

Terbit *online* pada laman web jurnal :
<https://ejournal.sttp-yds.ac.id/index.php/js/index>

SAINSTEK

| ISSN (Print) 2337-6910 | ISSN (Online) 2460-1039 |



Penerapan Algoritma Random Forest Untuk Prediksi Keterlambatan Proyek Konstruksi Berdasarkan Faktor Cuaca, Tenaga Kerja, dan Material

Fitra Yuda^a, Irzon Meiditra^a, Alifia Restu Selvanda^a, Fauziah^a, Yulia^a

^aProgram Studi Sistem Informasi & Teknologi Informasi, Institut Teknologi Rokan Hilir, Rokan Hilir, Riau, Indonesia.

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 02 Juni 2026

Revisi Akhir: 20 Juni 2026

Diterbitkan Online: 29 Juni 2026

KATA KUNCI

Keterlambatan Proyek;

Random Forest;

Machine Learning;

Cuaca;

Tenaga Kerja;

Material;

Sistem Prediksi

KORESPONDENSI

Telepon: -

E-mail: fitrayuda110199@gmail.com

ABSTRACT

Keterlambatan proyek konstruksi merupakan masalah yang sering terjadi karena pengaruh cuaca, tenaga kerja, material, dan ketidaksesuaian progres lapangan. Penelitian ini bertujuan menerapkan algoritma Random Forest untuk memprediksi potensi keterlambatan proyek konstruksi berdasarkan variabel cuaca, tenaga kerja, material, dan progres pekerjaan. Metode penelitian menggunakan pendekatan kuantitatif berbasis supervised machine learning dengan tahapan pengumpulan data, preprocessing, pembentukan label, pelatihan model, pengujian model, dan implementasi aplikasi berbasis web menggunakan PHP 7.4 serta MySQLi. Dataset prototipe terdiri atas 90 data training dan 18 data testing dengan variabel curah hujan, hari hujan, suhu, kelembapan, jumlah pekerja, absensi, produktivitas, keterlambatan material, stok material, target progres, realisasi progres, deviasi progres, durasi rencana, dan minggu pekerjaan. Hasil evaluasi menunjukkan accuracy 77,78%, precision 86,67%, recall 86,67%, F1-score 86,67%, dan ROC-AUC 66,67%. Faktor yang paling dominan adalah deviasi progres, keterlambatan material, produktivitas tenaga kerja, stok material, dan jumlah hari hujan. Aplikasi yang dibangun mampu menampilkan form prediksi, hasil klasifikasi, riwayat prediksi, serta laporan formal. Hasil penelitian menunjukkan bahwa Random Forest dapat digunakan sebagai alat bantu deteksi dini risiko keterlambatan proyek konstruksi, namun model tetap perlu divalidasi menggunakan data proyek riil yang lebih besar.

1. PENDAHULUAN

Keterlambatan proyek konstruksi menjadi salah satu isu penting dalam manajemen proyek karena berdampak langsung pada biaya, mutu, hubungan kontraktual, dan tingkat kepercayaan pemangku kepentingan. Dalam praktik lapangan, keterlambatan tidak hanya disebabkan oleh satu faktor, tetapi muncul dari gabungan kondisi yang saling berkaitan, seperti cuaca buruk, keterbatasan tenaga kerja, rendahnya produktivitas, keterlambatan material, perubahan progres pekerjaan, serta kurangnya kemampuan manajemen proyek dalam membaca risiko sejak awal. Situasi ini membuat proses pengendalian proyek membutuhkan pendekatan yang lebih berbasis data, bukan hanya bergantung pada pengalaman subjektif pelaksana.

Perkembangan kecerdasan buatan dan machine learning memberi peluang untuk membangun sistem prediksi yang dapat mendukung pengambilan keputusan

proyek. Pada bidang konstruksi, machine learning banyak digunakan untuk membaca pola historis proyek, mempelajari hubungan antarvariabel, serta menghasilkan prediksi terhadap potensi keterlambatan, pembengkakan biaya, produktivitas, dan risiko pelaksanaan. Datta et al. menjelaskan bahwa kecerdasan buatan dan machine learning semakin banyak digunakan dalam siklus hidup proyek konstruksi, mulai dari perencanaan, desain, pelaksanaan, hingga pemeliharaan [1].

Penelitian mengenai prediksi keterlambatan proyek menggunakan machine learning juga terus berkembang. Alsulamy membandingkan CatBoost, XGBoost, dan LGBM untuk memprediksi risiko keterlambatan proyek di Arab Saudi dan menunjukkan bahwa pendekatan berbasis data mampu memberikan hasil prediktif yang berguna bagi manajemen risiko proyek [2]. Egwim et al. juga membuktikan bahwa algoritma ensemble learning, termasuk Random Forest, memiliki kemampuan yang baik untuk memprediksi keterlambatan proyek konstruksi

karena mampu mengolah banyak variabel dan hubungan yang tidak selalu linear [3].

Berdasarkan latar belakang tersebut, penelitian ini berfokus pada penerapan algoritma Random Forest untuk memprediksi keterlambatan proyek konstruksi berdasarkan faktor cuaca, tenaga kerja, dan material. Fokus ini dipilih karena ketiga faktor tersebut merupakan variabel yang mudah ditemukan pada laporan proyek harian atau mingguan, sehingga berpotensi diterapkan pada sistem informasi proyek yang sederhana. Penelitian ini juga menghasilkan prototipe aplikasi berbasis web menggunakan PHP 7.4 dan MySQLi agar model prediksi tidak hanya berhenti pada tahap perhitungan, tetapi dapat digunakan melalui antarmuka sistem yang mudah dipahami oleh pengguna.

2. TINJAUAN PUSTAKA

Keterlambatan Proyek Konstruksi

Keterlambatan proyek konstruksi dapat dipahami sebagai kondisi ketika realisasi pekerjaan tidak sesuai dengan jadwal rencana. Dalam proyek konstruksi, keterlambatan sering muncul karena adanya ketidakpastian teknis dan nonteknis, seperti keterlambatan persetujuan, gangguan cuaca, kekurangan tenaga kerja, keterlambatan pengiriman material, serta rendahnya koordinasi di lapangan. Gondia et al. menjelaskan bahwa prediksi risiko keterlambatan dapat ditingkatkan melalui pemanfaatan algoritma machine learning karena data proyek mengandung pola yang dapat dipelajari secara komputasional [4].

Machine Learning untuk Prediksi Proyek

Machine learning merupakan pendekatan komputasi yang memungkinkan sistem mempelajari pola dari data dan menggunakan pola tersebut untuk membuat prediksi terhadap data baru. Dalam konteks proyek konstruksi, model machine learning dapat mempelajari hubungan antara variabel input, seperti curah hujan, jumlah pekerja, produktivitas, dan keterlambatan material, dengan label output berupa status terlambat atau tidak terlambat. Yaseen et al. menggunakan pendekatan hybrid Random Forest dan Genetic Algorithm untuk memprediksi risiko keterlambatan proyek dan menunjukkan bahwa model AI mampu menangani kompleksitas proyek konstruksi [5].

Random Forest

Random Forest merupakan algoritma ensemble learning yang membangun banyak decision tree dan menggabungkan hasil prediksi melalui mekanisme voting. Keunggulan Random Forest terletak pada kemampuannya mengurangi risiko overfitting, bekerja pada data numerik maupun kategorikal, serta menghasilkan informasi feature importance yang membantu peneliti memahami variabel dominan. Secara sederhana, setiap pohon dalam Random Forest belajar dari subset data dan subset fitur yang berbeda, kemudian hasil akhir ditentukan berdasarkan suara mayoritas. Pendekatan ini cocok digunakan pada data proyek yang

memiliki banyak variabel dan pola hubungan yang tidak sederhana.

Sistem Prediksi Berbasis Web

Sistem prediksi berbasis web menjadi bagian penting dalam penelitian ini karena hasil model perlu disajikan dalam bentuk aplikasi yang dapat diakses pengguna. Penelitian terkait sistem informasi oleh Yuda dan rekan menunjukkan bahwa pengembangan aplikasi berbasis web dapat membantu proses pengelolaan data dan layanan secara lebih terstruktur [14]. Selain itu, penelitian Fauziah et al. tentang pemanfaatan AI dalam pendidikan menunjukkan bahwa pemahaman terhadap teknologi AI perlu diperkuat melalui penerapan praktis yang mudah digunakan [15]. Dengan demikian, pengembangan aplikasi prediksi berbasis web menjadi relevan untuk mempertemukan pendekatan machine learning dengan kebutuhan operasional pengguna.

3. METODOLOGI

Penelitian ini menggunakan metode kuantitatif dengan pendekatan supervised machine learning. Tahapan penelitian meliputi identifikasi variabel, penyusunan dataset prototipe, preprocessing data, pembagian data training dan testing, pelatihan model Random Forest, evaluasi model, serta implementasi aplikasi berbasis web. Pendekatan ini dipilih karena tujuan utama penelitian adalah menghasilkan model klasifikasi yang dapat membedakan proyek berisiko terlambat dan tidak terlambat berdasarkan data input yang diberikan.

Data dan Variabel Penelitian

Data yang digunakan pada penelitian ini berupa dataset prototipe yang disusun berdasarkan struktur kebutuhan aplikasi prediksi. Dataset terdiri atas variabel cuaca, tenaga kerja, material, dan progres proyek. Variabel target diklasifikasikan menjadi dua kelas, yaitu 0 untuk tidak terlambat dan 1 untuk berisiko terlambat. Struktur data ini dapat dikembangkan lebih lanjut menggunakan laporan harian proyek, laporan mingguan, kurva-S, data kehadiran pekerja, catatan material masuk, dan data cuaca dari sumber resmi.

Tabel 1. Variabel Input dan Target Model

Kelompok	Variabel Utama
Cuaca	Curah hujan, hari hujan, suhu, kelembapan
Tenaga kerja	Jumlah pekerja, pekerja absen, produktivitas
Material	Keterlambatan material, stok material
Progres dan target	Target, realisasi, deviasi, durasi, minggu ke-, status terlambat

Tahapan Algoritma Random Forest

Tahapan algoritma dimulai dari pembersihan data dan penyamaan tipe data numerik. Setelah itu, dataset dibagi menjadi data latih dan data uji. Model Random Forest dibangun dengan membentuk sejumlah decision tree, kemudian setiap tree memberikan hasil klasifikasi. Hasil akhir diperoleh melalui voting mayoritas. Pada tahap evaluasi, model diuji menggunakan accuracy, precision, recall, F1-score, ROC-AUC, dan confusion matrix. Nilai

feature importance digunakan untuk melihat variabel yang paling berpengaruh terhadap hasil prediksi.

Tabel 2. Skema Evaluasi Model

Metrik	Fungsi Evaluasi
Accuracy	Mengukur proporsi prediksi benar terhadap seluruh data uji
Precision	Mengukur ketepatan model saat memprediksi kelas terlambat
Recall	Mengukur kemampuan model menemukan kasus terlambat
F1-score	Menggabungkan precision dan recall dalam satu ukuran
ROC-AUC	Mengukur kemampuan model membedakan dua kelas prediksi

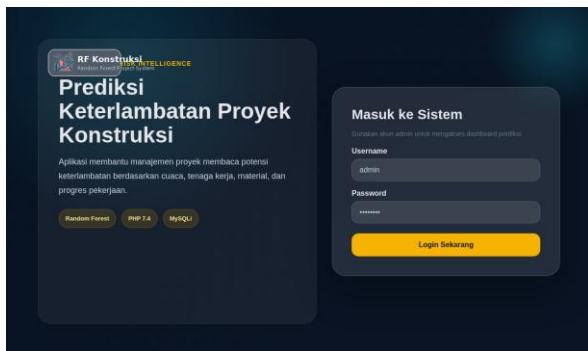
Implementasi Aplikasi

Aplikasi dikembangkan menggunakan PHP 7.4 dan MySQLi. Fitur utama aplikasi terdiri atas login, dashboard, form prediksi, proses klasifikasi, halaman hasil, riwayat prediksi, data training, serta cetak laporan formal. Data input dari pengguna disimpan ke basis data, kemudian diproses oleh modul Random Forest untuk menghasilkan label prediksi, probabilitas risiko, confidence model, dan faktor dominan. Desain aplikasi dibuat sederhana agar pengguna nonteknis tetap dapat memahami hasil prediksi.

4. HASIL DAN PEMBAHASAN

Implementasi Antarmuka Aplikasi

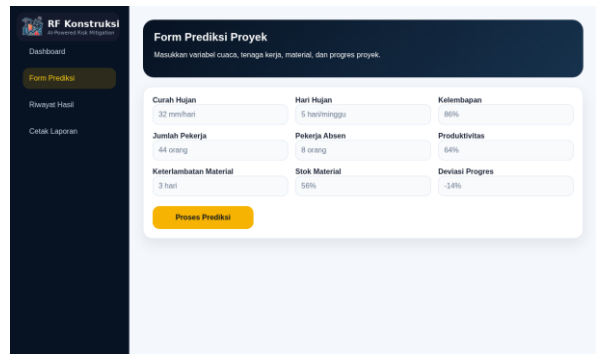
Hasil implementasi aplikasi menunjukkan bahwa sistem telah mampu menyajikan alur kerja prediksi secara terstruktur. Pengguna masuk melalui halaman login, kemudian diarahkan ke dashboard untuk melihat ringkasan data. Setelah itu, pengguna dapat mengisi form prediksi sesuai kondisi proyek. Hasil prediksi ditampilkan dalam bentuk klasifikasi risiko, probabilitas, confidence model, dan faktor dominan.



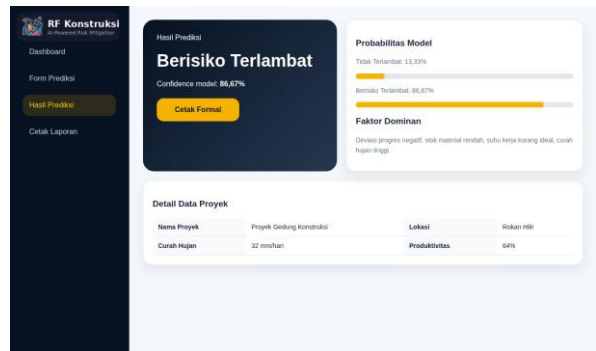
Gambar 1. Tampilan halaman login aplikasi

Gambar 2. Tampilan dashboard sistem prediksi

Pada halaman form prediksi, pengguna mengisi data cuaca, tenaga kerja, material, dan progres proyek. Variabel yang dimasukkan dibuat dalam bentuk numerik agar dapat langsung diproses oleh model. Hal ini menunjukkan bahwa aplikasi tidak hanya menjadi media input data, tetapi juga menjadi penghubung antara data proyek dan model machine learning.



Gambar 3. Tampilan form input prediksi



Gambar 4. Tampilan hasil prediksi risiko keterlambatan

Fitur cetak laporan digunakan untuk menghasilkan dokumen formal dari hasil prediksi. Laporan berisi identitas proyek, nilai variabel input, hasil klasifikasi, probabilitas model, faktor dominan, dan area tanda tangan. Fitur ini penting karena hasil prediksi dalam proyek konstruksi sering perlu dijadikan lampiran evaluasi, bukan hanya dilihat pada layar aplikasi.



Gambar 5. Tampilan cetak laporan formal

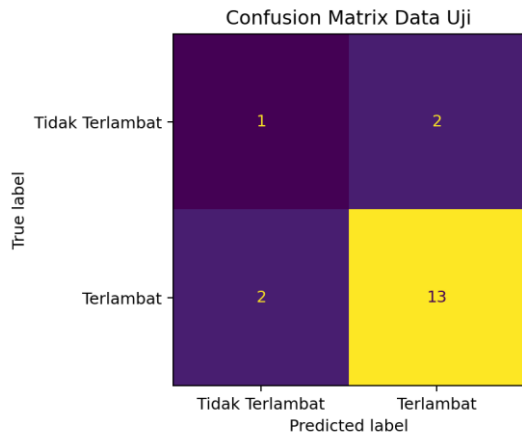
Hasil Evaluasi Model

Evaluasi model dilakukan menggunakan 18 data testing dari dataset prototipe. Hasil pengujian menunjukkan bahwa model Random Forest memperoleh accuracy sebesar 77,78%, precision sebesar 86,67%, recall sebesar 86,67%, F1-score sebesar 86,67%, dan ROC-AUC sebesar 66,67%. Nilai precision dan recall yang cukup tinggi menunjukkan bahwa model cukup baik dalam mengenali kelas berisiko terlambat, meskipun kemampuan pemisahan kelas secara keseluruhan masih perlu diperbaiki dengan data yang lebih besar.

Tabel 3. Hasil Evaluasi Model Random Forest

Metrik	Nilai
Accuracy	77,78%

Precision	86,67%
Recall	86,67%
F1-score	86,67%
ROC-AUC	66,67%

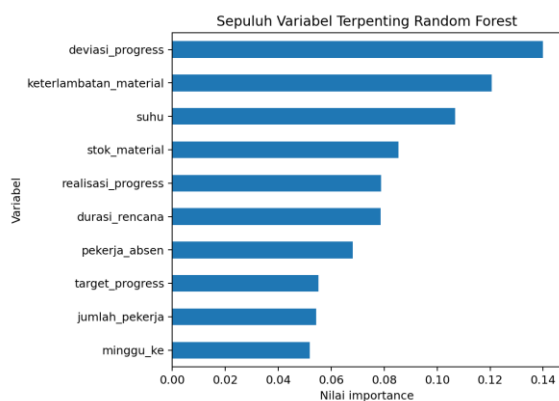


Gambar 6. Confusion matrix hasil pengujian

Berdasarkan confusion matrix, model menghasilkan 1 data benar untuk kelas tidak terlambat, 13 data benar untuk kelas terlambat, 2 data tidak terlambat yang diprediksi terlambat, dan 2 data terlambat yang diprediksi tidak terlambat. Pola ini menunjukkan bahwa model lebih kuat dalam mengenali kelas berisiko terlambat dibandingkan kelas tidak terlambat. Kondisi tersebut dapat terjadi karena data prototipe memiliki distribusi kelas yang belum seimbang.

Analisis Faktor Dominan

Analisis feature importance digunakan untuk mengetahui variabel yang paling sering berkontribusi dalam pengambilan keputusan model. Hasil analisis menunjukkan bahwa deviasi progres, keterlambatan material, produktivitas tenaga kerja, stok material, dan jumlah hari hujan menjadi faktor yang cukup berpengaruh. Secara logis, deviasi progres negatif menunjukkan bahwa realisasi pekerjaan tertinggal dari rencana. Keterlambatan material dan stok material rendah dapat menghambat pekerjaan fisik, sedangkan produktivitas tenaga kerja menentukan kemampuan proyek mengejar target mingguan.



Gambar 7. Feature importance variabel prediksi

Temuan ini sejalan dengan konsep manajemen proyek bahwa keterlambatan tidak hanya disebabkan oleh cuaca,

tetapi juga dipengaruhi oleh kesiapan sumber daya dan pengendalian progres. Random Forest membantu membaca pola tersebut dengan cara membandingkan banyak kemungkinan pemisahan data pada setiap pohon keputusan. Dengan demikian, hasil prediksi tidak hanya berupa label, tetapi juga dapat digunakan untuk memahami faktor risiko yang perlu diprioritaskan oleh pengelola proyek.

Pembahasan

Dari sisi ilmu komputer, penelitian ini menunjukkan bahwa Random Forest dapat menjadi algoritma yang cukup tepat untuk prototipe prediksi keterlambatan proyek karena mampu bekerja pada data tabular dan memberikan interpretasi melalui feature importance. Dibandingkan model tunggal seperti decision tree, Random Forest lebih stabil karena keputusan akhir berasal dari gabungan banyak pohon. Pendekatan ini juga lebih mudah diimplementasikan pada aplikasi sederhana dibandingkan model deep learning yang membutuhkan data lebih besar dan proses komputasi lebih kompleks.

Meskipun hasil evaluasi awal menunjukkan performa yang cukup baik, penelitian ini masih memiliki keterbatasan. Dataset yang digunakan masih berupa dataset prototipe sehingga belum sepenuhnya mewakili keragaman proyek konstruksi nyata. Selain itu, variabel eksternal seperti perubahan desain, pembayaran, ketersediaan alat, subkontraktor, dan perizinan belum dimasukkan dalam model. Oleh karena itu, pengembangan berikutnya perlu menggunakan data proyek aktual dalam jumlah lebih besar serta melakukan perbandingan dengan algoritma lain seperti Logistic Regression, XGBoost, LGBM, dan CatBoost sebagaimana dilakukan dalam penelitian terdahulu [2].

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, algoritma Random Forest dapat diterapkan untuk memprediksi keterlambatan proyek konstruksi berdasarkan faktor cuaca, tenaga kerja, material, dan progres pekerjaan. Model menghasilkan accuracy 77,78%, precision 86,67%, recall 86,67%, F1-score 86,67%, dan ROC-AUC 66,67% pada data testing prototipe. Hasil tersebut menunjukkan bahwa model cukup mampu mengenali proyek yang berisiko terlambat, terutama ketika terdapat deviasi progres negatif, keterlambatan material, produktivitas tenaga kerja rendah, stok material rendah, dan jumlah hari hujan tinggi.

Aplikasi berbasis PHP 7.4 dan MySQLi yang dikembangkan dalam penelitian ini telah mampu menjalankan proses input data, prediksi, penyimpanan riwayat, dan cetak laporan formal. Dengan demikian, sistem yang dibangun dapat menjadi alat bantu awal bagi manajemen proyek untuk membaca risiko keterlambatan secara lebih cepat. Namun, penelitian ini masih perlu dikembangkan menggunakan data proyek riil yang lebih besar, validasi lintas proyek, penyeimbangan kelas data, serta perbandingan dengan algoritma machine learning lain agar akurasi dan generalisasi model menjadi lebih kuat.

Saran untuk penelitian selanjutnya adalah menambahkan variabel perubahan desain, pembayaran, ketersediaan alat, perizinan, dan koordinasi proyek. Selain itu, sistem dapat dikembangkan dengan fitur import data dari Excel, visualisasi tren progres, dashboard risiko mingguan, dan integrasi data cuaca dari sumber resmi. Pengembangan tersebut akan membuat aplikasi lebih dekat dengan kebutuhan pengendalian proyek di lapangan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Institut Teknologi Rokan Hilir atas dukungan akademik dalam penyusunan penelitian ini. Ucapan terima kasih juga disampaikan kepada pihak-pihak yang telah membantu dalam proses penyusunan dataset prototipe, pengembangan aplikasi, dan evaluasi awal model prediksi keterlambatan proyek konstruksi.

DAFTAR PUSTAKA

- [1] S. D. Datta, M. Islam, M. H. R. Sobuz, S. Ahmed, and M. Kar, "Artificial intelligence and machine learning applications in the project lifecycle of the construction industry: A comprehensive review," *Heliyon*, vol. 10, no. 5, Art. no. e26888, 2024, doi: 10.1016/j.heliyon.2024.e26888.
- [2] S. Alsulamy, "Predicting construction delay risks in Saudi Arabian projects: A comparative analysis of CatBoost, XGBoost, and LGBM," *Expert Systems with Applications*, vol. 268, Art. no. 126268, 2025, doi: 10.1016/j.eswa.2024.126268.
- [3] C. N. Egwim, H. Alaka, L. O. Toriola-Coker, H. Balogun, and F. Sunmola, "Applied artificial intelligence for predicting construction projects delay," *Machine Learning with Applications*, vol. 6, Art. no. 100166, 2021, doi: 10.1016/j.mlwa.2021.100166.
- [4] A. Gondia, A. S. Siam, W. El-Dakhkhni, and A. H. Nassar, "Machine learning algorithms for construction projects delay risk prediction," *Journal of Construction Engineering and Management*, vol. 146, no. 1, Art. no. 04019085, 2020, doi: 10.1061/(ASCE)CO.1943-7862.0001736.
- [5] Z. M. Yaseen, Z. H. Ali, S. Q. Salih, and N. Al-Ansari, "Prediction of risk delay in construction projects using a hybrid artificial intelligence model," *Sustainability*, vol. 12, no. 4, Art. no. 1514, 2020, doi: 10.3390/su12041514.
- [6] K. Tian, Z. Zhu, J. Mbachu, A. Ghanbaripour, and M. Moorhead, "Artificial intelligence in risk management within the realm of construction projects: A bibliometric analysis and systematic literature review," *Journal of Innovation & Knowledge*, vol. 10, no. 3, Art. no. 100711, 2025, doi: 10.1016/j.jik.2025.100711.
- [7] S. Ivanova, A. Kuznetsov, R. Zverev, and A. Rada, "Artificial intelligence methods for the construction and management of buildings," *Sensors*, vol. 23, no. 21, Art. no. 8740, 2023, doi: 10.3390/s23218740.
- [8] A. P. Gurgun, K. Koc, and H. Kunkcu, "Exploring the adoption of technology against delays in construction projects," *Engineering, Construction and Architectural Management*, vol. 31, no. 3, pp. 1222–1253, 2024, doi: 10.1108/ECAM-06-2022-0566.
- [9] C. N. Egwim, H. Alaka, E. Demir, H. Balogun, and S. O. Ajayi, "Systematic review of critical drivers for delay risk prediction: towards a conceptual framework for BIM-based construction projects," *Frontiers in Engineering and Built Environment*, vol. 3, no. 1, pp. 16–31, 2023, doi: 10.1108/FEBE-05-2022-0017.
- [10] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [11] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [12] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 3rd ed. Sebastopol, CA: O'Reilly Media, 2022.
- [13] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed. Burlington, MA: Morgan Kaufmann, 2016.
- [14] F. Yuda, I. Meiditra, D. G. A. Candra, Fauziah, and C. Mutia, "Sistem informasi pemesanan lapangan pada Kenzie Futsal Padang menggunakan web," *Riau Journal of Computer Science*, vol. 11, no. 1, pp. 1–10, 2025, doi: 10.30606/rjocs.v11i1.2919.
- [15] Fauziah, I. Meiditra, C. Mutia, F. Yuda, M. Rasyid, R. Agustin, and S. S. Lubis, "Revolusi digital dalam pendidikan: Pemanfaatan teknologi AI (Artificial Intelligence) untuk meningkatkan kualitas pembelajaran," *Jurnal Pengabdian Masyarakat Ilmu Komputer*, vol. 2, no. 1, pp. 45–52, 2025, doi: 10.70248/jpmik.v2i1.1930.