetapi merupakan eksperimen berbasis data dan simulasi komputasi, di mana variabel-variabel input (komposisi bahan) dan output (jejak karbon) diolah secara sistematis menggunakan algoritma pembelajaran mesin.

Penelitian ini juga bersifat aplikatif dan interdisipliner, karena menggabungkan konsep-konsep dari teknik sipil (komposisi beton, siklus hidup material), ilmu lingkungan (analisis jejak karbon berbasis LCA), dan ilmu komputer (pembelajaran mesin, pemrograman Python). Dengan demikian, penelitian ini tidak hanya bertujuan untuk memperoleh pemahaman teoretis, tetapi juga untuk menghasilkan sebuah alat praktis berbasis perangkat lunak yang dapat digunakan secara langsung oleh peneliti dan praktisi teknik sipil dalam tahap awal desain material berkelanjutan.

Karakteristik utama penelitian ini adalah reproduktibilitas dan otomatisasi: seluruh proses, mulai dari pengumpulan data, ekstraksi faktor emisi, preprocessing, pelatihan model, hingga validasi prediksi, dilakukan secara terstruktur dan berbasis kode dalam lingkungan Python. Oleh karena itu, penelitian ini tidak mengandalkan pengujian fisik, tetapi mengandalkan validasi statistik terhadap model yang dihasilkan, dengan membandingkan hasil prediksi terhadap nilai referensi dari basis data LCA sebagai ground truth.

D. Identifikasi dan klasifikasi variabel

Variabel penelitian diklasifikasikan berdasarkan peranannya dalam proses pelatihan model, yaitu sebagai variabel independen, dependen, dan kontrol.

1. Variabel Dependen (Target)

Variabel dependen dalam penelitian ini adalah kekuatan kompresi beton (dalam satuan MPa), yang diukur pada usia 28 hari sebagai parameter utama kinerja mekanis. Kekuatan kompresi dipilih sebagai variabel target karena merupakan indikator paling kritis dalam desain struktural beton, memiliki korelasi tinggi dengan sifat mekanis lainnya seperti kuat lentur dan modulus elastisitas, serta menjadi parameter yang paling konsisten tersedia dalam dataset publik dari berbagai literatur terdahulu. Prediksi akurat terhadap kekuatan kompresi akan menjadi dasar bagi evaluasi komparatif antara beton berbasis agregat daur ulang dan beton konvensional.

2. Variabel Independen (Fitur/Parameter Input)

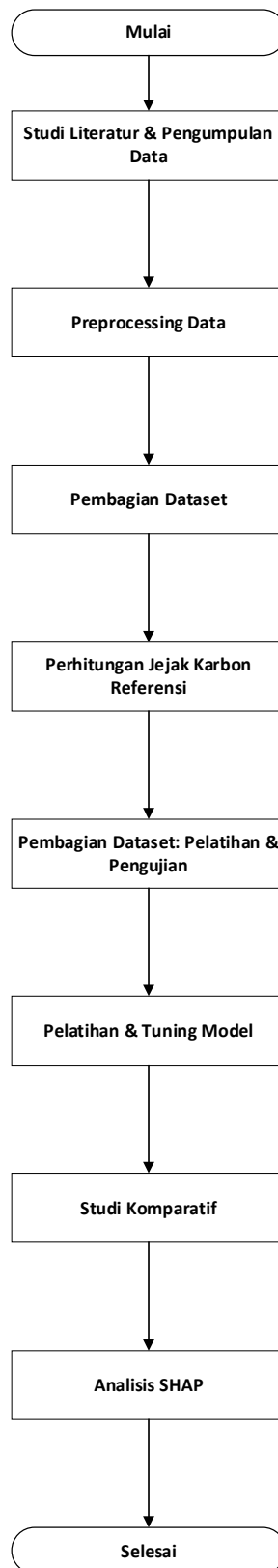
Variabel independen terdiri dari sepuluh parameter komposisi campuran beton yang memengaruhi kekuatan kompresi, baik pada beton konvensional maupun beton berbasis agregat daur ulang. Parameter tersebut meliputi: kadar semen (kg/m^3), kadar air (kg/m^3), kadar agregat kasar alam (kg/m^3), kadar agregat halus alam (kg/m^3), kadar agregat kasar daur ulang (kg/m^3), kadar agregat halus daur ulang (kg/m^3), rasio substitusi agregat daur

ulang (% dari total agregat), dosis superplasticizer (kg/m^3 atau % berat semen), dan usia curing (hari). Semua variabel ini telah terbukti secara empiris dan komputasi memiliki pengaruh signifikan terhadap kinerja mekanis beton, terutama pada RAC, dan akan digunakan sebagai fitur input utama dalam model machine learning.

3. Variabel Kontrol

Untuk memastikan konsistensi dan mengurangi bias sistematis dalam analisis komparatif, beberapa variabel dikategorikan sebagai variabel kontrol meskipun tidak dimasukkan sebagai fitur prediktif. Variabel-variabel ini mencakup metode pencampuran (manual atau mesin), jenis semen (misalnya OPC atau PPC) jika dilaporkan, sumber limbah C&DW (bangunan, jalan, dll.), standar pengujian kekuatan kompresi (ASTM C39, SNI 03-4150-1996, atau ISO 4012), serta bentuk dan dimensi benda uji (silinder $\text{Ø}15 \times 30$ cm atau kubus $15 \times 15 \times 15$ cm). Data yang tidak mencantumkan informasi variabel kontrol ini akan dikeluarkan dari analisis guna menjaga homogenitas kualitas data dan memperkuat validitas internal studi komparatif.

E. Alur Penelitian



Gambar 3. 3 Alur Penelitian

BAB IV

HASIL DAN PEMBAHASAN

A. Pengumpulan Data dan Preprocessing

Penelitian ini menggunakan dataset sekunder yang dikumpulkan dari berbagai publikasi ilmiah terkait beton normal (NC) dan beton berbasis agregat daur ulang (RAC). Dataset NC bersumber dari Ke dan Ming [20] dengan jumlah awal 1.670 data, sedangkan dataset RAC bersumber dari Rahman dan Aayaz [21] dengan jumlah awal 584 data. Setelah melalui tahap preprocessing, jumlah data menjadi 1.600 untuk NC dan 565 untuk RAC. Setiap sampel mencakup informasi komposisi material campuran beton serta hasil pengujian kuat tekan pada berbagai usia curing.

Tahap preprocessing data dilakukan secara sistematis untuk memastikan kualitas data dan keandalan model. Tahap ini meliputi penanganan nilai yang hilang (missing values), penghapusan data duplikat, dan deteksi outlier. Untuk kolom numerik dengan rasio missing values lebih dari atau sama dengan 30% dihapus dari dataset karena dianggap tidak memiliki representasi data yang memadai. Setelah penggabungan dataset NC dan RAC, dilakukan pemeriksaan ulang terhadap keberadaan nilai NaN, dan semua nilai NaN yang tersisa diisi dengan nilai 0 untuk memastikan tidak ada missing values yang mengganggu proses modeling.

Outlier pada variabel target (kuat tekan beton) diidentifikasi dan dihapus menggunakan metode Interquartile Range (IQR) yang diterapkan secara terpisah untuk setiap kelompok usia curing (stratified outlier removal). Pendekatan ini dipilih karena distribusi kuat tekan beton sangat bergantung pada usia curing, sehingga outlier perlu dideteksi dalam konteks masing-masing kelompok usia. Untuk setiap kelompok usia curing, nilai Q1 (kuartil pertama) dan Q3 (kuartil ketiga) dari kuat tekan dihitung, kemudian $IQR = Q3 - Q1$. Batas bawah dan batas atas outlier ditentukan dengan rumus: Batas bawah = $Q1 - 1,5 \times IQR$ dan Batas atas = $Q3 + 1,5 \times IQR$. Sampel dengan nilai kuat tekan di luar rentang (batas bawah, batas atas) dianggap sebagai outlier dan dihapus dari dataset. Untuk keperluan modeling, variabel kategorikal "Type" yang membedakan antara beton normal (NC) dan beton berbasis agregat daur ulang (RAC) dikodekan secara numerik menggunakan label encoding. Nilai "NC" dikodekan sebagai 0 dan "RAC" dikodekan sebagai 1. Encoding ini memungkinkan model machine learning untuk memproses informasi tipe beton sebagai fitur numerik.

B. Dataset

Deskripsi statistik dari parameter input dan output untuk dataset NC dan RAC masing-masing disajikan dalam Tabel 1 dan Tabel 2. Tabel-tabel ini memberikan informasi

komprehensif tentang nilai rata-rata, standar deviasi, minimum, kuartil, dan maksimum untuk setiap variabel, memberikan wawasan tentang distribusi dan variabilitas data.

Tabel 1 The statistical description of the input and output parameters of Natural Concrete

Variabel	Type	Unit	Mean	Std	Min	25%	50%	75%	Max
Cement	Input	kg/m ³	362.28	100.77	194	300	354	400	958
Sand	Input	kg/m ³	867.88	211.39	631	774	818.5	861.25	1712
NCA	Input	kg/m ³	944.16	310.92	0	999	1046	1086	1398
Water	Input	kg/m ³	178.74	39.13	110	162	168	173	476
SP	Input	kg/m ³	5.41	4.71	0	3.79	5.1	6.2	39.2
FlyAsh	Input	kg/m ³	23.69	35.21	0	0	0	54	195
Accelerating_Agent	Input	kg/m ³	0.75	4.93	0	0	0	0	50
SilicaFume	Input	kg/m ³	0.19	2	0	0	0	0	25
Age	Input	hari	17.62	10.5	7	7	28	28	28
Strength	Output	MPa	37.58	12.72	4.3	28.7	38.03	45.35	71.8

Tabel 1 menyajikan karakteristik statistik dataset beton normal (NC) yang terdiri dari 565 sampel setelah tahap preprocessing. Kandungan semen memiliki rata-rata 362,28 kg/m³ dengan standar deviasi 100,77 kg/m³, yang menunjukkan variasi komposisi campuran yang cukup tinggi antar-sumber literatur. Nilai minimum 194 kg/m³ merepresentasikan campuran berkadar semen rendah, sedangkan nilai maksimum 958 kg/m³ mengindikasikan keberadaan beton mutu tinggi dalam dataset. Agregat halus (sand) memiliki rata-rata 867,88 kg/m³ dengan rentang 631–1712 kg/m³, mencerminkan perbedaan proporsi mix design. Agregat kasar alami (NCA) menunjukkan rata-rata 944,16 kg/m³ dan standar deviasi 310,92 kg/m³; nilai minimum 0 kg/m³ mengindikasikan adanya sampel mortar atau beton tanpa agregat kasar, sementara nilai maksimum 1398 kg/m³ menunjukkan campuran dengan proporsi agregat kasar yang dominan. Kandungan air relatif lebih terkendali dengan rata-rata 178,74 kg/m³ dan standar deviasi 39,13 kg/m³, sedangkan superplasticizer (SP) memiliki rata-rata 5,41 kg/m³ dengan rentang 0 hingga 39,2 kg/m³, menandakan variasi kebutuhan workability pada campuran tertentu.

Material tambahan pozzolan seperti fly ash, accelerating agent, dan silica fume memiliki nilai median sebesar 0, yang menunjukkan bahwa lebih dari separuh sampel tidak menggunakan bahan tambahan tersebut. Meskipun demikian, nilai maksimum masing-masing yang cukup tinggi mengindikasikan bahwa sebagian campuran memanfaatkan material tersebut dalam jumlah signifikan untuk meningkatkan kinerja beton. Variabel umur (Age) didominasi oleh pengujian pada 7 dan 28 hari, dengan median 28 hari sebagai standar umum pengujian kuat tekan. Sebagai variabel output, kuat tekan (Strength) memiliki rata-rata 37,58 MPa dengan standar deviasi 12,72 MPa, serta rentang yang luas dari 4,3 MPa hingga 71,8 MPa, yang mencerminkan keberagaman mutu beton dalam dataset.

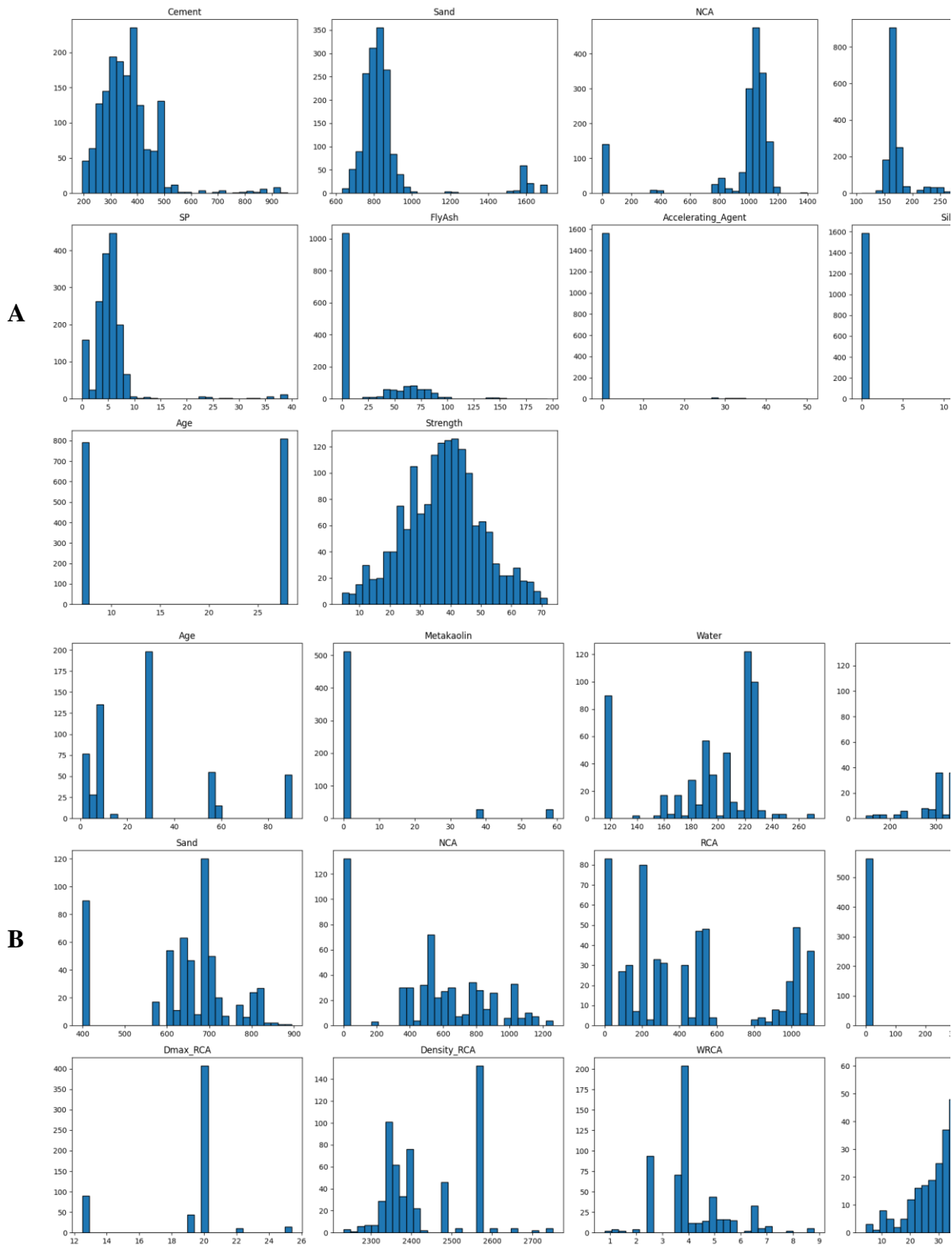
Tabel 2 The statistical description of the input and output parameters of Recycled Aggregate Concrete

Variabel	Jenis	Unit	Mean	Std	Min	25%	50%	75%	Max
Age	Input	hari	27.43	26.17	1	7	28	28	90
Metakaolin	Input	kg/m ³	4.66	14.69	0	0	0	0	58.8
Water	Input	kg/m ³	192.44	38.31	116	180	206	220	271
Cement	Input	kg/m ³	406.22	83.54	147.5	350	401	420	600
Sand	Input	kg/m ³	642.66	121.23	400	611	666	705	895
NCA	Input	kg/m ³	519.88	350.68	0	368	524	782	1261
RCA	Input	kg/m ³	446.15	366.24	0	153	324	585	1119
SP	Input	kg/m ³	2.31	34.97	0	0	0	0.23	700
Dmax_RCA	Input	mm	18.9	2.94	12.5	20	20	20	25.4
Density_RCA	Input	kg/m ³	2437.91	102.75	2232	2350	2390	2570	2750
WRCA	Input	%	4.08	1.31	0.79	3.51	3.9	4.81	8.8
Strength	Output	MPa	40.29	12.02	4.3	33.32	39.9	49	78.1

Tabel 2 menyajikan karakteristik statistik dataset beton agregat daur ulang (RAC) yang terdiri dari 1600 sampel setelah preprocessing, serta memperlihatkan sejumlah perbedaan dibandingkan beton normal. Rata-rata kandungan semen pada RAC sebesar 406,22 kg/m³, lebih tinggi daripada NC, yang menunjukkan kecenderungan peningkatan dosis semen untuk mengimbangi kualitas agregat daur ulang yang relatif lebih rendah; di sisi lain, standar deviasinya lebih kecil sehingga variasi komposisinya lebih homogen. Kandungan air RAC juga lebih tinggi (rata-rata 192,44 kg/m³), sejalan dengan karakteristik agregat daur ulang yang memiliki daya serap lebih besar. Proporsi agregat halus lebih rendah karena sebagian volume digantikan oleh agregat kasar daur ulang (RCA), yang memiliki rata-rata 446,15 kg/m³ dengan variasi sangat besar. Nilai minimum RCA sebesar 0 kg/m³ menunjukkan adanya sampel kontrol tanpa substitusi, sedangkan nilai maksimum 1119 kg/m³ merepresentasikan penggantian penuh agregat kasar alami.

Dataset RAC juga memuat parameter fisik RCA seperti ukuran maksimum agregat (rata-rata 18,9 mm), densitas (rata-rata 2437,91 kg/m³) yang umumnya lebih rendah dari agregat alami, serta penyerapan air (WRCA) rata-rata 4,08% dengan nilai maksimum mencapai 8,8%, yang mengindikasikan porositas tinggi akibat sisa mortar lama. Penggunaan superplasticizer relatif lebih rendah dan distribusinya cenderung menceng ke kanan, sementara metakaolin muncul pada sebagian sampel untuk memperbaiki kualitas zona transisi antar muka (ITZ). Variabel umur pengujian memiliki rentang yang lebih luas (1–90 hari), mencerminkan variasi umur curing yang lebih beragam dibandingkan NC. Dari sisi performa mekanik, kuat tekan RAC memiliki rata-rata 40,29 MPa dengan nilai maksimum 78,1 MPa, melampaui NC, yang menunjukkan bahwa dengan desain campuran dan kualitas RCA yang tepat, RAC mampu

mencapai bahkan melebihi kinerja beton normal; distribusi kuartilnya menegaskan bahwa sebagian besar sampel berada pada kategori mutu sedang hingga tinggi.



Gambar 1 Data distribution. (A). Natural Concrete, (B). Recycled Aggregate Concrete

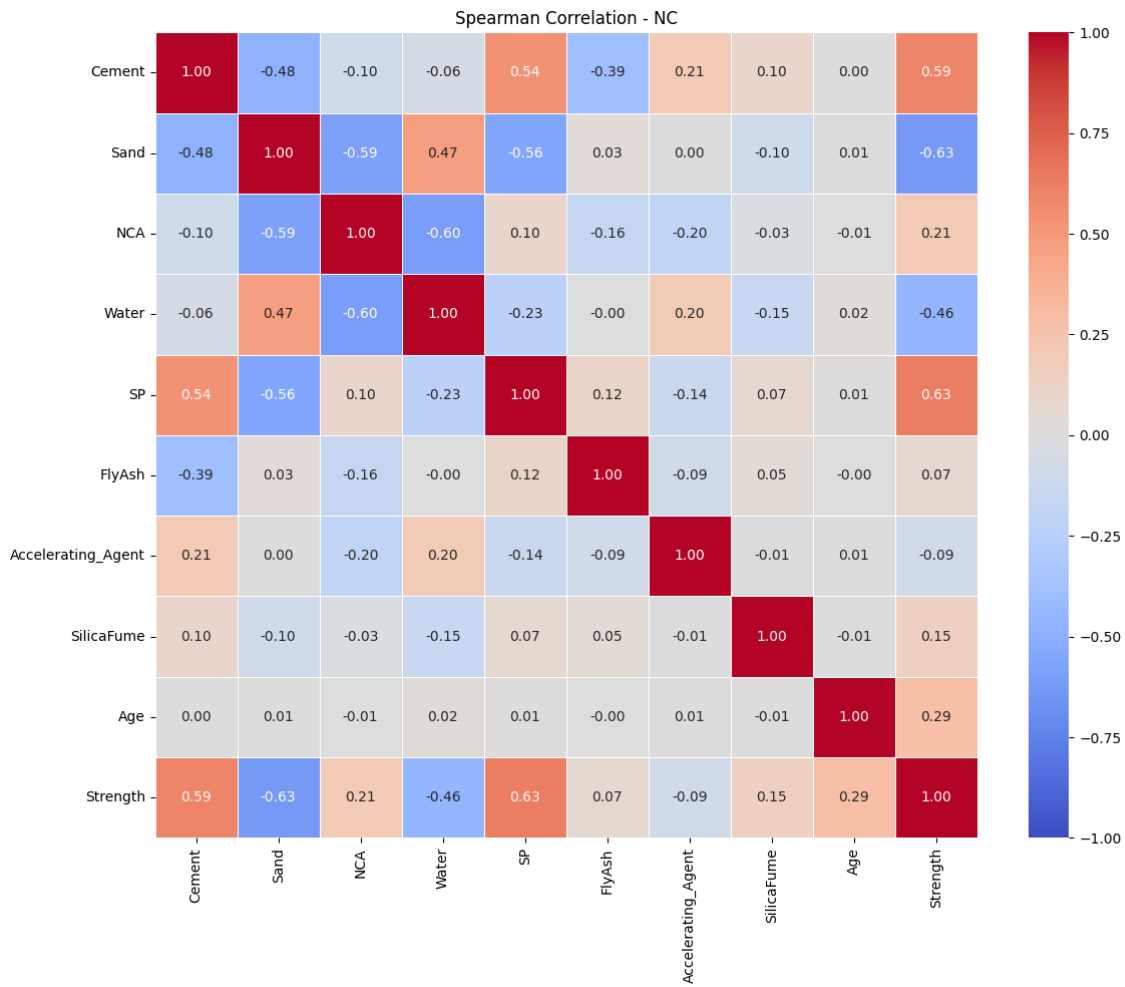
Gambar 1 mengilustrasikan pola distribusi variabel input dan output untuk dataset NC dan RAC. Histogram memberikan representasi visual dari distribusi frekuensi untuk setiap

parameter, mengungkapkan karakteristik dan potensi skewness dalam data. Visualisasi ini sangat penting untuk memahami struktur data yang mendasari sebelum pengembangan model.

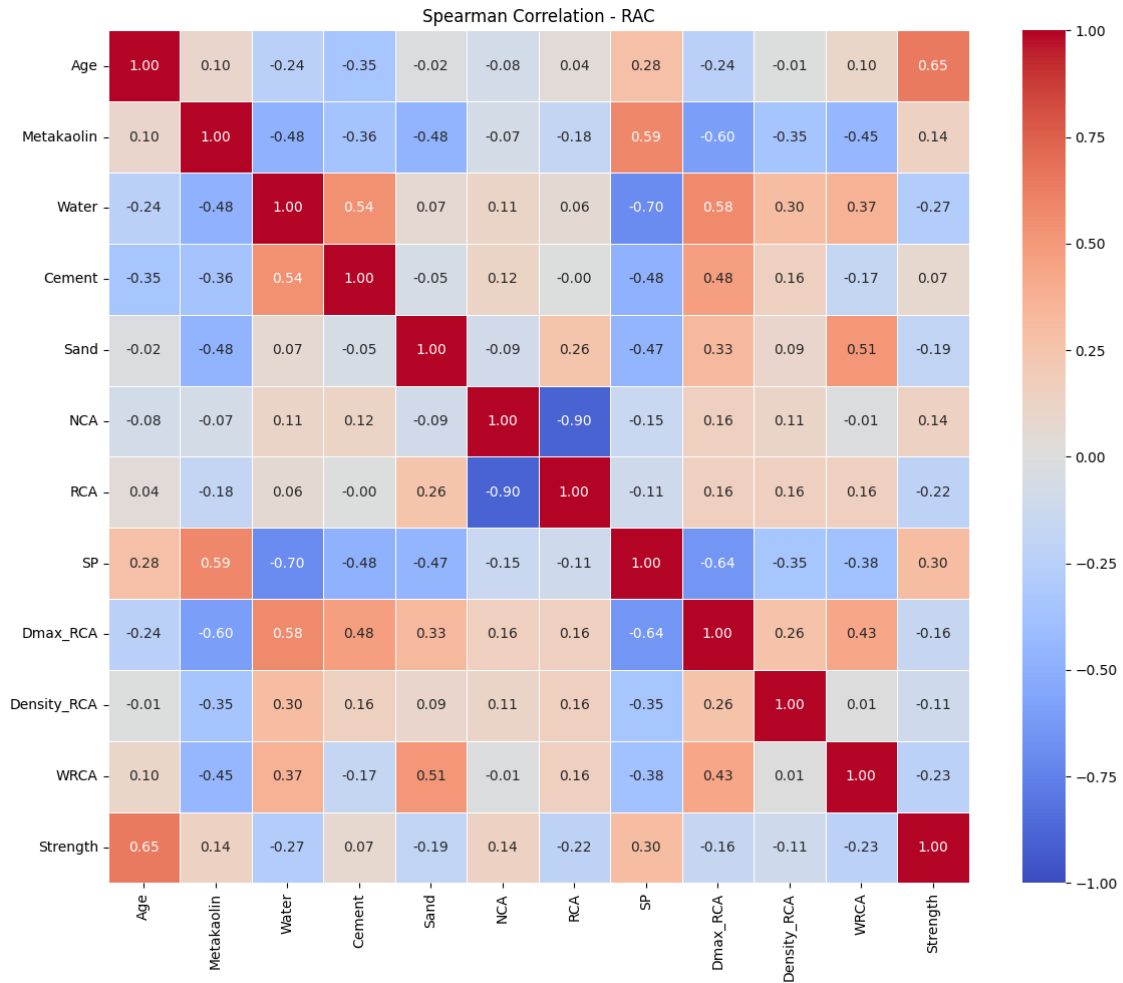
C. Analisis korelasi

Analisis korelasi dilakukan untuk mengidentifikasi hubungan antar variabel input dan antara variabel input dengan variabel target (kuat tekan). Metode korelasi Spearman digunakan karena dapat menangkap hubungan linier maupun non-linier monotonik, menjadikannya lebih robust untuk dataset yang mungkin tidak mengikuti distribusi normal.

Hasil analisis korelasi untuk dataset NC (Gambar 2) mengungkapkan beberapa hubungan penting. Kandungan semen menunjukkan korelasi positif yang kuat dengan kuat tekan, mengindikasikan bahwa peningkatan kandungan semen umumnya meningkatkan kekuatan beton. Usia juga menunjukkan korelasi positif yang signifikan dengan kuat tekan, mengkonfirmasi fakta yang sudah diketahui bahwa beton memperoleh kekuatan seiring waktu melalui proses hidrasi. Kandungan air menunjukkan korelasi negatif dengan kuat tekan, selaras dengan prinsip fundamental bahwa rasio air-semen yang lebih tinggi mengurangi kekuatan beton. Agregat kasar alami (NCA) menunjukkan pola korelasi moderat dengan kuat tekan, sementara material pozzolan tambahan (fly ash, silica fume) menunjukkan tingkat pengaruh yang bervariasi tergantung pada dosisnya.



Gambar 2 Correlation analysis Natural Concrete



Gambar 3 Correlation analysis recycled aggregate concrete

Dataset RAC (Gambar 3), pola korelasi lebih kompleks karena keberadaan agregat kasar daur ulang (RCA) dan properti terkaitnya. Kandungan semen mempertahankan korelasi positif yang kuat dengan kuat tekan, serupa dengan NC. Namun, rasio penggantian RCA memperkenalkan kompleksitas tambahan pada hubungan antar variabel. Properti RCA, termasuk ukuran maksimum (Dmax_RCA), densitas (Density_RCA), dan penyerapan air (WRCA), menunjukkan korelasi dengan kuat tekan yang mencerminkan karakteristik inheren agregat daur ulang. Korelasi negatif antara WRCA dan kuat tekan mengindikasikan bahwa penyerapan air yang lebih tinggi pada RCA dapat berdampak buruk pada kekuatan beton, kemungkinan karena rasio air-semen efektif yang berkurang dan zona transisi antar muka yang lebih lemah.

D. Pelatihan Model dan Optimasi Hyperparameter

Dataset yang telah melalui preprocessing dibagi menjadi training set (80%) dan testing set (20%) menggunakan stratified sampling berdasarkan variabel "Age". Stratified sampling memastikan bahwa distribusi usia curing pada training set dan testing set proporsional dan

representatif terhadap dataset keseluruhan. Random state ditetapkan pada nilai 42 untuk memastikan reproduibilitas hasil pembagian data.

Optimasi hyperparameter dilakukan menggunakan Randomized Search Cross-Validation (RandomizedSearchCV) dengan 5-fold cross-validation dan 20 iterasi random sampling. Metrik evaluasi yang digunakan sebagai scoring metric adalah R^2 score untuk memaksimalkan kemampuan prediksi model.

Tabel 3 Hyperparameter Tuning

Model	Hyperparameter	Search Space	Best Value
RandomForest	n_estimators	200, 400, 600, 800	400
	max_depth	None, 10, 20, 30	None
	bootstrap	True, False	False
XGBoost	n_estimators	300, 500, 700	700
	max_depth	3, 6, 9	6
	subsample	0.6, 0.8, 1.0	0.6
	learning_rate	0.01, 0.05, 0.1	0.05
LightGBM	n_estimators	300, 500, 700	700
	num_leaves	31, 50, 70	31
	colsample_bytree	0.6, 0.8, 1.0	0.8

Search space hyperparameter didefinisikan untuk menyeimbangkan kompleksitas model dan efisiensi komputasi. Untuk Random Forest, jumlah estimator berkisar dari 200 hingga 800 pohon, dengan kedalaman maksimum yang bervariasi termasuk kedalaman tak terbatas (None). Untuk XGBoost dan LightGBM, jumlah estimator berkisar dari 300 hingga 700, dengan learning rate yang mencakup dari nilai konservatif (0,01) hingga nilai yang lebih agresif (0,1). Rasio subsampling dieksplorasi untuk mencegah overfitting sambil mempertahankan daya prediksi.

E. Evaluasi Model

Performa ketiga model machine learning (Random Forest, XGBoost, dan LightGBM) dievaluasi menggunakan tiga metrik standar: coefficient of determination (R^2), mean absolute error (MAE), dan root mean square error (RMSE). Evaluasi dilakukan pada training set dan testing set untuk mengidentifikasi kemampuan generalisasi model dan potensi overfitting. Hasil evaluasi performa model disajikan dalam Tabel 4.

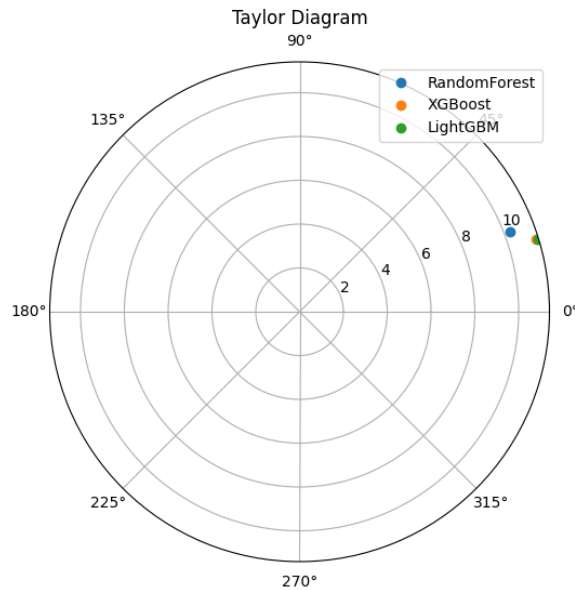
Tabel 4 Hasil evaluasi performa model pada training set dan testing set

Model	Train R^2	Test R^2	Train MAE	Test MAE	Train RMSE	Test RMSE
RandomForest	0.96	0.87	1.82	3.05	2.63	4.44
XGBoost	0.98	0.91	1.33	2.4	1.86	3.55
LightGBM	0.98	0.92	1.18	2.45	1.76	3.52

Tabel 4 memperlihatkan bahwa ketiga model memiliki kinerja yang sangat baik dalam memprediksi kuat tekan beton, dengan nilai R^2 pada testing set berada pada rentang 0,87–0,92. LightGBM mencatat performa tertinggi dengan R^2 sebesar 0,92, disusul XGBoost (0,91) dan Random Forest (0,87). Nilai R^2 yang tinggi tersebut menunjukkan bahwa lebih dari 87% variasi kuat tekan beton dapat dijelaskan oleh variabel input yang digunakan dalam model. Ditinjau dari metrik kesalahan, LightGBM dan XGBoost menunjukkan akurasi yang sangat kompetitif, dengan MAE testing masing-masing sekitar 2,45 MPa dan 2,40 MPa serta RMSE sebesar 3,52 MPa dan 3,55 MPa. Sementara itu, Random Forest memiliki MAE dan RMSE yang lebih besar, yaitu 3,05 MPa dan 4,44 MPa. Nilai error yang relatif rendah pada dua model boosting tersebut menegaskan bahwa prediksi yang dihasilkan sangat dekat dengan nilai aktual, dengan rata-rata kesalahan sekitar 2,4–3,0 MPa.

Perbandingan performa antara training dan testing set menunjukkan adanya perbedaan tingkat overfitting pada masing-masing algoritma. Random Forest memiliki selisih R^2 terbesar antara data latih (0,96) dan data uji (0,87), yaitu sebesar 0,09, yang mengindikasikan overfitting pada tingkat moderat. Sebaliknya, XGBoost dan LightGBM menunjukkan selisih yang lebih kecil, masing-masing 0,07 dan 0,06, sehingga memiliki kemampuan generalisasi yang lebih baik. Kondisi ini dapat dijelaskan oleh mekanisme regularisasi yang lebih kuat pada algoritma gradient boosting dibandingkan pendekatan bagging yang digunakan Random Forest, sehingga model boosting lebih stabil dalam menghadapi variasi data baru.

Perbedaan kinerja ketiga model juga berkaitan dengan karakteristik algoritmiknya. Random Forest membangun pohon keputusan secara independen melalui bootstrap sampling dan menggabungkannya dengan metode averaging, sehingga memberikan stabilitas namun kurang optimal dalam menangkap pola nonlinier yang kompleks. Sebaliknya, XGBoost dan LightGBM membangun pohon secara sekuensial dengan meminimalkan error dari iterasi sebelumnya, memungkinkan penangkapan pola hubungan komposisi beton dan kuat tekan secara lebih presisi. LightGBM sedikit mengungguli XGBoost karena menggunakan strategi pertumbuhan pohon leaf-wise yang lebih agresif dan berorientasi pada gain maksimum, meskipun berpotensi meningkatkan risiko overfitting. Dengan tuning hyperparameter yang tepat, risiko tersebut dapat dikendalikan sehingga LightGBM menjadi model terbaik berdasarkan R^2 testing tertinggi (0,92) dan RMSE terendah (3,52 MPa), serta paling layak untuk aplikasi prediksi kuat tekan beton konvensional maupun berbasis agregat daur ulang.



Gambar 4 Taylor diagram

Gambar 4 menyajikan Taylor Diagram yang memberikan representasi komprehensif dari performa ketiga model (Random Forest, XGBoost, dan LightGBM) dalam satu grafik polar. Taylor Diagram menampilkan tiga metrik statistik secara simultan: (1) correlation coefficient yang ditunjukkan oleh sudut polar (azimuth angle), (2) standard deviation yang ditunjukkan oleh jarak radial dari origin, dan (3) centered root mean square error (RMSE) yang ditunjukkan oleh jarak dari titik referensi observasi. Dalam Taylor Diagram, titik referensi (observed data) terletak pada sumbu horizontal dengan nilai standard deviation observasi. Semakin dekat posisi model ke titik referensi, semakin baik performanya karena mengindikasikan: (1) correlation coefficient mendekati 1 (sudut mendekati 0°), (2) standard deviation prediksi mendekati standard deviation observasi, dan (3) centered RMSE yang minimal.

Dari Gambar 4, dapat diamati bahwa ketiga model menempati posisi yang relatif dekat dengan titik referensi, mengindikasikan performa prediktif yang baik. LightGBM dan XGBoost berada paling dekat dengan titik referensi, menunjukkan correlation coefficient yang sangat tinggi (>0.95) dan standard deviation yang hampir identik dengan observasi. Posisi kedua model yang sangat berdekatan mengkonfirmasi bahwa performa mereka hampir setara, dengan keunggulan marginal pada LightGBM. Random Forest berada sedikit lebih jauh dari titik referensi, dengan correlation coefficient yang sedikit lebih rendah dan centered RMSE yang lebih besar, konsisten dengan nilai R^2 dan RMSE yang lebih rendah pada Tabel 4.

Standard deviation prediksi yang mendekati standard deviation observasi untuk LightGBM dan XGBoost mengindikasikan bahwa kedua model tidak hanya akurat dalam mean

prediction, tetapi juga mampu menangkap variabilitas dalam data dengan sangat baik. Hal ini penting karena menunjukkan bahwa model tidak underpredicting atau overpredicting variance, yang akan menghasilkan prediction interval yang reliable untuk aplikasi praktis. Taylor Diagram dengan jelas memvisualisasikan superioritas algoritma gradient boosting (XGBoost dan LightGBM) dibandingkan dengan bagging approach (Random Forest) untuk tugas prediksi kuat tekan beton.

F. Comparison of NC vs. RAC Performance

Guna mengevaluasi apakah model machine learning mampu memprediksi kuat tekan beton normal (NC) dan beton agregat daur ulang (RAC) dengan tingkat akurasi yang sebanding, dilakukan analisis komparatif pada subset data testing berdasarkan tipe beton. Model terbaik, yaitu LightGBM, digunakan untuk menghasilkan prediksi pada masing-masing subset, kemudian kinerjanya dievaluasi menggunakan metrik R^2 score. Ringkasan hasil perbandingan tersebut disajikan pada Tabel 5.

Tabel 5 Comparison performance

Group	R^2	Samples
NC	0.92	298
RAC	0.91	135

Tabel 5 menunjukkan bahwa model LightGBM mampu memprediksi kuat tekan pada beton normal (NC) dan beton agregat daur ulang (RAC) dengan tingkat akurasi yang sangat tinggi dan hampir identik. Pada subset NC, model menghasilkan nilai R^2 sebesar 0,92 dari 298 sampel testing, sedangkan pada RAC diperoleh R^2 sebesar 0,91 dari 135 sampel. Selisih yang sangat kecil, yaitu 0,01, menegaskan bahwa kemampuan generalisasi model konsisten pada kedua tipe beton. Meskipun jumlah sampel NC lebih besar dibandingkan RAC yang mencerminkan dominasi penelitian beton normal dalam literatur ukuran sampel RAC tetap memadai untuk evaluasi statistik yang representatif, dan hasil R^2 yang hampir sama menunjukkan bahwa model tidak bias terhadap perbedaan ukuran dataset.

Temuan ini menjadi penting karena RAC umumnya memiliki karakteristik yang lebih kompleks dibandingkan NC, seperti variasi kualitas agregat daur ulang, tingkat penyerapan air yang lebih tinggi, serta densitas yang lebih rendah. Walaupun demikian, model tetap mampu menangkap hubungan antara komposisi material dan kuat tekan dengan akurasi yang sebanding. Hal ini menunjukkan bahwa variabel input yang digunakan meliputi kandungan semen, air, agregat, superplasticizer, properti RCA, dan usia curing telah cukup representatif dalam menggambarkan faktor-faktor yang memengaruhi kuat tekan pada kedua jenis beton.

Kemampuan memprediksi RAC dengan akurasi setara NC memiliki implikasi praktis yang signifikan. Pendekatan machine learning dapat dimanfaatkan sebagai alat bantu yang andal dalam perancangan mix design RAC, yang selama ini cenderung memerlukan proses trial-and-error lebih intensif. Selain itu, konsistensi performa model memperkuat dukungan terhadap pemanfaatan RAC sebagai material konstruksi yang lebih berkelanjutan. Analisis lanjutan terhadap distribusi residual pada kedua kelompok dapat memberikan validasi tambahan terkait potensi bias; apabila error terdistribusi simetris dan terpusat di sekitar nol, maka hal tersebut semakin menegaskan robustitas dan kemampuan generalisasi model yang dikembangkan dalam penelitian ini.

G. Influence of Concrete Age

Usia curing merupakan salah satu faktor terpenting yang mempengaruhi kuat tekan beton karena proses hidrasi semen yang berlangsung secara kontinu seiring waktu. Untuk mengevaluasi kemampuan model dalam memprediksi kuat tekan pada berbagai tahap perkembangan kekuatan beton, dilakukan analisis performa model pada setiap kelompok usia curing. Hasil analisis disajikan dalam Tabel 6.

Tabel 6 Performance by age

Age (hari)	R ²	MAE	Samples
1	0.81	1.58	7
3	0.25	3.9	9
4	0.83	3.14	6
7	0.88	2.45	185
28	0.91	2.36	201
56	0.75	2.71	11
90	0.74	3.56	10

Tabel 6 memperlihatkan variasi performa model pada berbagai kelompok usia curing yang cukup menarik. Kinerja terbaik dicapai pada usia 28 hari dengan R² sebesar 0,91 dan MAE 2,36 MPa dari 201 sampel, disusul usia 7 hari dengan R² 0,88 dan MAE 2,45 MPa dari 185 sampel. Tingginya akurasi pada kedua usia ini dapat dijelaskan oleh dominannya representasi data karena 7 dan 28 hari merupakan standar pengujian kuat tekan yang paling umum dalam literatur dan praktik industri. Selain itu, pada rentang usia tersebut proses hidrasi semen telah mencapai kondisi yang relatif stabil, sehingga hubungan antara komposisi material dan kuat tekan menjadi lebih konsisten. Jumlah sampel yang besar juga memberikan informasi yang memadai bagi model untuk mempelajari pola secara optimal.

Pada usia sangat awal, yaitu 1 dan 4 hari, model menunjukkan performa moderat hingga baik dengan R² masing-masing 0,81 dan 0,83, meskipun jumlah sampel sangat terbatas (7 dan

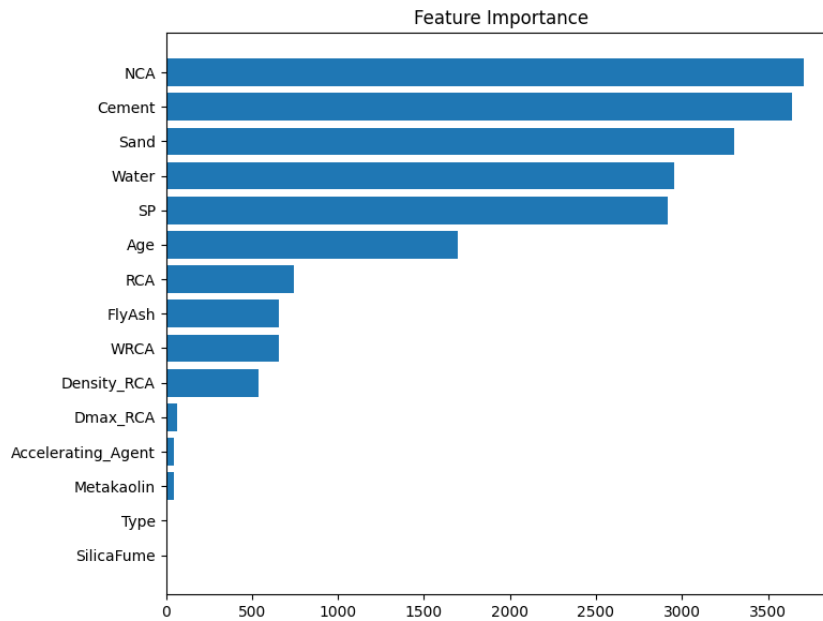
6 sampel). Nilai MAE pada usia 1 hari yang relatif rendah (1,58 MPa) menunjukkan bahwa secara absolut prediksi masih cukup akurat, walaupun variasi data menyebabkan R^2 tidak setinggi usia 28 hari. Kemampuan ini penting dalam konteks pengendalian mutu (quality control), terutama untuk evaluasi kekuatan awal beton dan pengambilan keputusan seperti waktu pembongkaran bekisting (stripping formwork).

Performa terendah terjadi pada usia 3 hari dengan R^2 hanya 0,25 dan MAE 3,9 MPa. Rendahnya akurasi ini kemungkinan besar disebabkan oleh jumlah sampel yang sangat terbatas (9 sampel), sehingga model tidak memperoleh pola yang cukup robust. Selain itu, usia 3 hari merupakan fase transisi kritis dalam proses hidrasi, di mana perkembangan kuat tekan sangat dipengaruhi oleh jenis semen, suhu curing, dan kondisi lingkungan. Heterogenitas sumber data dari berbagai literatur juga berpotensi meningkatkan variabilitas, sehingga prediksi menjadi kurang stabil pada kelompok usia ini.

Pada usia lanjut, yaitu 56 dan 90 hari, terjadi penurunan performa dengan R^2 masing-masing 0,75 dan 0,74, yang dapat dikaitkan dengan keterbatasan jumlah sampel (11 dan 10 data) serta pengaruh faktor jangka panjang seperti curing berkepanjangan, karbonasi, dan perkembangan mikrostruktur lanjutan yang mungkin belum sepenuhnya terwakili dalam fitur model. Meskipun MAE pada usia 56 hari masih relatif rendah (2,71 MPa), nilai MAE pada 90 hari meningkat menjadi 3,56 MPa, menandakan ketidakpastian prediksi yang lebih besar pada usia sangat lanjut. Secara keseluruhan, hasil analisis pengaruh usia curing menunjukkan bahwa model machine learning paling reliable untuk memprediksi kuat tekan beton pada usia standar pengujian (7 dan 28 hari) yang memiliki representasi data yang baik. Untuk usia yang sangat awal (<7 hari) atau sangat lanjut (>56 hari) dengan jumlah sampel terbatas, prediksi model perlu diinterpretasikan dengan lebih hati-hati dan sebaiknya disertai dengan interval kepercayaan yang lebih lebar. Hasil ini memberikan guidance penting untuk aplikasi praktis model: penggunaan optimal pada usia 7-28 hari untuk desain mix proporsi dan quality control, sementara untuk usia ekstrem perlu dikombinasikan dengan pengujian laboratorium atau data tambahan.

H. Feature Importance and Interpretability

Untuk memahami faktor-faktor yang paling berpengaruh terhadap prediksi kuat tekan beton, dilakukan analisis feature importance menggunakan built-in feature importance dari model LightGBM terbaik dan analisis SHAP (SHapley Additive exPlanations) untuk interpretasi yang lebih mendalam. Hasil analisis feature importance disajikan dalam Gambar 6 dan analisis SHAP disajikan dalam Gambar 7.



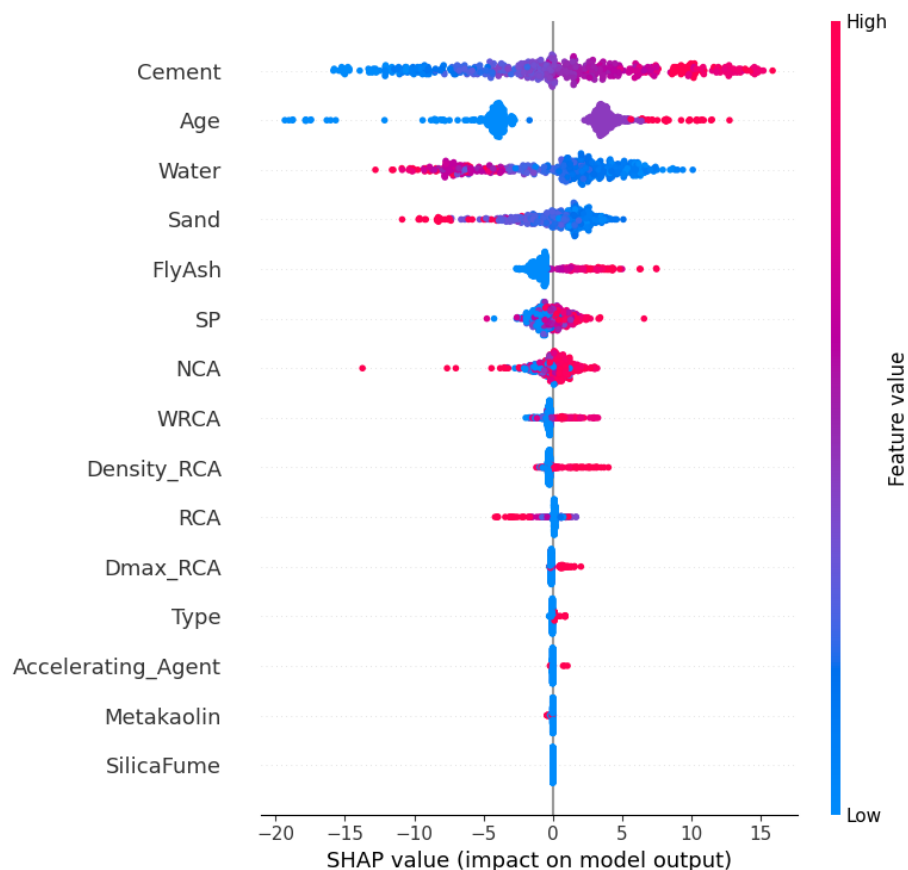
Gambar 5 Feature Importance

Gambar 5 menampilkan ranking feature importance yang menunjukkan kontribusi relatif setiap variabel input terhadap prediksi model LightGBM. Feature importance dihitung berdasarkan total gain yang dihasilkan oleh setiap fitur dalam proses splitting pada seluruh pohon keputusan dalam ensemble. Fitur dengan importance tinggi mengindikasikan bahwa fitur tersebut secara konsisten memberikan kontribusi besar dalam mengurangi error prediksi model. Berdasarkan analisis feature importance, usia curing (Age) menunjukkan importance tertinggi, mengkonfirmasi peran kritis waktu dalam pengembangan kuat tekan beton melalui proses hidrasi semen yang berlanjut seiring waktu. Hal ini konsisten dengan teori hidrasi semen yang menunjukkan bahwa kekuatan beton berkembang secara logaritmik seiring waktu, dengan peningkatan signifikan pada minggu-minggu awal dan perlambatan gradual pada usia lanjut. Importance Age yang tinggi juga mencerminkan bahwa model telah berhasil menangkap pola temporal dalam pengembangan kekuatan beton.

Kandungan semen (Cement) menunjukkan importance yang sangat tinggi, ranking kedua atau ketiga tergantung pada subset data yang dianalisis. Hal ini sangat masuk akal karena semen adalah komponen aktif utama dalam beton yang bertanggung jawab untuk pembentukan kekuatan melalui proses hidrasi. Relationship antara kandungan semen dan kuat tekan umumnya bersifat positif dan hampir linier dalam rentang tertentu, menjadikan cement sebagai predictor yang sangat kuat dan konsisten. Kandungan air (Water) juga menunjukkan importance yang signifikan, yang dapat dijelaskan oleh perannya dalam menentukan rasio air-semen (water-cement ratio), salah satu parameter terpenting dalam desain mix proporsi beton.

Rasio air-semen yang lebih rendah umumnya menghasilkan beton dengan kuat tekan lebih tinggi karena porositas yang lebih rendah dan struktur mikroskopik yang lebih padat. Model telah berhasil menangkap hubungan negatif ini antara kandungan air dan kuat tekan.

Agregat kasar alami (NCA) dan agregat kasar daur ulang (RCA) menunjukkan importance yang moderat. Untuk RCA, properti fisik RCA seperti densitas (Density_RCA), penyerapan air (WRCA), dan ukuran maksimum (Dmax_RCA) juga berkontribusi pada prediksi, mencerminkan pengaruh kualitas agregat daur ulang terhadap kinerja mekanis beton. WRCA yang tinggi dapat mengindikasikan porositas tinggi dan kualitas agregat yang lebih rendah, yang berdampak negatif pada kuat tekan. Superplasticizer (SP) menunjukkan importance yang relatif lebih rendah dibandingkan dengan komponen material utama, yang konsisten dengan perannya sebagai bahan tambah yang mempengaruhi workability lebih dari kekuatan langsung. Namun, SP dapat memiliki efek tidak langsung pada kuat tekan dengan memungkinkan pengurangan air (water reduction) tanpa mengorbankan workability, sehingga meningkatkan rasio air-semen efektif.



Gambar 6 SHAP analysis

Gambar 6 menampilkan SHAP summary plot yang memberikan visualisasi komprehensif mengenai pengaruh masing-masing fitur terhadap prediksi model pada setiap

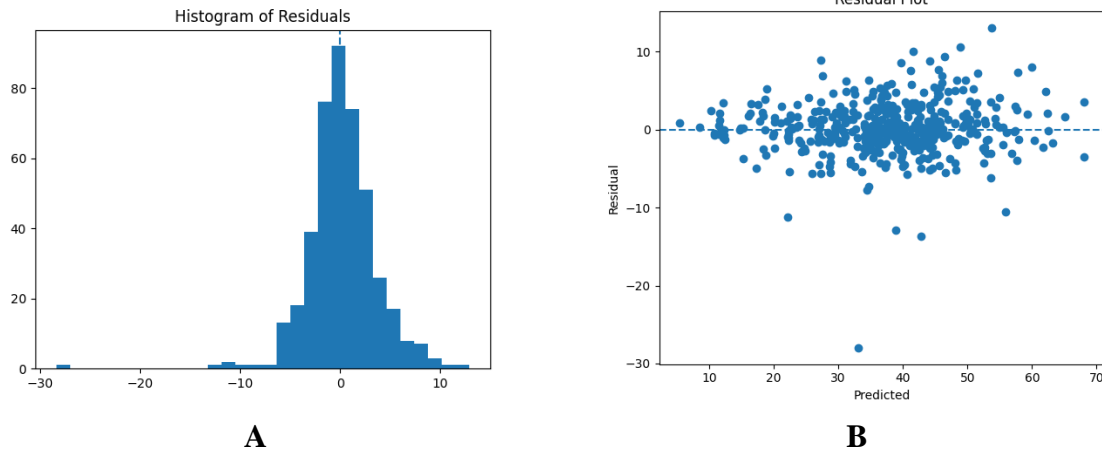
observasi individual. Berbeda dengan feature importance konvensional yang hanya menyajikan peringkat pengaruh secara global, SHAP values mampu menunjukkan besaran (magnitude) sekaligus arah (direction) kontribusi setiap fitur terhadap output model. Setiap titik pada plot merepresentasikan satu sampel, dengan posisi horizontal menunjukkan nilai SHAP (kontribusi terhadap prediksi kuat tekan), sedangkan warna menunjukkan nilai fitur (merah untuk nilai tinggi dan biru untuk nilai rendah). Pendekatan ini memungkinkan interpretasi yang lebih mendalam terhadap bagaimana model mengambil keputusan pada level individual maupun global.

Hasil analisis menunjukkan bahwa Age merupakan fitur paling dominan, di mana nilai usia curing yang tinggi secara konsisten memberikan kontribusi positif besar terhadap prediksi kuat tekan, menegaskan hubungan positif yang kuat antara umur beton dan perkembangan kekuatan. Cement juga memperlihatkan pola positif yang hampir monoton, dengan kandungan semen tinggi meningkatkan prediksi kuat tekan, meskipun efeknya dapat dipengaruhi oleh rasio air-semen. Sebaliknya, Water cenderung memberikan kontribusi negatif ketika nilainya tinggi, mengonfirmasi bahwa kelebihan air berpotensi menurunkan kuat tekan, walaupun pengaruhnya tetap kontekstual tergantung kombinasi variabel lain dalam campuran.

Untuk variabel agregat, baik NCA maupun RCA menunjukkan pola kontribusi yang lebih kompleks dan kontekstual, tercermin dari sebaran SHAP values yang cukup luas akibat interaksi antara proporsi, kualitas, dan rasio substitusi agregat. Secara khusus pada RCA, nilai WRCA yang tinggi cenderung berkontribusi negatif terhadap prediksi kuat tekan, sesuai dengan karakteristik agregat daur ulang berporositas tinggi. Kombinasi analisis feature importance dan SHAP tidak hanya meningkatkan interpretabilitas model, tetapi juga memberikan arahan praktis dalam optimasi mix design, seperti pengendalian rasio air-semen, penggunaan kandungan semen yang memadai, curing yang optimal, serta pemilihan agregat daur ulang dengan daya serap air rendah untuk mencapai kuat tekan yang maksimal.

Error Analysis

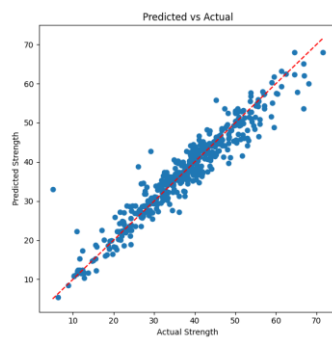
Analisis error dilakukan untuk memahami karakteristik dan distribusi kesalahan prediksi model, mengidentifikasi pola sistematis dalam error, dan mengevaluasi asumsi-asumsi yang mendasari model regresi. Visualisasi error disajikan dalam Gambar 7 dan Gambar 8.



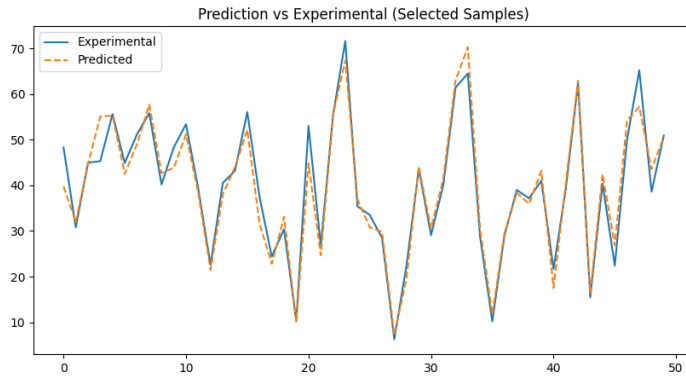
Gambar 7 A) histogram residual, (B) scatter plot residual vs predicted values

Gambar 8 menampilkan dua jenis visualisasi error: (A) histogram residual dan (B) scatter plot residual vs predicted values. Histogram residual (Gambar 7A) menunjukkan distribusi frekuensi dari residual (selisih antara nilai aktual dan nilai prediksi). Distribusi residual yang ideal untuk model regresi adalah distribusi normal (Gaussian) yang terpusat di sekitar nol, mengindikasikan bahwa error bersifat random dan tidak bias. Dari Gambar 7A, dapat diamati bahwa distribusi residual mendekati bentuk normal dengan puncak di sekitar nol dan bentuk yang relatif simetris. Hal ini mengindikasikan bahwa model tidak memiliki bias sistematis yang signifikan dalam prediksi - tidak ada kecenderungan untuk konsisten over-predict atau under-predict kuat tekan. Sebagian besar residual terkonsentrasi dalam rentang ± 5 MPa, dengan hanya sedikit outlier yang memiliki error absolut yang sangat besar. Bentuk distribusi yang mendekati normal juga mengkonfirmasi bahwa asumsi homoscedasticity (variance error yang konstan) sebagian besar terpenuhi.

Scatter plot residual vs predicted values (Gambar 7B) memberikan insight tambahan tentang pola error. Plot ini menampilkan residual pada sumbu Y dan nilai prediksi pada sumbu X. Pola yang ideal adalah random scatter di sekitar garis horizontal $y=0$, tanpa pola atau trend yang jelas. Dari Gambar 8B, dapat diamati bahwa residual tersebar secara relatif random di sekitar nol, dengan beberapa pengamatan range prediksi 20-30 MPa residual menunjukkan spread yang relatif konstan dan terdistribusi di sekitar nol, mengindikasikan performa prediksi yang baik pada range kuat tekan rendah. Range prediksi 30-50 MPa ini adalah range dengan jumlah sampel terbanyak (berdasarkan densitas titik), dan residual menunjukkan spread yang sangat konsisten. Tidak ada pola funnel atau trend yang jelas, mengkonfirmasi bahwa model sangat reliable pada range kuat tekan yang paling umum. Range prediksi >50 MPa terdapat kecenderungan slight underprediction (residual positif) pada beberapa sampel dengan kuat tekan sangat tinggi.



A



B

Gambar 8 (A) Scatter plot actual vs predicted strenght, (B)) line plot perbandingan prediksi dan aktual

Gambar 8 menyajikan dua bentuk visualisasi error tambahan, yaitu (A) scatter plot antara nilai prediksi dan nilai aktual serta (B) line plot perbandingan prediksi dan aktual pada sampel terpilih. Scatter plot pada Gambar 8A merupakan metode fundamental dalam evaluasi model regresi, dengan nilai aktual (observed) pada sumbu X dan nilai prediksi (predicted) pada sumbu Y. Garis diagonal ($y = x$) merepresentasikan kondisi prediksi sempurna. Semakin dekat titik-titik data terhadap garis tersebut, semakin tinggi tingkat akurasi model. Pada grafik terlihat bahwa sebagian besar titik terkonsentrasi sangat dekat dengan garis diagonal, mengonfirmasi nilai R^2 sebesar 0,92 yang sebelumnya dijelaskan. Pada rentang kuat tekan 20-70 MPa yang paling umum dalam praktik konstruksi deviasi titik sangat kecil, menunjukkan bahwa model sangat andal pada rentang kekuatan yang paling relevan secara praktis.

Analisis lebih rinci terhadap scatter plot menunjukkan beberapa pola penting. Pada rentang kuat tekan 10-30 MPa, titik-titik masih berada di sekitar garis diagonal tanpa pola overprediction atau underprediction yang sistematis. Pada rentang 30-50 MPa, yang memiliki kepadatan data tertinggi, titik-titik sangat rapat mengikuti garis diagonal sehingga mencerminkan akurasi optimal model. Sementara itu, pada kuat tekan >60 MPa, terlihat kecenderungan slight underprediction, di mana beberapa titik berada di bawah garis diagonal. Hal ini kemungkinan disebabkan oleh keterbatasan jumlah sampel pada rentang ekstrem tersebut dalam data pelatihan.

Gambar 8B menampilkan line plot perbandingan nilai prediksi dan aktual pada sejumlah sampel acak dari testing set. Grafik ini memperlihatkan dua garis-garis penuh untuk nilai aktual dan garis putus-putus untuk prediksi dengan sumbu X sebagai indeks sampel dan sumbu Y sebagai kuat tekan (MPa). Kedua garis menunjukkan pola yang sangat mirip, di mana puncak dan lembah muncul pada posisi yang sama. Hal ini menandakan bahwa model mampu

menangkap fluktuasi nilai kuat tekan secara konsisten. Perbedaan vertikal antar garis relatif kecil pada sebagian besar sampel, sedangkan deviasi yang lebih besar hanya muncul pada titik-titik tertentu yang berkorespondensi dengan outlier yang telah teridentifikasi sebelumnya. Penting dicatat bahwa deviasi tersebut tidak bersifat sistematis, karena model kembali mengikuti pola aktual pada sampel berikutnya.

Secara keseluruhan, analisis error melalui scatter plot dan line plot menegaskan bahwa model LightGBM memiliki performa prediktif yang sangat baik dengan error yang bersifat acak dan tidak bias. Konsistensi hasil ini, yang juga sejalan dengan analisis residual sebelumnya, menunjukkan tidak adanya pola kesalahan sistematis yang mengindikasikan masalah fundamental pada model. Dalam aplikasi praktis, model dapat diandalkan untuk memprediksi kuat tekan beton pada rentang 20-60 MPa dengan tingkat kepercayaan tinggi dan deviasi relatif kecil (sekitar $\pm 3-4$ MPa untuk sebagian besar sampel). Namun, untuk komposisi yang sangat tidak umum atau target kuat tekan >70 MPa, hasil prediksi sebaiknya divalidasi melalui pengujian laboratorium atau dilengkapi margin keamanan tambahan dalam desain struktural.

BAB V

KESIMPULAN

A. Kesimpulan dan Saran

Penelitian ini berhasil mengembangkan dan mengevaluasi model machine learning berbasis algoritma tree-based—Random Forest, XGBoost, dan LightGBM—untuk memprediksi kuat tekan beton agregat daur ulang (RAC) dan beton normal (NC). Analisis dilakukan terhadap 2165 sampel (1600 NC dan 565 RAC) yang telah melalui tahapan preprocessing secara sistematis. Secara umum, ketiga model menunjukkan performa prediktif yang sangat baik pada data pengujian. LightGBM memberikan hasil terbaik dengan nilai R^2 sebesar 0,92, MAE 2,45 MPa, dan RMSE 3,52 MPa, diikuti oleh XGBoost ($R^2 = 0,91$) dan Random Forest ($R^2 = 0,87$). Pendekatan gradient boosting terbukti lebih efektif dibandingkan metode bagging dalam menangkap hubungan nonlinier antara komposisi material dan kuat tekan beton.

Kemampuan generalisasi model juga menunjukkan konsistensi yang tinggi pada kedua jenis beton. LightGBM mampu memprediksi kuat tekan beton normal ($R^2 = 0,92$) dan beton agregat daur ulang ($R^2 = 0,91$) dengan perbedaan yang sangat kecil. Hal ini menunjukkan bahwa model memiliki robust generalization capability serta bahwa variabel input yang digunakan telah cukup representatif dalam menggambarkan faktor-faktor yang memengaruhi kuat tekan. Analisis berdasarkan usia curing memperlihatkan bahwa performa model paling optimal pada umur 7 hari ($R^2 = 0,88$) dan 28 hari ($R^2 = 0,91$), sedangkan akurasi sedikit menurun pada usia sangat awal (<7 hari) dan sangat lanjut (>56 hari) akibat keterbatasan jumlah data serta kompleksitas proses hidrasi pada tahap tersebut.

Hasil analisis feature importance dan SHAP menunjukkan bahwa usia curing (Age), kandungan semen (Cement), dan kandungan air (Water) merupakan faktor paling dominan dalam menentukan kuat tekan beton. Pada beton RAC, parameter tambahan seperti penyerapan air agregat daur ulang (WRCA) juga memberikan kontribusi signifikan. Hubungan yang teridentifikasi memperlihatkan bahwa Age dan Cement berpengaruh positif terhadap peningkatan kuat tekan, sedangkan Water dan WRCA cenderung memberikan pengaruh negatif. Analisis residual mengindikasikan bahwa kesalahan prediksi bersifat acak, tidak bias, dan mayoritas berada dalam rentang ± 5 MPa, dengan kecenderungan sedikit underprediction pada beton mutu sangat tinggi (>70 MPa).

Secara praktis, model yang dikembangkan dapat diandalkan untuk memprediksi kuat tekan pada rentang 20–60 MPa dengan tingkat kepercayaan yang baik (± 3 -4 MPa untuk sebagian besar sampel). Model ini berpotensi diintegrasikan ke dalam sistem desain mix

proporsi maupun quality control untuk mempercepat estimasi kuat tekan tanpa harus menunggu pengujian laboratorium 28 hari. Selain itu, keberhasilan model dalam memprediksi RAC dengan akurasi setara NC memberikan kontribusi penting terhadap praktik konstruksi berkelanjutan. Ke depan, pengembangan dataset yang lebih luas, penambahan variabel baru, eksplorasi pendekatan deep learning, validasi eksperimental, serta pengembangan aplikasi berbasis web atau mobile menjadi arah penelitian lanjutan.

DAFTAR PUSTKA

- [1] A. T. Balasbaneh, W. Sher, J. Li, and A. Ashour, "Systematic Review of Construction Waste Management Scenarios: Informing Life Cycle Sustainability Analysis," 2025. doi: 10.1007/s43615-024-00424-z.
- [2] S. Lee, H. Chang, and J. Lee, "Construction and demolition waste management and its impacts on the environment and human health: Moving forward sustainability enhancement," 2024. doi: 10.1016/j.scs.2024.105855.
- [3] I. Papamichael, I. Voukkali, P. Loizia, and A. A. Zorpas, "Construction and demolition waste framework of circular economy: A mini review," *Waste Management and Research*, vol. 41, no. 12, 2023, doi: 10.1177/0734242X231190804.
- [4] G. C. R. Hasibuan *et al.*, "Integrating circular economy into construction and demolition waste management: A bibliometric review of sustainable engineering practices in the built environment," *Case Studies in Chemical and Environmental Engineering*, vol. 11, 2025, doi: 10.1016/j.cscee.2025.101159.
- [5] X. Yuan *et al.*, "Machine Learning Prediction Models to Evaluate the Strength of Recycled Aggregate Concrete," *Materials*, vol. 15, no. 8, 2022, doi: 10.3390/ma15082823.
- [6] D. Wang *et al.*, "Mechanical performance of recycled aggregate concrete in green civil engineering: Review," *Case Studies in Construction Materials*, vol. 19, 2023, doi: 10.1016/j.cscm.2023.e02384.
- [7] Y. Ji, D. Wang, and J. Wang, "Study of recycled concrete properties and prediction using machine learning methods," *Journal of Building Engineering*, vol. 94, 2024, doi: 10.1016/j.jobe.2024.110067.
- [8] H. G. Gebremariam, S. Taye, and A. G. Tarekegn, "Compressive strength prediction of carbonated recycled aggregate concrete using regression based machine learning models," *Sci. Rep.*, 2026, doi: 10.1038/s41598-026-36197-8.
- [9] X. Zhang, C. Dai, W. Li, and Y. Chen, "Prediction of compressive strength of recycled aggregate concrete using machine learning and Bayesian optimization methods," *Front. Earth Sci. (Lausanne)*, vol. 11, 2023, doi: 10.3389/feart.2023.1112105.
- [10] M. Hosseinzadeh, M. Dehestani, and A. Hosseinzadeh, "Prediction of mechanical properties of recycled aggregate fly ash concrete employing machine learning algorithms," *Journal of Building Engineering*, vol. 76, 2023, doi: 10.1016/j.jobe.2023.107006.

- [11] M. S. Khan *et al.*, “Explainable AutoML models for predicting the strength of high-performance concrete using Optuna, SHAP and ensemble learning,” *Front. Mater.*, vol. 12, 2025, doi: 10.3389/fmats.2025.1542655.
- [12] A. Jaglan and R. R. Singh, “Recycled Aggregate Concrete Incorporating GGBS and Polypropylene Fibers Using RSM and Machine Learning Techniques,” *Buildings*, vol. 15, no. 1, 2025, doi: 10.3390/buildings15010066.
- [13] D. Demirtürk, Ö. Mintemur, and A. Arslan, “Optimizing LightGBM and XGBoost Algorithms for Estimating Compressive Strength in High-Performance Concrete,” *Arab. J. Sci. Eng.*, 2025, doi: 10.1007/s13369-025-10217-7.
- [14] A. Ahmad, K. A. Ostrowski, M. Maślak, F. Farooq, I. Mehmood, and A. Nafees, “Comparative study of supervised machine learning algorithms for predicting the compressive strength of concrete at high temperature,” *Materials*, vol. 14, no. 15, 2021, doi: 10.3390/ma14154222.
- [15] N. D. Hoang, “Leveraging a Hybrid Machine Learning Approach for Compressive Strength Estimation of Roller-Compacted Concrete with Recycled Aggregates,” *Mathematics*, vol. 12, no. 16, 2024, doi: 10.3390/math12162542.
- [16] Q. Wang, R. Dai, H. Zhang, H. Zheng, and X. Liang, “Machine learning-based prediction method for drying shrinkage of recycled aggregate concrete,” *Journal of Building Engineering*, vol. 96, 2024, doi: 10.1016/j.jobbe.2024.110493.
- [17] Y. Zhang, W. Ren, Y. Chen, Y. Mi, J. Lei, and L. Sun, “Predicting the compressive strength of high-performance concrete using an interpretable machine learning model,” *Sci. Rep.*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-79502-z.
- [18] N. Elshaboury, A. Al-Sakkaf, E. M. Abdelkader, and G. Alfalah, “Construction and Demolition Waste Management Research: A Science Mapping Analysis,” 2022. doi: 10.3390/ijerph19084496.
- [19] L. Huang, H. Cui, S. Li, and E. Guo, “Hybrid machine learning for elastic modulus prediction in recycled aggregate concrete,” *Multiscale and Multidisciplinary Modeling, Experiments and Design*, vol. 8, no. 7, 2024, doi: 10.1007/s41939-025-00909-w.
- [20] 明邱, “Dataset of compressive strength and slump of normal concrete,” Feb. 2024. doi: 10.17632/zrsbhndz9f.1.
- [21] Md. M. Rahman and R. Aayaz, “Recycled aggregate-metakeolin enhanced concrete Compressive Strength DataSet,” Feb. 2025. doi: 10.17632/zs42wxbyvf.1.

LAMPIRAN

JURNAL SINTA 2 <https://journal.iistr.org/index.php/JNEST/index>

The screenshot displays the 'Submissions' section of a journal's website. At the top, there is a dark blue header with the journal's name 'Journal of Novel Engineering Science and Technology' and navigation icons. Below the header, the 'Submissions' page is shown with two tabs: 'My Queue' (containing 1 item) and 'Archives'. A 'Help' icon is visible. The main content area is titled 'My Assigned' and includes a search bar, 'Filters', and 'New Submission' buttons. A single submission is listed with the ID '2092', the author 'Sujat et al.', and the title 'Integrated Prediction Model for Normal and Recycled Aggregate Concrete Strength Using Ensem...'. Action buttons for 'Submission' and 'View' are provided for this entry.