

Pemodelan Prediksi Biaya Pemeliharaan Konstruksi Berbasis *Machine Learning* Menggunakan Multiple Linear Regression dan Random Forest

Cahya Esther P.W.^{1*}, Farida Rachmawati², Yusroniya Eka Putri R.W.³

¹Postgraduate Student at Institute Teknologi Sepuluh Nopember, Surabaya, Indonesia

²Institute Teknologi Sepuluh Nopember, Surabaya, Indonesia

Corresponding Author's e-mail : cahyaesther@gmail.com

ARMADA
JURNAL PENELITIAN MULTIDISIPLIN

e-ISSN: 2964-2981

ARMADA : Jurnal Penelitian Multidisiplin

<https://ejournal.45mataram.ac.id/index.php/armada>

Vol. 04, No. 07 Juli, 2026

Page: 2979-2991

DOI:

<https://doi.org/10.55681/armada.v4i7.3135>

Article History:

Received: Mei 13, 2026

Revised: Juni 07, 2026

Accepted: Juli 19, 2026

Abstract : During the tendering phase, contractors must at least estimate the whole project's costs. However, contractors often focus primarily on the execution phase, whilst the maintenance and handover phases are not adequately accounted for. Such behaviour risks reducing profits or even resulting in losses. This study aims to develop a data-driven maintenance cost prediction model, specifically for water infrastructure projects, and to compare the performance of the Multiple Linear Regression (MLR) and Random Forest Regressor (RF) prediction models. MLR and RF are machine learning-based methods with contrasting data processing characteristics. MLR processes data linearly, whilst the RF processes data randomly. The RF and MLR models utilise raw data and logarithmic transformations. This study produced model using data from 18 projects. The best model performance on the 18 data points was achieved by RF 10 – Log – KFold, with the lowest MAPE (21.19%), MAE (5.25×10^8) and RMSE (9.2×10^8). In this study, the prediction model using the RF algorithm performed better than the MLR algorithm. However, the findings of this study cannot be generalised due to data limitations, which appear to have influenced the results.

Keywords : Construction Maintenance Costs, Multiple Linear Regression, Random Forest Regressor, Machine Learning

Abstrak : Pada tahap tender, kontraktor setidaknya harus memperkirakan seluruh biaya proyek secara akurat sejak awal sebelum pelaksanaan pekerjaan dimulai. Namun, kontraktor sering kali lebih berfokus pada tahap pelaksanaan, sedangkan tahap pemeliharaan dan serah terima belum diperhitungkan secara memadai. Perilaku tersebut berisiko mengurangi keuntungan atau bahkan menimbulkan kerugian. Penelitian ini bertujuan mengembangkan model prediksi biaya pemeliharaan berbasis data, khususnya untuk proyek infrastruktur air, serta membandingkan kinerja model prediksi Regresi Linear Berganda (MLR) dan Random Forest Regressor (RF). MLR dan RF merupakan metode berbasis pembelajaran mesin dengan karakteristik pengolahan data yang berbeda dalam memproses pola dan hubungan antarvariabel. MLR mengolah data secara linear, sedangkan RF mengolah data secara acak. Model RF dan MLR menggunakan data mentah serta transformasi logaritmik. Penelitian ini menghasilkan model dengan menggunakan data dari 18 proyek. Kinerja model terbaik pada 18 data yang tersedia dicapai oleh RF 10 – Log – KFold, dengan nilai MAPE terendah (21,19%), MAE

($5,25 \times 10^8$), dan RMSE ($9,2 \times 10^8$). Dalam penelitian ini, model prediksi menggunakan algoritma RF menunjukkan kinerja lebih baik dibandingkan algoritma MLR. Namun, temuan penelitian ini belum dapat digeneralisasikan karena keterbatasan data yang tampaknya memengaruhi hasil penelitian.

Kata Kunci : Biaya Pemeliharaan Konstruksi, Regresi Linear Berganda, Random Forest Regressor, Pembelajaran Mesin

PENDAHULUAN

Proyek Konstruksi membutuhkan pengendalian biaya produksi untuk mengelola dan menyelesaikan proyek. Biaya proyek terdiri atas biaya langsung dan biaya tidak langsung. Seluruh biaya tersebut direncanakan mulai dari masa tender hingga masa pemeliharaan sebelum hasil kerja diserahterimakan. Dalam beberapa kasus, biaya proyek yang diberi perhatian khusus dari masa tender adalah biaya pelaksanaan proyek, di sisi lain biaya pemeliharaan sebelum serah terima dikesampingkan perhitungannya. Padahal, fase pemeliharaan sebelum serah terima merupakan fase krusial. Produk konstruksi harus dapat diterima sesuai dengan spesifikasi yang ditetapkan. Pada fase pemeliharaan terdapat kemungkinan hal-hal tak terduga yang dapat mengganggu proses serah terima nantinya. Tantangan-tantangan teknis dan non teknis yang mungkin terjadi pada masa pemeliharaan sebelum serah terima harus diperhitungkan dalam struktur biaya proyek. Salah satu penyebab peningkatan biaya pada masa pemeliharaan adalah kesalahan proses produksi yang menyebabkan pengulangan pekerjaan (*rework*). Rework dapat disebabkan oleh beberapa faktor antara lain kurangnya komunikasi antar kontraktor, perubahan desain dari pemilik pekerjaan, serta kesalahan metode kerja oleh kontraktor (Yuni, 2021).

Penelitian mengenai prediksi biaya yang berkolaborasi dengan machine learning telah mulai dikembangkan dalam skala global. Puspitasari, dkk (2025) mengembangkan machine learning dengan metode Random Forest pada proyek konstruksi dengan jenis produk bangunan Gedung. Namun, penelitian ini lebih mengulas dari sisi pemrograman. Sedangkan, sisi manajemen proyek dikesampingkan. Realisasi biaya proyek juga tidak diperhitungkan dalam penelitian ini. Penelitian lain tentang prediksi biaya konstruksi dilakukan oleh Habib et al (2025) yang berkesimpulan tentang pentingnya otomatisasi dalam prediksi biaya konstruksi yang berkolaborasi dengan kecerdasan buatan bagi perusahaan konstruksi. Random Forest dan Gradien Boosting merupakan algoritma yang digunakan pada penelitian ini. Namun, penelitian ini tidak mempertimbangkan biaya pasca konstruksi atau biaya pemeliharaan. Shoar, et al (2022) melakukan penelitian terkait prediksi biaya layanan utama konstruksi yaitu desain dan pengawasan menggunakan metode Random Forest Regressor dan membandingkan hasilnya dengan metode Multiple Linier Regression dan Support Vector Regression (SVR). Penelitian Shoar, et al (2022) membuktikan bahwa Random Forest memiliki kinerja terbaik dibandingkan MLR dan SVR serta efektif digunakan untuk estimasi biaya tahap awal.

Penelitian ini bertujuan untuk membuat model prediksi biaya pemeliharaan proyek-proyek infrastruktur keairan berdasarkan variabel-variabel yang relevan menggunakan metode Multiple Linier Regression dan Random Forest Regressor, serta menganalisis dan membandingkan kinerja model prediksi Multiple Linier Regression dan Random Forest Regressor dalam memprediksi biaya pemeliharaan pada data uji. Selain itu, penelitian ini menambahkan modul validasi *K-Fold Cross Validation* dan *Leave-One-Out Cross Validation* sebagai validator model. Penelitian ini berkontribusi dalam menambah kajian terhadap rencana biaya pemeliharaan proyek infrastruktur keairan dengan batasan data yang digunakan berasal dari salah satu Perusahaan general kontraktor di Indonesia. Kedua, penelitian ini dapat memberikan rekomendasi alternatif pemodelan pencadangan biaya pemeliharaan yang harus diestimasi oleh proyek-proyek serupa di masa depan dalam lingkup Perusahaan. Secara keilmuan, penelitian ini menambah wawasan keilmuan terkait manajemen biaya pada masa pemeliharaan dan serah terima proyek.

METODE PENELITIAN

Penelitian ini menggunakan data historis proyek infrastruktur keairan pada sebuah perusahaan kontraktor di Indonesia yang telah diserahterimakan pada periode 2023–2025 dan dibatasi pada kontrak harga satuan (*unit price*). Dari 33 data mentah proyek, dilakukan pembersihan data agar sesuai dengan kriteria penelitian. Variabel dependen adalah biaya pemeliharaan proyek, sedangkan variabel independen meliputi nilai proyek, tingkat kerusakan bangunan, dan jenis produk bangunan. Nilai proyek menggambarkan skala serta kompleksitas pekerjaan, sedangkan tingkat kerusakan diperoleh dari persentase biaya perbaikan terhadap total biaya pemeliharaan yang dikonversi ke skala 1–5. Jenis produk bangunan diklasifikasikan menjadi bendungan, dermaga dan fasilitas pantai, serta bangunan dan normalisasi sungai. Nilai proyek dan tingkat kerusakan merupakan variabel numerik, sedangkan jenis produk bangunan merupakan variabel kategorikal yang diubah menjadi variabel *dummy*. Data yang telah dibersihkan digunakan untuk membangun model *Multiple Linear Regression* dan *Random Forest Regressor*, kemudian dibagi menjadi data latih dan data uji. Validasi dilakukan menggunakan *5-Fold Cross-Validation* dan LOOCV untuk mengurangi risiko *overfitting*, sedangkan kinerja model dievaluasi berdasarkan RMSE, MAE, dan MAPE. Hasil kedua model selanjutnya dibandingkan, dianalisis, dan divisualisasikan sebagai dasar penarikan kesimpulan dan penyusunan rekomendasi.

MLR akan menghasilkan model matematis sesuai dengan variabel-variabel independen yang membentuknya. Pada penelitian ini model prediksi MLR diharapkan membentuk model :

$$y = ax_1 + bx_2 + cx_3 + d \quad (1)$$

Keterangan :

- y : Prediksi biaya pemeliharaan proyek
- x_1 : Nilai Proyek
- x_2 : Durasi Pemeliharaan
- x_3 : Jenis Produk Bangunan
- a, b, c, d : Koefisien Regresi

Random Forest Regressor (RF) merupakan metode berbasis kumpulan *decision tree* yang menghasilkan prediksi akhir melalui perataan keluaran setiap pohon. Karena jumlah data terbatas, kompleksitas model dikendalikan menggunakan parameter *max_depth* dan *min_samples_leaf*. Dataset dibagi menjadi 80% data latih dan 20% data uji, kemudian data latih diambil secara acak menjadi sejumlah subsampel untuk membangun setiap pohon keputusan yang memprediksi biaya pemeliharaan (Ashraf, 2023). Stabilitas model RF dan *Multiple Linear Regression* (MLR) diuji menggunakan *5-Fold Cross-Validation*, yaitu data dibagi menjadi lima lipatan dan setiap lipatan digunakan secara bergantian sebagai data validasi, serta *Leave-One-Out Cross-Validation* (LOOCV), yaitu satu data dijadikan data uji dan seluruh data lainnya sebagai data latih pada setiap iterasi. Kinerja kedua model selanjutnya dievaluasi menggunakan *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), dan *Mean Absolute Percentage Error* (MAPE), dengan nilai kesalahan yang lebih rendah menunjukkan kemampuan prediksi yang lebih baik.

$$RMSE = \sqrt{\frac{1}{n} \sum (y_{true} - y_{pred})^2} \quad (2)$$

MAE bermanfaat untuk memahami besaran error relatif terhadap nilai aktual. MAE dihitung dengan rumus :

$$MAE = \frac{1}{n} \sum |y_{true} - y_{pred}| \quad (3)$$

MAPE mengukur rata-rata presentase kesalahan absolut antara nilai aktual dan nilai prediksi. MAPE diperoleh dengan rumus :

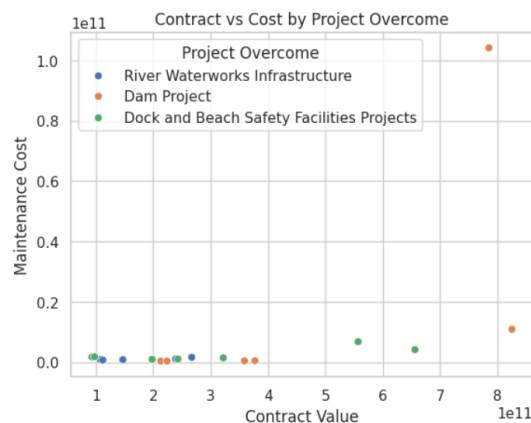
$$MAPE = \frac{1}{n} \sum \left| \frac{y_{true} - y_{pred}}{y_{true}} \right| \times 100\% \quad (4)$$

Keterangan :

y_{true} : nilai aktual
 y_{pred} : nilai prediksi
 n : jumlah data

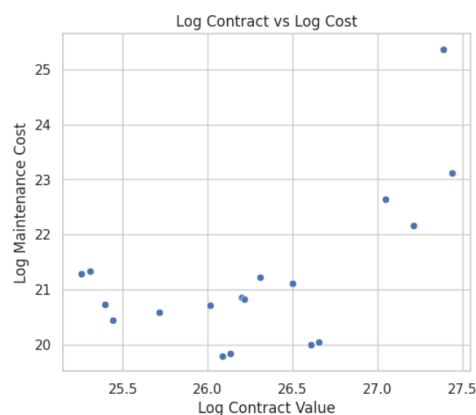
HASIL DAN PEMBAHASAN

Dalam penelitian ini terkumpul 33 data historis proyek infrastruktur keairan dari sebuah Perusahaan kontraktor di Indonesia selama tahun 2023 sampai dengan 2025. Namun hanya terdapat 18 data proyek yang sesuai dengan kriteria proyek yang ditentukan yaitu jenis kontrak *Unit Price* dan telah *Final Hand Over* (FHO) dalam rentang tahun 2023 hingga 2025, serta memiliki rincian biaya pemeliharaan yang dapat dikonversi dalam skala untuk variabel tingkat kerusakan bangunan. Data terdiri atas 5 proyek bangunan dan normalisasi sungai, 6 proyek bendungan, dan 7 proyek dermaga dan fasilitas pantai. Persebaran data divisualisasikan melalui Gambar 1.



Gambar 1. Persebaran Data Penelitian

Dari persebaran data nampak bahwa terdapat data yang menyimpang dari titik-titik lainnya. Agar data dapat diolah dengan lebih baik, dilakukan transformasi data menggunakan transformasi logaritmik untuk mempersempit sebaran data dan menangani data yang menyimpang ekstrem. Gambar 2 menunjukkan hasil transformasi logaritmik dalam persebaran data penelitian.



Gambar 2. Transformasi Log Persebaran Data Penelitian

Untuk memodelkan prediksi biaya pemeliharaan, penelitian ini menggunakan 3 variabel independen yaitu nilai kontrak, tingkat kerusakan bangunan, dan jenis bangunan. Selanjutnya, data asli dan transformasi log dibuat model prediksi menggunakan *Multiple Linier Regression* (MLR) dan *Random Forest* (RF). Terdapat 2 variabel independen numerikal yaitu nilai kontrak dan durasi pemeliharaan serta 1 variabel kategorikal yaitu jenis bangunan. Pembuatan pemodelan menggunakan *machine learning* dengan Bahasa pemrograman Python. Karena keterbatasan data, pemodelan menggunakan modul validasi tambahan yaitu *5-Fold Cross Validation* dan LOOCV.

Dengan modul validasi maka data akan mengalami lebih banyak pengujian dan latihan meskipun data yang digunakan terbatas.

Pemodelan MLR

Pada pemodelan MLR akan menghasilkan model matematis penyusun nilai prediksi yang tersusun atas koefisien dan konstanta yang mempengaruhi model. Terdapat 5 koefisien dan konstanta yang berpengaruh pada model yaitu; nilai kontrak (ax_1), tingkat kerusakan bangunan (bx_2), bangunan bendungan (cx_3), bangunan dan normalisasi sungai (dx_4), dermaga dan fasilitas pantai (ex_5). Model matematis hasil dari pemodelan sebagai berikut :

Data Asli :

$$y = 1,38 \cdot 10^{-2} x_1 - 2,89 \cdot 10^6 x_2 - 1,37 \cdot 10^9 x_3 - 4,45 \cdot 10^8 x_4 - 1,17 \cdot 10^9 x_5$$

Data Transformasi Logaritmik :

$$y = 1,13x_1 + 0,12x_2 - 9,72x_3 - 8,78x_4 - 8,43x_5$$

Nilai positif pada koefisien meningkatkan nilai prediksi sedangkan nilai koefisien negatif menurunkan hasil model prediksi. Model tersebut divalidasi menggunakan *5-Fold Cross Validation* dan LOOCV, Secara keseluruhan, masih terdapat hasil prediksi negatif pada pemodelan menggunakan data asli. Sedangkan, pada pemodelan data transformasi logaritmik menghasilkan nilai prediksi yang seluruhnya bernilai positif.

Pemodelan RF

Pemodelan RF membutuhkan pembatas dalam pemodelan agar hasil model prediksi tidak *overfitting* maupun *underfitting*. Pada pemodelan awal RF jumlah pohon keputusan ($n_estimators$) sebanyak 10 dan 100 pohon dan kedalaman tiap pohon (max_depth) sebanyak 5. Model RF menghasilkan *features importance* yang menentukan urutan kepentingan variabel pembentuk model prediksi disajikan dalam Tabel 1. Pemodelan RF baik menggunakan data mentah maupun transformasi logaritmik menghasilkan hasil prediksi bernilai positif.

Tabel 1. Features Importance Pemodelan RF

No	Feature	RF 100 Raw	RF 100 Log	RF 10 Raw	RF 10 Log
1	Nilai Kontrak	0,9277	0,7850	0,9377	0,7354
2	Skala Biaya Perbaikan	0,0143	0,0117	0,0089	0,0274
3	Jenis Hasil Produk_Bangunan Bendungan	0,0365	0,1842	0,0336	0,1955
4	Jenis Hasil Produk_Bangunan dan Normalisasi Sungai	0,0024	0,0029	0,0004	0,0032
5	Jenis Hasil Produk_Dermaga dan Fasilitas Pantai	0,0192	0,0161	0,0194	0,0385

Model MLR dan RF dievaluasi kinerjanya menggunakan MAE, RMSE, dan MAPE. Hasil dari evaluasi kinerja ditabulasikan dalam Tabel 2.

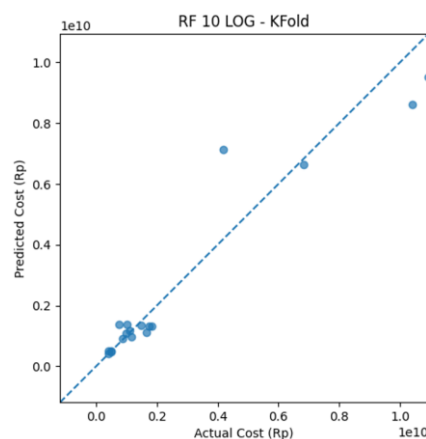
Tabel 2. Perbandingan Evaluasi Kinerja Model

No	Model	MAE	RMSE	MAPE (%)
1	MLR RAW - KFold	1.61E+09	2,03E+09	113,30
2	MLR LOG - KFold	1,94E+09	3,24E+09	77,29
3	MLR RAW - LOOCV	1,63E+09	2,09E+09	116,91
4	MLR LOG - KFold	1,93E+09	3,23E+09	82,86
5	RF 10 LOG - KFold	5.25E+08	9.2E+08	21.19
6	RF 10 LOG - LOOCV	9.37E+08	1.7E+09	27.30
7	RF 10 RAW - KFold	7.86E+08	1.19E+09	53.83
8	RF 10 RAW - LOOCV	8.4E+08	1.44E+09	33.07
9	RF LOG - KFold	1E+09	1.86E+09	29.26
10	RF LOG - LOOCV	9.96E+08	1.9E+09	24.48

11	RF RAW - KFold	8.15E+08	1.35E+09	34.02
12	RF RAW - LOOCV	7.9E+08	1.43E+09	28.42

Baik pada model MLR maupun RF, tampak bahwa model yang menggunakan data dengan transformasi logaritmik menunjukkan kinerja yang lebih baik pada data dalam penelitian ini. Model RF 10 – Log – KFold memiliki nilai MAPE terendah (21,19%); sejalan dengan itu, model ini juga memiliki nilai MAE ($5,25 \times 10^8$) dan RMSE ($9,2 \times 10^8$) terendah. Dibandingkan dengan model RF dengan 100 pohon, model RF 10 menunjukkan kinerja yang lebih baik. Hal ini mungkin disebabkan oleh jumlah data yang relatif kecil, sehingga setiap pohon belajar dari sampel *bootstrap* yang hampir identik. Akibatnya, menambahkan lebih banyak pohon tidak memperbaiki model, melainkan justru meningkatkan risiko *overfitting*. Pada model dengan 18 titik data ini, RF dengan 10 pohon sudah cukup untuk menangkap pola-pola utama model tersebut.

Di sisi lain, metode validasi LOOCV tidak selalu menghasilkan kinerja yang lebih baik untuk setiap kumpulan model. Dalam hal ini, kinerja model terbaik dicapai dengan menggunakan validasi *K-Fold*. Hal ini mungkin disebabkan oleh fakta bahwa data uji yang digunakan dalam validasi *K-Fold* lebih luas daripada yang digunakan dalam LOOCV. Dengan LOOCV, hanya satu datum yang dipilih sebagai data uji; masalah dapat muncul jika data uji yang dipilih dalam pembelajaran mesin terdiri dari nilai penyimpangan (*outliers*), yang dapat meningkatkan kesalahan pengamatan dan memengaruhi kinerja model. Namun, dengan *K-Fold Cross Validation*, kinerja model cenderung lebih stabil karena diuji pada beberapa kumpulan data secara bersamaan. Visualisasi model RF 10 – Log – KFold ditampilkan pada Gambar 3.



Gambar 3. Visualisasi Hasil Prediksi RF 10 LOG - KFold

Pembahasan

Penelitian ini menunjukkan bahwa pengembangan model prediksi biaya pemeliharaan proyek infrastruktur keairan sangat dipengaruhi oleh kualitas data, bentuk distribusi variabel, pemilihan algoritma, konfigurasi hiperparameter, dan metode validasi. Dari 33 proyek yang teridentifikasi, hanya 18 proyek memenuhi kriteria kontrak harga satuan, telah mencapai *Final Hand Over* pada periode 2023–2025, dan memiliki rincian biaya pemeliharaan yang memadai. Seleksi tersebut meningkatkan keseragaman karakteristik data yang digunakan dalam pemodelan, tetapi sekaligus memperlihatkan bahwa ketersediaan data historis proyek yang lengkap masih menjadi persoalan penting. Lee dan Yun (2024) menegaskan bahwa keberhasilan model prediksi biaya konstruksi tidak hanya ditentukan oleh algoritma, tetapi juga oleh kelengkapan, konsistensi, penanganan nilai hilang, dan pengelolaan data menyimpang sebelum proses pembelajaran dilakukan. Oleh karena itu, pembersihan data dalam penelitian ini merupakan tahapan substantif yang menentukan validitas model, bukan sekadar prosedur teknis pendahuluan.

Komposisi data yang terdiri atas lima proyek bangunan dan normalisasi sungai, enam proyek bendungan, serta tujuh proyek dermaga dan fasilitas pantai menunjukkan distribusi kategori yang relatif berimbang, meskipun jumlah observasi pada setiap kelompok masih terbatas. Keberagaman jenis bangunan penting karena karakteristik teknis, lingkungan operasional, bentuk kerusakan,

dan kebutuhan pemeliharaan setiap produk konstruksi berbeda. Yun (2022) menemukan bahwa kesalahan prediksi biaya dapat berubah menurut jenis konstruksi dan kelengkapan faktor yang digunakan dalam model. Hal ini mengindikasikan bahwa jenis bangunan layak dimasukkan sebagai prediktor, tetapi pengaruhnya baru dapat dipelajari secara lebih stabil apabila setiap kategori didukung oleh jumlah proyek yang lebih besar dan karakteristik teknis yang lebih rinci.

Pengaruh transformasi logaritmik terhadap data dan model

Persebaran data awal memperlihatkan adanya nilai ekstrem yang letaknya jauh dari sebagian besar observasi. Kondisi tersebut dapat terjadi karena nilai kontrak dan biaya pemeliharaan proyek konstruksi umumnya memiliki rentang yang lebar. Proyek berskala besar dapat memiliki nilai beberapa kali lebih tinggi daripada proyek lain sehingga mendominasi proses pembentukan model. Transformasi logaritmik yang digunakan dalam penelitian ini berhasil mempersempit rentang data dan menghasilkan persebaran yang lebih terkendali. Temuan tersebut sejalan dengan penelitian Yun (2022), yang menunjukkan bahwa deviasi besar dalam data biaya dapat menurunkan kestabilan pembelajaran dan meningkatkan kesalahan prediksi apabila data langsung digunakan tanpa prapemrosesan. Penelitian Lee dan Yun (2024) juga menegaskan bahwa perlakuan terhadap data menyimpang berpengaruh nyata terhadap kualitas model prediksi biaya konstruksi.

Manfaat transformasi logaritmik terlihat paling jelas pada pemodelan MLR. Model yang menggunakan data asli masih menghasilkan prediksi biaya bernilai negatif, sedangkan model berbasis data logaritmik menghasilkan prediksi positif. Biaya pemeliharaan secara konseptual tidak mungkin bernilai negatif sehingga hasil MLR pada data mentah menunjukkan bahwa bentuk linear pada skala asli belum mampu mewakili struktur data dengan baik. Prediksi negatif dapat muncul ketika garis regresi diekstrapolasi ke kombinasi nilai prediktor yang tidak tertangkap secara memadai dalam sampel kecil. Transformasi logaritmik membatasi ruang prediksi setelah hasil ditransformasikan kembali ke satuan awal dan mengurangi dominasi proyek bernilai sangat besar. Temuan ini memperkuat pandangan bahwa prapemrosesan harus dipilih berdasarkan karakteristik distribusi data dan bukan diterapkan sebagai prosedur otomatis.

Walaupun transformasi logaritmik memperbaiki bentuk prediksi, hasil evaluasi menunjukkan bahwa model MLR logaritmik belum mengungguli RF. Nilai MAPE MLR masih berada pada rentang 77,29–82,86%, sementara MAPE MLR berbasis data mentah mencapai 113,30–116,91%. Hasil tersebut menunjukkan bahwa hubungan antara biaya pemeliharaan dan variabel pembentuknya kemungkinan tidak sepenuhnya linear. Nilai kontrak dapat berinteraksi dengan jenis bangunan dan tingkat kerusakan, sehingga kenaikan nilai proyek tidak selalu menghasilkan perubahan biaya pemeliharaan dalam proporsi yang tetap. Penelitian Park *et al.* (2022) menunjukkan bahwa model ensemble dapat meningkatkan prediksi biaya konstruksi karena mampu menggabungkan pola yang tidak sepenuhnya ditangkap oleh satu model dasar. Turkyilmaz dan Polat (2024) juga menemukan bahwa algoritma yang paling baik dalam prediksi biaya dapat berbeda menurut struktur data, risiko proyek, dan bentuk hubungan antarvariabel.

Koefisien MLR menunjukkan bahwa nilai kontrak memiliki arah hubungan positif terhadap biaya pemeliharaan, baik pada data asli maupun transformasi logaritmik. Secara substantif, hasil tersebut dapat dipahami karena proyek dengan nilai kontrak lebih besar cenderung memiliki volume pekerjaan, jumlah komponen, tingkat kompleksitas, serta luas paparan kerusakan yang lebih tinggi. Namun, koefisien negatif pada tingkat kerusakan dan variabel kategori bangunan tidak dapat langsung dimaknai bahwa peningkatan kerusakan atau jenis bangunan tertentu selalu menurunkan biaya pemeliharaan. Arah dan besar koefisien kategori dipengaruhi oleh teknik pengodean *dummy*, kategori acuan, keberadaan konstanta, korelasi antarvariabel, serta ukuran sampel. Dengan jumlah observasi yang terbatas, koefisien dapat berubah cukup besar ketika satu atau dua proyek dikeluarkan dari data pelatihan. Oleh sebab itu, model MLR dalam penelitian ini lebih tepat diposisikan sebagai model pembandingan yang transparan daripada sebagai dasar untuk menyimpulkan hubungan kausal.

Keunggulan Random Forest dalam memprediksi biaya pemeliharaan

Model RF menghasilkan prediksi positif pada data mentah maupun data yang ditransformasikan. Kemampuan tersebut berkaitan dengan karakter RF yang membentuk prediksi dari rata-rata sejumlah pohon keputusan. Berbeda dengan MLR yang mengharuskan hubungan

linear, RF dapat membentuk pembagian data berdasarkan ambang tertentu dan menangkap interaksi serta pola nonlinier tanpa menentukan bentuk persamaan sejak awal. Keunggulan algoritma ensemble dalam estimasi biaya konstruksi juga ditemukan oleh Park *et al.* (2022), sedangkan Alazawy *et al.* (2025) memperlihatkan bahwa RF dapat digunakan untuk memprediksi biaya konstruksi sekaligus menjelaskan kontribusi variabel melalui metode interpretabilitas model.

Model RF 10-Log-KFold menghasilkan kinerja terbaik dengan MAPE 21,19%, MAE sebesar $(5,25 \times 10^8)$, dan RMSE sebesar $(9,2 \times 10^8)$. Nilai tersebut berarti bahwa, pada skema validasi yang digunakan, penyimpangan absolut prediksi secara rata-rata setara dengan 21,19% dari biaya pemeliharaan aktual. MAPE tidak tepat ditafsirkan sebagai tingkat akurasi sebesar 78,81%, tetapi sebagai besaran kesalahan persentase rata-rata. MAE menunjukkan bahwa rata-rata selisih absolut prediksi sekitar Rp525 juta, sedangkan RMSE yang lebih tinggi, sekitar Rp920 juta, mengindikasikan masih terdapat beberapa kesalahan prediksi besar yang memperoleh bobot lebih kuat karena dikuadratkan. Perbedaan MAE dan RMSE tersebut konsisten dengan keberadaan proyek yang memiliki nilai biaya jauh lebih tinggi dibandingkan proyek lainnya.

Kinerja MAPE 21,19% lebih rendah dibandingkan seluruh model lain dalam penelitian ini sehingga RF 10-Log-KFold merupakan konfigurasi relatif terbaik pada dataset yang tersedia. Hasil tersebut juga berada pada kisaran yang masih dapat dibandingkan dengan penelitian Yun (2022), yang menghasilkan kesalahan prediksi biaya konstruksi sekitar 16,80–17,67% dengan dataset jauh lebih besar. Perbandingan ini tidak dimaksudkan untuk menyatakan kedua model memiliki tingkat kemampuan yang sama karena objek, jumlah data, variabel, dan algoritmanya berbeda. Namun, perbandingan tersebut menunjukkan bahwa capaian model penelitian ini cukup menjanjikan mengingat hanya menggunakan 18 proyek dan tiga kelompok prediktor utama.

Keunggulan RF berbasis data logaritmik menunjukkan adanya efek gabungan antara algoritma nonlinier dan pengendalian rentang data. RF memang relatif tahan terhadap perbedaan skala prediktor, tetapi transformasi pada variabel biaya dapat tetap membantu ketika target memiliki distribusi sangat menceng dan terdapat proyek bernilai ekstrem. Transformasi logaritmik mengurangi jarak antara proyek terbesar dan terkecil, sehingga fungsi kesalahan tidak terlalu didominasi oleh beberapa proyek bernilai tinggi. Oleh karena itu, hasil penelitian ini menegaskan bahwa pemilihan algoritma dan prapemrosesan harus dinilai sebagai satu rangkaian, bukan sebagai keputusan yang berdiri sendiri.

Interpretasi kepentingan variabel

Hasil *feature importance* menunjukkan bahwa nilai kontrak merupakan prediktor dominan pada seluruh konfigurasi RF, dengan nilai kepentingan antara 0,7354 dan 0,9377. Dominasi ini mengindikasikan bahwa sebagian besar kemampuan RF dalam membedakan biaya pemeliharaan berasal dari variasi skala proyek. Temuan tersebut masuk akal secara operasional karena nilai kontrak merangkum berbagai unsur yang berkaitan dengan ukuran pekerjaan, kuantitas material, kompleksitas pelaksanaan, dan besarnya aset yang harus dipelihara. Alazawy *et al.* (2025) juga menemukan bahwa variabel yang merepresentasikan skala dan komponen biaya utama memiliki kontribusi besar dalam model RF untuk prediksi biaya konstruksi.

Setelah transformasi logaritmik, kepentingan nilai kontrak menurun, sedangkan kontribusi kategori bendungan meningkat menjadi 0,1842 pada RF 100-Log dan 0,1955 pada RF 10-Log. Perubahan ini menunjukkan bahwa ketika dominasi rentang nilai kontrak dikurangi, model mulai menangkap perbedaan karakter biaya berdasarkan jenis produk bangunan. Bangunan bendungan dapat mempunyai kebutuhan pemeliharaan yang berbeda karena struktur, fungsi hidraulik, komponen mekanis, kondisi fondasi, dan konsekuensi kerusakannya. Namun, hasil tersebut harus dibaca sebagai pola dalam data perusahaan yang diteliti, bukan sebagai bukti bahwa seluruh proyek bendungan secara umum selalu memiliki risiko pemeliharaan lebih tinggi.

Kontribusi skala biaya perbaikan relatif rendah, yaitu antara 0,0089 dan 0,0274. Rendahnya nilai tersebut dapat berarti bahwa variasi biaya pemeliharaan lebih banyak dijelaskan oleh nilai kontrak dan jenis produk, tetapi juga dapat disebabkan oleh penyederhanaan tingkat kerusakan ke dalam skala 1–5. Konversi tersebut memudahkan pemodelan, tetapi menghilangkan variasi rinci dalam jenis, lokasi, volume, dan tingkat keparahan kerusakan. Pada penelitian selanjutnya, tingkat kerusakan sebaiknya dibentuk dari indikator teknis independen, misalnya jumlah temuan, kategori

kerusakan struktural dan nonstruktural, luas atau volume pekerjaan perbaikan, serta hasil inspeksi lapangan.

Nilai *feature importance* juga tidak boleh ditafsirkan sebagai pengaruh kausal. Pengukuran bawaan RF dapat memberi keuntungan kepada variabel numerik kontinu yang memiliki lebih banyak kemungkinan titik pemisah dibandingkan variabel kategorikal. Thakur dan Biswas (2024) menunjukkan bahwa metode kepentingan variabel RF dapat mengandung bias dan merekomendasikan pengujian dengan *permutation importance* untuk memperoleh interpretasi yang lebih kuat. Oleh karena itu, dominasi nilai kontrak dalam penelitian ini perlu dikonfirmasi melalui *permutation importance*, SHAP, analisis sensitivitas, atau pengulangan model dengan beberapa *random seed*.

Jumlah pohon dan stabilitas model

Model dengan 10 pohon menghasilkan kesalahan lebih rendah dibandingkan model dengan 100 pohon pada konfigurasi penelitian ini. Akan tetapi, hasil tersebut tidak cukup untuk menyimpulkan bahwa penambahan pohon secara langsung menyebabkan *overfitting*. Pada RF, penambahan pohon umumnya ditujukan untuk menstabilkan hasil perataan, sedangkan kompleksitas setiap pohon lebih banyak ditentukan oleh *max depth*, jumlah minimum sampel pada daun, jumlah fitur yang dipertimbangkan pada setiap pemisahan, dan ukuran sampel *bootstrap*. Lange et al. (2025) menunjukkan bahwa hubungan antara jumlah pohon dan stabilitas RF bersifat nonlinier: penambahan pohon biasanya meningkatkan stabilitas hingga mencapai titik ketika peningkatan berikutnya menjadi sangat kecil.

Dengan hanya 18 observasi, perbedaan performa antara 10 dan 100 pohon kemungkinan mencerminkan sensitivitas terhadap komposisi lipatan, sampel *bootstrap*, dan pengacakan model. Model 10 pohon dapat secara kebetulan membentuk kombinasi yang lebih sesuai dengan lipatan pengujian tertentu, tetapi belum tentu lebih stabil apabila proses diulang menggunakan *random seed* yang berbeda. Oleh sebab itu, konfigurasi 10 pohon dapat dinyatakan sebagai konfigurasi dengan kinerja terbaik dalam eksperimen ini, tetapi belum sebagai jumlah pohon optimal secara universal. Pengembangan berikutnya perlu menguji rentang *n_estimators* yang lebih luas dan melaporkan rata-rata serta simpangan baku hasil dari beberapa pengulangan.

Perbandingan K-Fold dan LOOCV

Kinerja K-Fold yang lebih baik daripada LOOCV pada model terbaik menunjukkan bahwa penggunaan lima kelompok data menghasilkan evaluasi yang lebih sesuai dengan struktur dataset penelitian ini. Dalam LOOCV, setiap iterasi hanya menguji satu proyek. Apabila proyek tersebut memiliki nilai kontrak atau biaya pemeliharaan yang ekstrem, kesalahan satu prediksi dapat memberikan pengaruh besar terhadap nilai evaluasi. Sebaliknya, setiap lipatan pada 5-Fold memuat beberapa proyek sehingga evaluasi dilakukan terhadap kelompok data yang lebih beragam.

Austin et al. (2025) menunjukkan bahwa LOOCV dapat menghadapi bias distribusional karena setiap model dilatih pada data yang hampir sama, sedangkan label observasi yang dikeluarkan dapat memiliki hubungan negatif dengan rata-rata data pelatihannya. Kondisi tersebut dapat mengganggu evaluasi model dan pemilihan hiperparameter, khususnya pada data terbatas. Temuan ini mendukung hasil penelitian bahwa LOOCV tidak otomatis memberikan estimasi kinerja yang lebih baik hanya karena menggunakan hampir seluruh data untuk pelatihan.

Meskipun demikian, hasil K-Fold juga perlu ditafsirkan hati-hati karena setiap lipatan hanya terdiri atas sekitar tiga sampai empat proyek. Komposisi jenis bangunan dan proyek bernilai ekstrem pada setiap lipatan dapat mengubah nilai MAE, RMSE, dan MAPE secara nyata. Untuk memperkuat reliabilitas, penelitian berikutnya dapat menggunakan *repeated K-Fold Cross-Validation*, pembagian lipatan yang mempertahankan proporsi jenis proyek, dan *nested cross-validation* apabila dilakukan penyetelan hiperparameter. Langkah tersebut akan memisahkan proses pemilihan konfigurasi dari proses penilaian akhir sehingga risiko estimasi yang terlalu optimistik dapat dikurangi.

Secara teoretis, penelitian ini menunjukkan bahwa hubungan antara nilai kontrak, tingkat kerusakan, jenis bangunan, dan biaya pemeliharaan proyek keairan cenderung lebih tepat dipelajari menggunakan model yang mampu menangkap pola nonlinier. Kinerja RF yang lebih baik

daripada MLR memperlihatkan bahwa peningkatan biaya pemeliharaan tidak selalu mengikuti perubahan prediktor secara proporsional. Penelitian ini juga memperkuat pentingnya prapemrosesan data, karena transformasi logaritmik tidak hanya memperbaiki persebaran data, tetapi juga mencegah prediksi biaya negatif dan meningkatkan kinerja sebagian besar konfigurasi model.

Secara praktis, model RF 10-Log-KFold dapat dikembangkan sebagai alat bantu awal bagi kontraktor untuk memperkirakan cadangan biaya pemeliharaan sejak tahap tender. Model memungkinkan perusahaan memanfaatkan pengalaman proyek terdahulu secara lebih sistematis daripada hanya mengandalkan pertimbangan subjektif estimator. Namun, hasil prediksi tidak seharusnya langsung digunakan sebagai nilai anggaran final. Model perlu dipadukan dengan penilaian insinyur, karakteristik kontrak, kondisi lokasi, risiko hidrologi dan kelautan, mutu pekerjaan, masa pemeliharaan, serta ketentuan pemberi kerja. Dengan demikian, model berfungsi sebagai sistem pendukung keputusan, bukan pengganti penilaian profesional.

Pada tingkat organisasi, hasil penelitian menegaskan perlunya basis data proyek yang terstandar. Perusahaan perlu mencatat nilai kontrak, biaya pemeliharaan, jenis dan volume kerusakan, tanggal kejadian, penyebab kerusakan, lokasi proyek, durasi pemeliharaan, metode konstruksi, kondisi lingkungan, dan tindakan perbaikan dengan format yang konsisten. Basis data yang lebih lengkap akan meningkatkan kemampuan model sekaligus mendukung pembelajaran organisasi dari proyek yang telah selesai.

Penelitian ini menggunakan 18 proyek dari satu perusahaan kontraktor dan terbatas pada proyek infrastruktur keairan dengan kontrak harga satuan yang telah mencapai FHO. Batasan tersebut memperkuat keseragaman konteks analisis, tetapi hasil model belum dapat digeneralisasikan pada kontrak *lump sum*, proyek yang belum FHO, jenis infrastruktur lain, atau perusahaan dengan sistem pengendalian mutu yang berbeda. Penelitian lanjutan perlu melibatkan beberapa perusahaan, rentang tahun yang lebih panjang, dan jumlah proyek yang lebih besar agar kestabilan serta daya generalisasi model dapat diuji.

Jumlah prediktor yang digunakan masih terbatas. Nilai kontrak, tingkat kerusakan, dan jenis bangunan belum sepenuhnya menangkap faktor lain yang dapat memengaruhi biaya pemeliharaan, seperti durasi masa pemeliharaan, lokasi, kondisi geografis, kompleksitas desain, kualitas pekerjaan, metode pelaksanaan, cuaca, perubahan pekerjaan, kapasitas kontraktor, dan respons terhadap kerusakan. Penambahan variabel perlu dilakukan secara selektif agar informasi meningkat tanpa menghasilkan model yang terlalu kompleks terhadap jumlah data.

Skala tingkat kerusakan dihitung dari persentase biaya perbaikan terhadap total biaya pemeliharaan. Karena total biaya pemeliharaan juga merupakan variabel keluaran, pembentukan prediktor tersebut berpotensi mengandung informasi yang berasal dari komponen target. Walaupun nilai kepentingannya dalam RF relatif rendah, penelitian berikutnya sebaiknya mengukur tingkat kerusakan secara independen melalui hasil inspeksi teknis sebelum biaya pemeliharaan diketahui. Langkah ini akan memperkuat validitas prediktif dan memastikan bahwa seluruh variabel masukan benar-benar tersedia pada saat model digunakan dalam tahap tender.

Model belum diuji menggunakan dataset eksternal yang sepenuhnya terpisah. Penggunaan K-Fold dan LOOCV merupakan pilihan yang relevan untuk memanfaatkan dataset kecil, tetapi belum menggantikan validasi pada proyek baru. Oleh karena itu, model terbaik dalam penelitian ini harus dipandang sebagai model awal yang menjanjikan. Validasi prospektif terhadap proyek berikutnya, pelaporan interval prediksi, pengujian beberapa *random seed*, dan analisis kestabilan *feature importance* diperlukan sebelum model diterapkan sebagai bagian rutin dari proses estimasi biaya perusahaan.

KESIMPULAN DAN SARAN

Berdasarkan analisis pemodelan biaya pemeliharaan proyek, terdapat beberapa kesimpulan antara lain : 1) Hasil terbaik untuk Model Ensemble pada 18 data dicapai oleh RF 10 – Log – KFold, dengan nilai MAPE (21,19%), MAE ($5,25 \times 10^8$), dan RMSE ($9,2 \times 10^8$) terendah; 2) Pada 18 data tersebut, *K-Fold Cross Validation* lebih stabil karena setiap pengujian melibatkan beberapa titik

data sekaligus, sebaliknya, LOOCV, yang hanya menggunakan satu titik data uji, memiliki risiko lebih tinggi untuk memasukkan nilai-nilai ekstrem; 3) Dalam penelitian ini, model RF menunjukkan kinerja yang lebih baik daripada model MLR; 4) Hasil pemodelan bergantung pada data yang digunakan, sehingga tidak dapat dibuat generalisasi apa pun.

Penelitian ini memiliki keterbatasan yang disebabkan oleh terbatasnya data yang tersedia; oleh karena itu, untuk penelitian lebih lanjut, kami mengusulkan: 1) Disarankan agar penelitian selanjutnya memanfaatkan kumpulan data proyek yang lebih besar sehingga model dapat secara komprehensif menangkap variasi karakteristik biaya pemeliharaan dan meningkatkan kemampuan generalisasi model; 2) Disarankan untuk tidak mengabaikan data *outlier* dalam analisis; 3) Penelitian selanjutnya dapat menambahkan variabel lain yang lebih berkorelasi dengan biaya pemeliharaan proyek.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Tuhan Yang Maha Kuasa dan pihak yang telah memberi dukungan baik moril maupun materil terhadap pelaksanaan kegiatan ini.

DAFTAR PUSTAKA

- Abd, A.M., Jasim, N.A., dan Naseef, F.S. (2019), "Developing a Prediction Model to Predict The Construction Project Cost by Using Multiple Linier Regression Technique", *Indian Journal of Science Technology*, Vol.12, No.7, hal. 1-6.
- Alazawy, S.F.M., Ahmed, M.A., Raheem, S.H., Imran, H., Bernardo, L.F.A., dan Pinto, H.A.S. (2025), "Explainable Machine Learning to Predict the Construction Cost of Power Plant Based on Random Forest and Shapley Method", *CivilEng*, Vol.6, No.21, hal. 1-18.
- Alkhuadhan, A.N.A. dan Naimi, S. (2023), "A New Model for Cost Estimation Project Using Hybrid Importance Regression Ensemble Method", *Acta Logistica – International Scientific Journal about Logistics*, Vol.10, No.2, hal. 191-198.
- Ashraf, A., Rady, M., dan Mahfouz, S.Y. (2024), "Price Prediction of Residential Buildings Using Random Forest and Artificial Neural Network", *HRBC Journal*, Vol.20, No.1, hal. 23-41.
- Aung, T., Liana, S.R., Htet, A., dan Bhaumik, A. (2023), "Using Machine Learning to Predict Cost Overrun in Construction Projects", *Journal of Technology Innovation and Energy*, Vol.2, No.2, hal. 1-7.
- Beste, T. (2020), "Effect of Systematic Completion on Public Construction Projects", *Public Construction Projects*, Vol.39, No.3/4, hal. 156-171.
- Breiman, L. (2001), "Random Forest", *Machine Learning*, Vol. 45, hal. 5-32.
- Chakraborty, D., Elhegazy, H., Elzarka, H., Gutierrez, L. (2020), "A Novel Construction Cost Prediction Model Using Hybrid Natural and Light Gradient Boosting", *Advanced Engineering Informatics*, Vol. 46.
- Chen, L., Tiang, S.S., Chong, K.S., Sharma, A., Berghout, T., dan Lim, W.H. (2025), "Predicting Construction Costs with Machine Learning: A Comparative Study on Ensemble and Linear Models", *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, Vol.7, No.3, hal. 624-645.
- Faqih, F. dan Zayed, T. (2021), "Defect-Based Building Condition Assessment", *Building and Environment*, Vol.191, No.107575, hal. 1-16.
- Farhanuddin, Sihombing, S.E.K., Yahfizham, (2024), "Komparasi Multiple Linier Regression dan Random Forest Regression dalam Memprediksi Anggaran Biaya Manajemen Proyek Sistem Informasi", *Journal of Computers and Digital Business*, Vol.3, No.2, hal. 86-97.
- Habib, O., Abouhamad, M., Bayoumi A.M., (2025), "Ensemble Learning Framework for Forecasting Construction Costs", *Automation in Construction*, Vol 170.
- Hashemi, S.T., Ebadati, O.M., Kaur, H. (2020), "Cost Estimation and Prediction in Construction Projects: A Systematic Review on Machine Learning Techniques", *SN Applied Sciences*, Vol.2, No.1703.

- Hauashdh,A., Nagapan, S., Jailani, J., dan Alzaeemi, S. (2024), "Predictive Model for Corrective Maintenance Costs : Empowering Decision-Making in Building Renovation", IOP Conf. Series : Earth and Environmental Science, Vol.1347, No.1.
- Huang C.H., dan Hsieh, S.H., (2020), "Predicting BIM Labor Cost With Random Forest and Simple Linier Regression", Automation in Construction, Vol. 118
- Jung,S., Pyeon, J.H., Lee, H.S., Park, M., Yoon, I., dan Rho, J. (2020), "Construction Cost Estimation Using a Case-Based Reasoning Hybrid Genetic Algorithm Based on Local Search Method", Suistanability, Vol.12, No.7920, hal. 1-17.
- Mahamid, I., (2011), "Early Cost Estimating for Road Construction Projects Using Multiple Regression Techniques", Australasian Journal of Construction Economics Building, Vol. 11, No. 4, hal 87-101.
- M. Firmansyah, Masrun, dan I Dewa K.Y.S. (2021), "Esensi Perbedaan Metode Kualitatif dan Kuantitatif", Elastisitas Jurnal Ekonomi Pembangunan, Vol.3, No.2, hal. 156-159.
- Meharie, M.G., dan Shai, N. (2020), "Predicting Highway Construction Costs : Comparison of The Performance of Random Forest, Neural Network and Support Vector Machine Models", Journal of Soft Computing in Civil Engineering, Vol.4, No.2, hal. 103-112.
- Nejad, A.A.F., Arabikhan, F., dan Ahelerooff, S. (2024), "Optimizing Project Time and Cost Prediction Using a Hybrid XGBoost and Simulated Annealing Algorithm", Machines, Vol.12, No.867, hal. 1-25.
- Ni Kadek S.E.Y. (2021), "Estimasi Besarnya Biaya Proyek Akibat Terjadinya Rework pada Pekerjaan Finishing", Jurnal Rekayasa Konstruksi Mekanika Sipil (JRKMS), Vol.4, No.1, hal. 55-65.
- Ottaviani, F.M., dan Marco, A.D., (2021), "Multiple Linier Regression Model for Improved Project Cost Forecasting", Procedia Computer Science, Vol 196, hal. 808-815.
- Penacerrada, P. (2021), Strategies to Improve The Handover of Building Adaptation Projects, Master's Thesis, Norwegian University of Science and Technology, Norwegia.
- Project Management Institute (2017),Project Management Book of Knowledge 6th Edition, Project Management Institute, Inc., United Stated of America.
- Project Management Institute (2021),Project Management Book of Knowledge 7th Edition, Project Management Institute, Inc., United Stated of America.
- Puspitasari, D., Nada, N.Q., Harjanta, A.T.J., (2025), "Penerapan Algoritma Random Forest untuk Prediksi Biaya Konstruksi Berbasis Web", Jurnal Ilmiah ILKOMINFO – Ilmu Komputer & Informatika, Vol.8, No.2, hal. 314-326.
- Schultz, C. S., Jorgensen, K., Bonke, S. dan Rasmussen, G.M.G. (2014), "Building Defect in Danish Construction: Project Characteristics Influencing the Occurrence of Defect at Handover", Architectural Engineering and Design Management, doi: 10.1080/17452007.2014.990352kamu, hal. 1-17.
- Shirkavand, I., Lohne, J., dan Laedre, O. (2016), "Defect Handover in Norwegian Construction Projects", Procedia – Social and Behavioral Sciences, Vol.226, No.3-11, hal. 1-10.
- Shoar, S., Chileshe, N., Edwards, J.D., (2022), "Machine Learning-Aided Engineering Services' Cost Overruns Prediction in High-Rise Residential Building Projects: Application of Random Forest Regression", Journal of Building Engineering, Vol. 50.
- Taiwo, R., Hussein, M., dan Zayed, T. (2022), "An Integrated Approach of Simulation and Regression Analysis for Assesing Productivity in Modular Integrated Construction Projects", Buildings, Vol.12, No.11, hal. 1-21.
- Terry, G.R. (1958), "Principles of Management Irwin Series in Industrial Engineering and Management", Columbia : University of Columbia.
- Yamusa, M.A., Lawal, H.S., Abdulrahman, R.S., Salisu, A.S., Saka, A., Abubakar, M., dan Abdullahi, M. (2025), "A Machine-Learning Model for Estimating Construction Renovation Costs", International Journal of Building Pathology and Adaption, doi: 10.1108/IJBPA-10-2024-0213, hal. 1-29.
- Yun, S.H. (2022), "Predictive Performance of building Construction Estimation : An Analysis Based on ANN Model", Journall of System and Management Sciences, Vol.12, No.2, hal. 325-335.

- Zima, K. dan Biel, S. (2019), "The Concept of Method of Detecting and Analyzing Construction Defects in Residential Buildings", *Archives of Civil Engineering*, Vol.65, No.4, hal. 81-95.
- Zhang, J., Yuan, J., Mahmoudi, A., Ji, W., dan Fang, Q. (2024), "A Data-Driven Framework for Conceptual Cost Estimation of Infrastructure Projects using XGBoost and Bayesian Optimization", *Journal of Asian Architecture and Building Engineering*, doi: 10.1080/13467581.2023.2294871, hal. 1-24.