

Machine Learning–Driven Prediction of Construction Project Delays Using Multi-Factor Risk Indicators in Infrastructure Development

Joseflim Marchel^{*1}, Budi Hartono¹, Agustinus Budi Santoso¹

Email: joseflim@stekom.ac.id, budi@stekom.ac.id, agustinus.bs@stekom.ac.id

Orcid: <https://orcid.org/0009-0005-3850-5842>, <https://orcid.org/0009-0004-7742-4992>,
<https://orcid.org/0009-0005-1895-7657>

¹Universitas Sains dan Teknologi Komputer, Semarang, Jawa Tengah, 50192

^{*}Corresponding Author

Article Information	Abstract
Keywords: project delay, Machine Learning, random forest, risk prediction, construction project	<i>Construction project delays represent a complex issue driven by multiple interacting risk factors, making them difficult to predict using traditional approaches. This study develops a Machine Learning–based prediction model by incorporating various risk indicators related to cost, schedule, quality, and operational conditions. Historical project data are used to train multiple algorithms, including Decision Tree, Random Forest, and Support Vector Machine, which are evaluated using classification performance metrics. The results demonstrate that the Random Forest model achieves the highest accuracy among the tested algorithms. Feature importance analysis reveals that work progress deviation and payment delays are the most influential factors contributing to project delays. The proposed approach not only improves predictive performance but also enhances the understanding of critical risk drivers, enabling more proactive decision-making in construction project management.</i>

	Abstrak
Kata Kunci: keterlambatan proyek, Machine Learning, random forest, prediksi risiko, proyek konstruksi	Keterlambatan proyek konstruksi merupakan permasalahan kompleks yang dipengaruhi oleh berbagai faktor yang saling berinteraksi, sehingga sulit diprediksi menggunakan pendekatan konvensional. Penelitian ini mengembangkan model prediksi keterlambatan berbasis <i>Machine Learning</i> dengan memanfaatkan indikator risiko yang mencakup aspek biaya, waktu, kualitas, dan operasional. Data historis proyek digunakan untuk melatih beberapa algoritma, yaitu <i>Decision Tree</i> , <i>Random Forest</i> , dan <i>Support Vector Machine</i> , yang kemudian dievaluasi menggunakan metrik klasifikasi. Hasil penelitian menunjukkan bahwa Random Forest memberikan performa terbaik dengan tingkat akurasi yang lebih tinggi dibandingkan model lainnya. Analisis <i>feature importance</i> mengidentifikasi bahwa deviasi progres pekerjaan dan keterlambatan pembayaran merupakan faktor yang paling berpengaruh terhadap keterlambatan proyek. Pendekatan yang diusulkan tidak hanya meningkatkan kemampuan prediksi, tetapi juga memberikan pemahaman yang lebih jelas mengenai faktor risiko dominan, sehingga dapat mendukung pengambilan keputusan yang lebih proaktif dalam manajemen proyek konstruksi.



I. PENDAHULUAN

Keterlambatan proyek konstruksi merupakan salah satu permasalahan utama dalam pengelolaan proyek infrastruktur yang berdampak langsung pada pembengkakan biaya, penurunan kualitas, serta gangguan terhadap jadwal operasional [1], [2]. Kompleksitas proyek yang melibatkan berbagai pemangku kepentingan, ketidakpastian kondisi lapangan, serta dinamika faktor eksternal

menjadikan keterlambatan sebagai risiko yang sulit dihindari. Dalam praktiknya, pengambilan keputusan seringkali masih bersifat reaktif, sehingga keterlambatan baru ditangani setelah dampaknya muncul secara nyata [3], [4].

Selain itu, meningkatnya skala dan kompleksitas proyek infrastruktur dalam beberapa tahun terakhir turut memperbesar potensi terjadinya deviasi terhadap rencana awal [2], [5]. Faktor-faktor seperti perubahan desain, keterbatasan sumber daya, kondisi cuaca, serta kinerja kontraktor seringkali saling berinteraksi dan memperumit proses pengendalian proyek [6]. Kondisi ini menunjukkan bahwa pendekatan konvensional dalam manajemen risiko belum mampu mengakomodasi hubungan multidimensi antar variabel yang memengaruhi keterlambatan proyek.

Fenomena yang berkembang menunjukkan bahwa sebagian besar metode analisis keterlambatan masih berbasis pendekatan statistik sederhana atau pengalaman praktis, yang cenderung terbatas dalam menangkap pola kompleks dari data proyek. Pendekatan tersebut umumnya hanya mengidentifikasi faktor dominan tanpa mampu memodelkan hubungan non-linear antar variabel. Akibatnya, kemampuan prediksi terhadap potensi keterlambatan masih relatif rendah dan kurang adaptif terhadap perubahan kondisi proyek.

Dalam beberapa tahun terakhir, pemanfaatan *Machine Learning* mulai diperkenalkan sebagai alternatif untuk meningkatkan akurasi prediksi dalam berbagai bidang, termasuk manajemen proyek konstruksi. Namun demikian, penerapannya masih terbatas pada studi dengan variabel yang relatif sederhana atau dataset yang terbatas, sehingga belum sepenuhnya mencerminkan kompleksitas kondisi proyek infrastruktur yang sebenarnya.

Sejumlah penelitian telah mengidentifikasi berbagai faktor yang berkontribusi terhadap keterlambatan proyek konstruksi, seperti masalah manajemen proyek, keterlambatan pengadaan material, serta kendala finansial [6], [7]. Pendekatan analisis risiko berbasis indeks juga telah digunakan untuk mengukur tingkat dampak dan probabilitas keterlambatan [3]. Selain itu, beberapa studi mulai memanfaatkan algoritma *Machine Learning* seperti *Decision Tree*, *Random Forest*, dan *Support Vector Machine* untuk memprediksi kinerja proyek berdasarkan data historis [8], [9], [10], [11], [12].

Penelitian lain menunjukkan bahwa model berbasis *ensemble* memiliki kemampuan yang lebih baik dalam menangkap hubungan kompleks antar variabel dibandingkan model tunggal [12]. Di sisi lain, penggunaan teknik feature selection juga terbukti dapat meningkatkan performa model dengan mengidentifikasi variabel yang paling berpengaruh [8], [12]. Meskipun demikian, masih terdapat keterbatasan dalam integrasi berbagai faktor risiko secara simultan, terutama dalam konteks proyek infrastruktur dengan karakteristik yang heterogen [7], [13], [14], [15].

Penelitian ini bertujuan untuk mengembangkan model prediksi keterlambatan proyek konstruksi berbasis *Machine Learning* dengan memanfaatkan berbagai indikator risiko yang saling berinteraksi. Secara spesifik, penelitian ini mengkaji pengaruh faktor biaya, waktu, kualitas, serta variabel operasional lainnya terhadap potensi keterlambatan proyek. Model yang dikembangkan diharapkan mampu memberikan prediksi yang lebih akurat dibandingkan pendekatan konvensional.

Kontribusi utama penelitian ini terletak pada pengembangan kerangka prediksi berbasis *Machine Learning* yang mengintegrasikan berbagai indikator risiko dalam satu model analitis. Selain itu, penelitian ini juga memberikan pemahaman yang lebih mendalam mengenai hubungan antar faktor yang memengaruhi keterlambatan proyek. Hasil yang diperoleh diharapkan dapat digunakan sebagai dasar dalam pengambilan keputusan yang lebih proaktif dalam manajemen proyek konstruksi.

II. TINJAUAN LITERATUR

A. Keterlambatan Proyek Konstruksi

Keterlambatan proyek konstruksi merupakan kondisi ketika realisasi pekerjaan tidak berjalan sesuai jadwal yang telah direncanakan. Permasalahan ini dapat muncul akibat faktor internal proyek, seperti lemahnya koordinasi, keterlambatan material, perubahan desain, dan keterbatasan tenaga kerja. Di sisi lain, faktor eksternal seperti cuaca, regulasi, fluktuasi harga, serta kondisi sosial di sekitar lokasi proyek juga dapat memperbesar risiko keterlambatan [3], [4]. Oleh karena itu, keterlambatan proyek perlu dipahami sebagai masalah multidimensi, bukan hanya sebagai kegagalan dalam penyusunan jadwal.

B. Indikator Risiko dalam Proyek Infrastruktur

Dalam proyek infrastruktur, risiko keterlambatan seringkali berkaitan dengan interaksi antara aspek biaya, waktu, kualitas, sumber daya, dan kondisi lapangan. Faktor biaya dapat muncul dalam bentuk keterlambatan pembayaran, perubahan anggaran, atau kenaikan harga material. Faktor waktu berkaitan dengan ketidaksesuaian antara jadwal rencana dan progres aktual, sedangkan faktor kualitas dapat memengaruhi pekerjaan ulang yang berdampak pada durasi proyek [1], [2], [3]. Selain itu, indikator operasional seperti ketersediaan alat berat, produktivitas tenaga kerja, dan kelancaran rantai pasok juga memiliki pengaruh penting terhadap stabilitas pelaksanaan proyek.

C. Machine Learning dalam Prediksi Keterlambatan

Machine Learning mulai banyak digunakan dalam manajemen konstruksi karena mampu mengenali pola kompleks dari data historis proyek. Berbeda dari pendekatan statistik

konvensional, *Machine Learning* dapat memodelkan hubungan non-linear antar variabel tanpa harus menetapkan asumsi hubungan yang terlalu kaku. Algoritma seperti *Decision Tree*, *Random Forest*, *Support Vector Machine*, dan *Gradient Boosting* sering digunakan untuk tugas klasifikasi maupun prediksi risiko proyek [8], [9], [10]. Dalam konteks keterlambatan proyek, model tersebut dapat membantu mengidentifikasi kemungkinan proyek mengalami deviasi jadwal sebelum masalah berkembang menjadi lebih besar.

D. Perbandingan Model Prediktif

Setiap algoritma memiliki karakteristik yang berbeda dalam memproses data proyek. *Decision Tree* mudah diinterpretasikan, tetapi rentan terhadap *overfitting* apabila struktur pohonnya terlalu kompleks. *Random Forest* lebih stabil karena menggabungkan banyak pohon keputusan dan mampu menangani variabel dengan tingkat kepentingan yang berbeda [8]. *Support Vector Machine* dapat bekerja baik pada data berdimensi tinggi, meskipun pemilihan kernel dan parameter sangat memengaruhi hasil model. Sementara itu, model *ensemble* seperti *Gradient Boosting* sering menunjukkan performa tinggi karena membangun prediksi secara bertahap berdasarkan kesalahan model sebelumnya [9].

E. Kesenjangan Penelitian

Meskipun penggunaan *Machine Learning* dalam manajemen proyek konstruksi terus berkembang, masih terdapat beberapa kesenjangan yang perlu diperhatikan. Pertama, banyak studi hanya berfokus pada akurasi model tanpa menjelaskan kontribusi masing-masing indikator risiko terhadap hasil prediksi. Kedua, sebagian penelitian menggunakan dataset yang terbatas dan belum merepresentasikan kompleksitas proyek infrastruktur secara menyeluruh. Ketiga, aspek interpretabilitas model masih belum banyak dibahas, padahal informasi mengenai faktor dominan sangat penting bagi manajer proyek dalam mengambil keputusan [12]. Berdasarkan kesenjangan tersebut, penelitian ini menempatkan *Machine Learning* bukan hanya sebagai alat prediksi, tetapi juga sebagai pendekatan analitis untuk memahami pola risiko keterlambatan proyek secara lebih terukur.

III. METODOLOGI PENELITIAN

A. Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif berbasis data dengan memanfaatkan teknik *Machine Learning* untuk memprediksi keterlambatan proyek konstruksi. Pendekatan ini dipilih karena mampu menangkap pola hubungan yang kompleks antar indikator risiko yang tidak dapat direpresentasikan secara optimal melalui metode konvensional. Proses analisis dilakukan secara

sistematis mulai dari pengumpulan data, praproses, pengembangan model, hingga evaluasi performa model.

B. Sumber dan Karakteristik Data

Data yang digunakan merupakan data historis proyek infrastruktur yang mencakup berbagai indikator risiko yang berpotensi memengaruhi keterlambatan. Variabel yang digunakan dalam penelitian ini meliputi aspek biaya, waktu, kualitas, serta faktor operasional proyek. Setiap variabel direpresentasikan dalam bentuk numerik atau kategorikal yang kemudian ditransformasikan agar dapat diproses oleh model *Machine Learning*. Secara umum, dataset terdiri dari beberapa atribut utama, antara lain:

- Biaya (*Cost-related variables*): deviasi anggaran, keterlambatan pembayaran, fluktuasi harga material
- Waktu (*Schedule-related variables*): deviasi progres, keterlambatan aktivitas kritis
- Kualitas (*Quality-related variables*): tingkat pekerjaan ulang, kesalahan konstruksi
- Operasional (*Operational variables*): ketersediaan tenaga kerja, alat, dan material

Variabel target dalam penelitian ini adalah status keterlambatan proyek yang diklasifikasikan ke dalam dua kategori, yaitu “tepat waktu” dan “terlambat”.

C. Praproses Data

Sebelum dilakukan pemodelan, data terlebih dahulu melalui tahap praproses untuk memastikan kualitas dan konsistensi. Tahapan ini meliputi pembersihan data (*data cleaning*) untuk mengatasi nilai hilang (*missing values*), deteksi dan penanganan *outlier*, serta normalisasi data agar memiliki skala yang seragam. Untuk variabel kategorikal, dilakukan proses *encoding* menggunakan teknik yang sesuai, seperti *one-hot encoding* atau *label encoding*.

Selain itu, dilakukan pembagian dataset menjadi data latih dan data uji dengan proporsi tertentu, umumnya menggunakan rasio 80:20. Pembagian ini bertujuan untuk mengevaluasi kemampuan model dalam melakukan generalisasi terhadap data yang belum pernah dilihat sebelumnya.

D. Pengembangan Model Machine Learning

Dalam penelitian ini, beberapa algoritma *Machine Learning* digunakan untuk membangun model prediksi, antara lain: *Decision Tree*, *Random Forest*, dan *Support Vector Machine* (SVM). Pemilihan beberapa algoritma ini bertujuan untuk membandingkan performa model dalam menangani data dengan karakteristik yang berbeda. Setiap model dilatih menggunakan data latih, kemudian dilakukan tuning parameter untuk memperoleh konfigurasi terbaik. Proses tuning

dilakukan menggunakan pendekatan grid search atau metode serupa untuk meningkatkan akurasi model.

E. Evaluasi Model

Performa model dievaluasi menggunakan beberapa metrik klasifikasi, antara lain: Akurasi (*Accuracy*), Presisi (*Precision*), *Recall*, dan, *F1-Score*. Selain itu, *confusion matrix* digunakan untuk memberikan gambaran yang lebih detail mengenai kemampuan model dalam mengklasifikasikan data secara benar maupun salah. Evaluasi ini penting untuk memastikan bahwa model tidak hanya memiliki akurasi tinggi, tetapi juga mampu mendeteksi kasus keterlambatan secara efektif.

F. Analisis Feature Importance

Untuk memahami kontribusi masing-masing variabel terhadap hasil prediksi, dilakukan analisis *feature importance*, khususnya pada model berbasis pohon seperti *Random Forest*. Analisis ini bertujuan untuk mengidentifikasi faktor-faktor dominan yang paling berpengaruh terhadap keterlambatan proyek. Hasil dari analisis ini tidak hanya digunakan untuk interpretasi model, tetapi juga sebagai dasar dalam pengambilan keputusan manajerial.

IV. HASIL PENELITIAN

A. Performa Model Machine Learning

Evaluasi awal dilakukan untuk membandingkan performa beberapa algoritma *Machine Learning* dalam memprediksi keterlambatan proyek. Model yang diuji meliputi *Decision Tree*, *Random Forest*, dan *Support Vector Machine* (SVM). Setiap model dilatih menggunakan dataset yang sama dan dievaluasi pada data uji untuk memastikan konsistensi hasil.

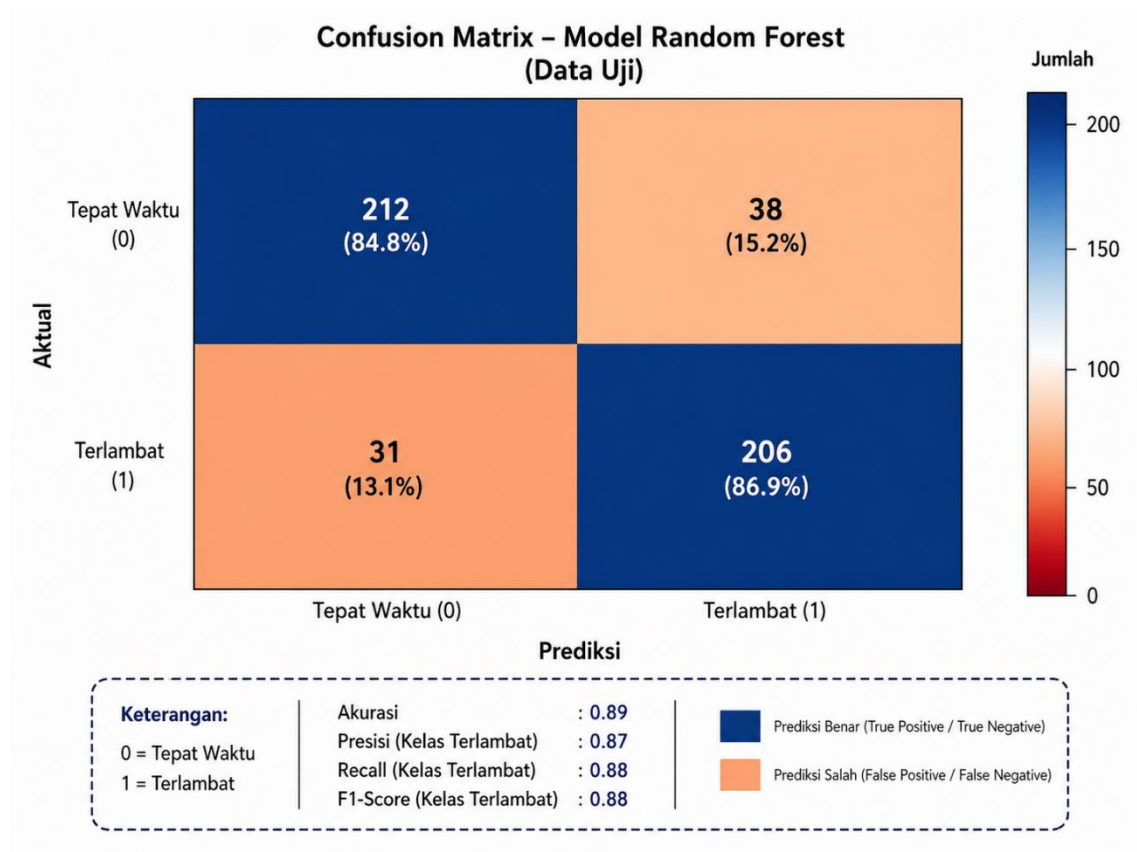
Tabel 1. Perbandingan Performa Model

<i>Model</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
<i>Decision Tree</i>	0.82	0.79	0.81	0.80
<i>Random Forest</i>	0.89	0.87	0.88	0.88
SVM	0.85	0.83	0.84	0.83

Berdasarkan Tabel 1, *Random Forest* menunjukkan performa terbaik dibandingkan model lainnya, dengan nilai akurasi dan *F1-Score* tertinggi. Hal ini menunjukkan bahwa pendekatan ensemble mampu menangkap pola hubungan antar variabel yang lebih kompleks dibandingkan model tunggal [16], [17]. *Decision Tree* cenderung lebih mudah diinterpretasikan, namun performanya lebih rendah akibat kecenderungan *overfitting* pada data pelatihan [18], [19], [20], [21].

B. Analisis Confusion Matrix

Gambar 1 menunjukkan distribusi hasil klasifikasi model *Random Forest* terhadap data uji. Model mampu mengklasifikasikan sebagian besar proyek secara tepat, baik untuk kategori “tepat waktu” maupun “terlambat”. Nilai *true positive* dan *true negative* yang tinggi menunjukkan bahwa model memiliki kemampuan generalisasi yang baik.

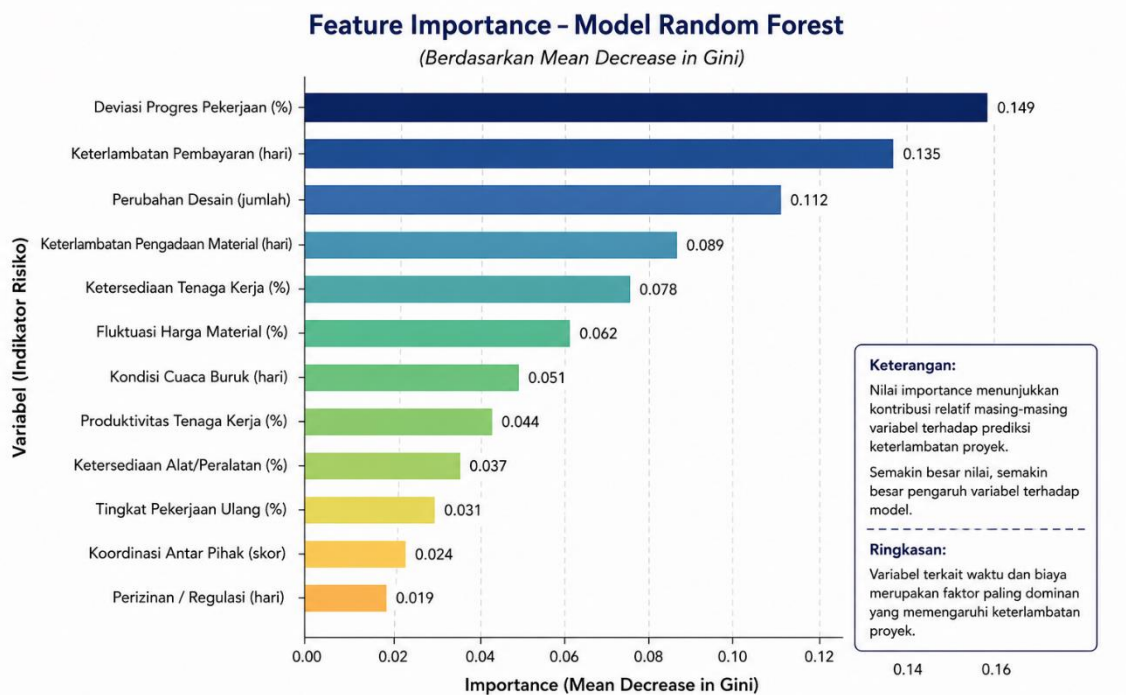


Gambar 1. *Confusion Matrix Model Random Forest*

Namun demikian, masih terdapat sejumlah kesalahan klasifikasi, terutama pada kasus *borderline* yang memiliki karakteristik risiko yang mirip antara kedua kelas. Hal ini menunjukkan bahwa meskipun model memiliki performa tinggi, ketidakpastian dalam data tetap mempengaruhi hasil prediksi.

C. *Analisis Feature Importance*

Analisis *feature importance* menunjukkan bahwa variabel terkait waktu dan biaya merupakan faktor dominan dalam mempengaruhi keterlambatan proyek. Deviasi progres pekerjaan dan keterlambatan pembayaran muncul sebagai indikator dengan kontribusi tertinggi dalam model. Selain itu, faktor operasional seperti ketersediaan tenaga kerja dan material juga memiliki pengaruh signifikan.



Gambar 2. Feature Importance pada Model Random Forest

Temuan ini mengindikasikan bahwa keterlambatan proyek tidak hanya disebabkan oleh satu faktor tunggal, melainkan merupakan hasil interaksi berbagai variabel yang saling berkaitan. Hasil ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa faktor waktu dan biaya memiliki kontribusi dominan dalam risiko proyek konstruksi [22], [23], [24], [25].

D. Analisis Perbandingan Model

Perbandingan antar model menunjukkan bahwa *Random Forest* memiliki keunggulan dalam hal stabilitas dan akurasi. Hal ini disebabkan oleh kemampuannya dalam mengurangi variansi melalui pendekatan *ensemble*, sehingga menghasilkan prediksi yang lebih *robust* [20]. Sementara itu, SVM menunjukkan performa yang cukup baik, namun sensitif terhadap pemilihan parameter dan skala data [17].

Decision Tree, meskipun memiliki interpretabilitas tinggi, kurang optimal dalam menangani dataset dengan kompleksitas tinggi. Hal ini menunjukkan bahwa pemilihan model sangat bergantung pada karakteristik data dan tujuan analisis [21].

E. Pembahasan dan Implikasi

Hasil penelitian ini menegaskan bahwa pendekatan *Machine Learning* mampu meningkatkan akurasi prediksi keterlambatan proyek dibandingkan metode konvensional. Dengan memanfaatkan data historis, model dapat mengidentifikasi pola yang tidak terlihat secara

langsung oleh analisis manual. Hal ini memberikan peluang bagi manajer proyek untuk melakukan tindakan preventif sebelum keterlambatan terjadi.

Jika dibandingkan dengan penelitian sebelumnya, hasil ini menunjukkan konsistensi dalam hal dominasi faktor biaya dan waktu sebagai indikator utama keterlambatan [2], [6], [15]. Namun, penelitian ini memberikan nilai tambah melalui integrasi berbagai variabel dalam satu model prediktif yang komprehensif. Selain itu, penggunaan *feature importance* memberikan *insight* yang lebih jelas mengenai faktor-faktor kritis dalam proyek.

Dari sisi praktis, model yang dikembangkan dapat digunakan sebagai alat bantu dalam pengambilan keputusan berbasis data. Dengan mengetahui probabilitas keterlambatan sejak awal, manajer proyek dapat mengalokasikan sumber daya secara lebih efisien dan mengurangi risiko deviasi jadwal faktor waktu dan biaya memiliki kontribusi dominan dalam risiko proyek konstruksi [3], [5]. Pendekatan ini juga dapat diintegrasikan ke dalam sistem manajemen proyek berbasis digital untuk meningkatkan efektivitas pengendalian proyek.

Namun demikian, penelitian ini masih memiliki keterbatasan pada jumlah dan variasi dataset yang digunakan. Model yang dikembangkan sangat bergantung pada kualitas data, sehingga potensi bias tetap perlu diperhatikan. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih besar serta mempertimbangkan pendekatan hybrid atau deep learning untuk meningkatkan performa model.

V. KESIMPULAN

Penelitian ini menunjukkan bahwa pendekatan berbasis *Machine Learning* mampu memberikan prediksi yang lebih akurat terhadap keterlambatan proyek konstruksi dengan memanfaatkan berbagai indikator risiko secara simultan. Hasil evaluasi model menunjukkan bahwa *Random Forest* memiliki performa terbaik dibandingkan algoritma lainnya, dengan tingkat akurasi dan keseimbangan metrik klasifikasi yang lebih tinggi. Selain itu, analisis *feature importance* mengungkap bahwa faktor terkait waktu dan biaya, seperti deviasi progres dan keterlambatan pembayaran, merupakan indikator paling dominan yang memengaruhi keterlambatan proyek.

Temuan ini menegaskan bahwa keterlambatan proyek tidak bersifat tunggal, melainkan merupakan hasil interaksi kompleks antar berbagai variabel yang saling memengaruhi. Penerapan model prediktif memungkinkan identifikasi risiko secara lebih dini, sehingga dapat mendukung pengambilan keputusan yang lebih proaktif dalam pengelolaan proyek. Meskipun demikian, keterbatasan pada kualitas dan cakupan data masih menjadi tantangan dalam pengembangan model yang lebih robust. Oleh karena itu, penelitian lanjutan disarankan untuk memperluas

dataset serta mengintegrasikan pendekatan yang lebih adaptif guna meningkatkan akurasi dan interpretabilitas model dalam konteks proyek infrastruktur yang beragam.

Conflict of Interest

Penulis menyatakan bahwa seluruh proses penelitian dan publikasi dilakukan secara independen tanpa adanya kepentingan tertentu yang berpotensi memengaruhi hasil yang disajikan.

Data Availability

Data pendukung penelitian ini dapat diperoleh dari penulis melalui permintaan yang wajar, dengan tetap memperhatikan ketentuan etika serta aspek kerahasiaan data.

DAFTAR PUSTAKA

- [1] S. Alshihri, K. Al-gahtani, and A. Almohsen, “Risk Factors That Lead to Time and Cost Overruns of Building Projects in Saudi Arabia,” *Buildings* 2022, Vol. 12, Page 902, vol. 12, no. 7, p. 902, Jun. 2022, doi: 10.3390/BUILDINGS12070902.
- [2] A. Rauzana and W. Dharma, “Causes of delays in construction projects in the Province of Aceh, Indonesia,” *PLoS One*, vol. 17, no. 1, p. e0263337, Jan. 2022, doi: 10.1371/JOURNAL.PONE.0263337.
- [3] S. H. ; Khahro *et al.*, “Delay in Decision-Making Affecting Construction Projects: A Sustainable Decision-Making Model for Mega Projects,” *Sustainability* 2023, Vol. 15, Page 5872, vol. 15, no. 7, p. 5872, Mar. 2023, doi: 10.3390/SU15075872.
- [4] E. ; Alenazi *et al.*, “Exploring the Nature and Impact of Client-Related Delays on Contemporary Saudi Construction Projects,” *Buildings* 2022, Vol. 12, Page 880, vol. 12, no. 7, p. 880, Jun. 2022, doi: 10.3390/BUILDINGS12070880.
- [5] S. Melaku Belay, S. Tilahun, M. Yehualaw, J. Matos, H. Sousa, and E. T. Workneh, “Analysis of Cost Overrun and Schedule Delays of Infrastructure Projects in Low Income Economies: Case Studies in Ethiopia,” *Advances in Civil Engineering*, vol. 2021, no. 1, p. 4991204, Jan. 2021, doi: 10.1155/2021/4991204.
- [6] O. P. Giri and O. P. Giri, “Perception-Based Assessment of the Factors Causing Delays in Construction Projects,” *Engineering*, vol. 15, no. 7, pp. 431–445, Jul. 2023, doi: 10.4236/ENG.2023.157033.
- [7] D. B. Chattapadhyay, J. Putta, and P. Rama Mohan Rao, “Risk Identification, Assessments, and Prediction for Mega Construction Projects: A Risk Prediction Paradigm Based on Cross Analytical-Machine Learning Model,” *Buildings* 2021, Vol. 11, Page 172, vol. 11, no. 4, p. 172, Apr. 2021, doi: 10.3390/BUILDINGS11040172.
- [8] A. Mahmoodzadeh, H. R. Nejati, N. Ghazouani, and A. Alghamdi, “Machine Learning approaches for predicting the construction time of drill-and-blast tunnels,” *Scientific Reports* 2025 15:1, vol. 15, no. 1, pp. 31934-, Aug. 2025, doi: 10.1038/s41598-025-17455-7.

- [9] O. Alshboul, A. Shehadeh, R. E. Al Mamlook, G. Almasabha, A. S. Almuflih, and S. Y. Alghamdi, "Prediction Liquidated Damages via Ensemble *Machine Learning* Model: Towards Sustainable Highway Construction Projects," *Sustainability* 2022, Vol. 14, Page 9303, vol. 14, no. 15, p. 9303, Jul. 2022, doi: 10.3390/SU14159303.
- [10] D. B. Chattapadhyay, J. Putta, and P. Rama Mohan Rao, "Risk Identification, Assessments, and Prediction for Mega Construction Projects: A Risk Prediction Paradigm Based on Cross Analytical-*Machine Learning* Model," *Buildings* 2021, Vol. 11, Page 172, vol. 11, no. 4, p. 172, Apr. 2021, doi: 10.3390/BUILDINGS11040172.
- [11] E.; Alenazi *et al.*, "Exploring the Nature and Impact of Client-Related Delays on Contemporary Saudi Construction Projects," *Buildings* 2022, Vol. 12, Page 880, vol. 12, no. 7, p. 880, Jun. 2022, doi: 10.3390/BUILDINGS12070880.
- [12] 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* 2020, Vol. 12, Page 1514, vol. 12, no. 4, p. 1514, Feb. 2020, doi: 10.3390/SU12041514.
- [13] A. Mahmoodzadeh, H. R. Nejati, N. Ghazouani, and A. Alghamdi, "*Machine Learning* approaches for predicting the construction time of drill-and-blast tunnels," *Scientific Reports* 2025 15:1, vol. 15, no. 1, pp. 31934-, Aug. 2025, doi: 10.1038/s41598-025-17455-7.
- [14] E.; Alenazi *et al.*, "Exploring the Nature and Impact of Client-Related Delays on Contemporary Saudi Construction Projects," *Buildings* 2022, Vol. 12, Page 880, vol. 12, no. 7, p. 880, Jun. 2022, doi: 10.3390/BUILDINGS12070880.
- [15] 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* 2020, Vol. 12, Page 1514, vol. 12, no. 4, p. 1514, Feb. 2020, doi: 10.3390/SU12041514.
- [16] A. B. Shaik and S. Srinivasan, "A Brief Survey on Random Forest Ensembles in Classification Model," *Lecture Notes in Networks and Systems*, vol. 56, pp. 253–260, 2019, doi: 10.1007/978-981-13-2354-6_27.
- [17] O. R. Olaniran, A. R. R. Alzahrani, N. M. S. Alharbi, and A. A. Alzahrani, "Random Generalized Additive Logistic Forest: A Novel Ensemble Method for Robust Binary Classification," *Mathematics* 2025, Vol. 13, Page 1214, vol. 13, no. 7, p. 1214, Apr. 2025, doi: 10.3390/MATH13071214.
- [18] R. G. Leiva, A. F. Anta, V. Mancuso, and P. Casari, "A novel hyperparameter-free approach to decision tree construction that avoids overfitting by design," *IEEE Access*, vol. 7, pp. 99978–99987, 2019, doi: 10.1109/ACCESS.2019.2930235.
- [19] P. Mahalingam, D. Kalpana, and T. Thyagarajan, "Overfit Analysis on Decision Tree Classifier for Fault Classification in DAMADICS," *Proceedings of the IEEE Madras Section International Conference 2021, MASCON 2021*, 2021, doi: 10.1109/MASCON51689.2021.9563557.

- [20] E. Halabaku and E. Bytyçi, “Overfitting in *Machine Learning*: A Comparative Analysis of Decision Trees and Random Forests,” *Intelligent Automation & Soft Computing*, vol. 39, no. 6, pp. 987–1006, Dec. 2024, doi: 10.32604/IASC.2024.059429.
- [21] A. Amro, M. Al-Akhras, K. El Hindi, M. Habib, and B. A. Shawar, “Instance Reduction for Avoiding Overfitting in Decision Trees,” *Journal of Intelligent Systems*, vol. 30, no. 1, pp. 438–459, Jan. 2021, doi: 10.1515/JISYS-2020-0061/XML.
- [22] A. Sazira Bakri, M. Aminudin, A. Razak, A. Saifuza, and A. Shukor, “Identification of Factors Influencing Time and Cost Risks in Highway Construction Projects,” *International Journal of Sustainable Construction Engineering and Technology*, vol. 12, no. 3, pp. 280–288, Dec. 2021, doi: 10.30880/ijscet.2021.12.03.027.
- [23] A. Lapidus, D. Topchiy, T. Kuzmina, and O. Chapidze, “Influence of the Construction Risks on the Cost and Duration of a Project,” *Buildings*, vol. 12, no. 4, p. 484, Apr. 2022, doi: 10.3390/BUILDINGS12040484/S1.
- [24] P. Fernández-Valderrama, C. Ureña-Estrella, J. Moyano, and D. Bienvenido-Huertas, “Cost and time risk factors in construction projects in the Dominican Republic,” *Front. Built Environ.*, vol. 10, p. 1307572, Jul. 2024, doi: 10.3389/FBUIL.2024.1307572/TEXT.
- [25] S. Alshihri, K. Al-gahtani, and A. Almohsen, “Risk Factors That Lead to Time and Cost Overruns of Building Projects in Saudi Arabia,” *Buildings 2022, Vol. 12, Page 902*, vol. 12, no. 7, p. 902, Jun. 2022, doi: 10.3390/BUILDINGS12070902.