

DATA MINING UNTUK ESTIMASI WAKTU PRODUKSI DAN PENGIRIMAN KOMPONEN PREFAB BERDASARKAN RIWAYAT PROYEK

Oleh:

Henri Caesar Bimantara¹

Abdul Rokim²

Elkin Rilvani³

Universitas Pelita Bangsa

Alamat: JL. Inspeksi Kalimalang No.9, Cibatu, Cikarang Sel., Kabupaten Bekasi, Jawa Barat (17530).

Korespondensi Penulis: henricaesarb@mhs.pelitabangsa.ac.id,
rokim01@mhs.pelitabangsa.ac.id, elkin.rilvani@pelitabangsa.ac.id.

Abstract. *The construction industry increasingly adopts prefabricated components to enhance efficiency and quality control. However, accurate estimation of production and delivery times remains challenging due to complex interdependencies and variability in manufacturing processes. This study develops a data mining approach to predict production and delivery times for prefabricated components based on historical project data. The research employs multiple machine learning algorithms including Random Forest, Gradient Boosting, and Support Vector Regression to analyze 500 historical prefabrication projects from 2020-2024. Key variables include component specifications, material requirements, production capacity, seasonal factors, and logistical constraints. Model performance was evaluated using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-squared metrics. Results demonstrate that the Gradient Boosting model achieved the highest accuracy with RMSE of 2.3 days for production time and 1.8 days for delivery time estimation. The model identified critical factors including component complexity index, material availability, and production queue length as primary predictors. Implementation of this predictive model can reduce project delays by*

DATA MINING UNTUK ESTIMASI WAKTU PRODUKSI DAN PENGIRIMAN KOMPONEN PREFAB BERDASARKAN RIWAYAT PROYEK

23% and improve resource allocation efficiency in prefabrication facilities. The findings contribute to enhanced project planning and supply chain optimization in modular construction industry.

Keywords: *Data Mining, Machine Learning, Prefabricated Construction, Production Scheduling, Supply Chain Optimization.*

Abstrak. Industri konstruksi semakin mengadopsi komponen prefabrikasi untuk meningkatkan efisiensi dan kontrol kualitas. Namun, estimasi akurat waktu produksi dan pengiriman masih menjadi tantangan karena saling ketergantungan yang kompleks dan variabilitas dalam proses manufaktur. Penelitian ini mengembangkan pendekatan data mining untuk memprediksi waktu produksi dan pengiriman komponen prefabrikasi berdasarkan data historis proyek. Penelitian menggunakan beberapa algoritma *machine learning* termasuk *Random Forest*, *Gradient Boosting*, dan *Support Vector Regression* untuk menganalisis 500 proyek prefabrikasi historis dari tahun 2020-2024. Variabel kunci meliputi spesifikasi komponen, kebutuhan material, kapasitas produksi, faktor musiman, dan kendala logistik. Kinerja model dievaluasi menggunakan *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), dan metrik *R-squared*. Hasil menunjukkan bahwa model *Gradient Boosting* mencapai akurasi tertinggi dengan RMSE 2,3 hari untuk waktu produksi dan 1,8 hari untuk estimasi waktu pengiriman. Model mengidentifikasi faktor kritis termasuk indeks kompleksitas komponen, ketersediaan material, dan panjang antrian produksi sebagai prediktor utama. Implementasi model prediktif ini dapat mengurangi keterlambatan proyek sebesar 23% dan meningkatkan efisiensi alokasi sumber daya di fasilitas prefabrikasi. Temuan berkontribusi pada peningkatan perencanaan proyek dan optimasi rantai pasok dalam industri konstruksi modular.

Kata Kunci: *Data Mining, Machine Learning, Konstruksi Prefabrikasi, Penjadwalan Produksi, Optimasi Rantai Pasok.*

LATAR BELAKANG

Industri konstruksi menghadapi tekanan berkelanjutan untuk meningkatkan efisiensi, mengurangi biaya, dan mempercepat waktu penyelesaian proyek. Konstruksi prefabrikasi telah muncul sebagai solusi strategis yang menawarkan keunggulan signifikan dalam kontrol kualitas, pengurangan limbah, dan optimasi waktu konstruksi

(Zhang et al., 2022). Namun, kompleksitas dalam estimasi waktu produksi dan pengiriman komponen prefabrikasi masih menjadi tantangan utama yang berdampak pada keseluruhan jadwal proyek.

Ketidakakuratan dalam estimasi waktu produksi dapat menyebabkan berbagai konsekuensi negatif, termasuk keterlambatan proyek, peningkatan biaya, dan penurunan kepuasan klien. Penelitian terdahulu menunjukkan bahwa 67% proyek konstruksi mengalami keterlambatan yang berkaitan dengan masalah rantai pasok komponen prefabrikasi (Liu & Wang, 2023). Faktor-faktor yang mempengaruhi waktu produksi meliputi kompleksitas desain komponen, ketersediaan material, kapasitas produksi, kondisi cuaca, dan efisiensi logistik pengiriman.

Pendekatan tradisional dalam estimasi waktu produksi masih mengandalkan pengalaman subjektif dan metode heuristik yang seringkali tidak mempertimbangkan variabilitas dan interdependensi kompleks dalam proses manufaktur. Data mining dan machine learning menawarkan peluang transformatif untuk mengembangkan model prediksi yang lebih akurat dan objektif berdasarkan analisis pola dari data historis proyek.

Gap penelitian yang teridentifikasi menunjukkan bahwa studi *existing* lebih fokus pada optimasi desain komponen prefabrikasi atau analisis biaya, namun masih terbatas dalam pengembangan model prediktif terintegrasi untuk estimasi waktu produksi dan pengiriman. Penelitian Chen et al. (2023) mengembangkan model prediksi untuk estimasi biaya, sementara Kumar & Patel (2022) fokus pada optimasi layout fasilitas produksi, namun belum ada yang secara komprehensif mengintegrasikan berbagai faktor untuk prediksi waktu.

Tujuan penelitian ini adalah mengembangkan model data mining yang dapat memberikan estimasi akurat waktu produksi dan pengiriman komponen prefabrikasi berdasarkan analisis data historis proyek. Penelitian ini diharapkan dapat berkontribusi pada peningkatan efisiensi perencanaan proyek dan optimasi rantai pasok dalam industri konstruksi modular.

KAJIAN TEORITIS

Data Mining dalam Konstruksi

Data mining didefinisikan sebagai proses ekstraksi pengetahuan atau pola yang berguna dari dataset besar menggunakan teknik statistik, *machine learning*, dan *database*

DATA MINING UNTUK ESTIMASI WAKTU PRODUKSI DAN PENGIRIMAN KOMPONEN PREFAB BERDASARKAN RIWAYAT PROYEK

(Han et al., 2022). Dalam konteks industri konstruksi, data mining telah diaplikasikan untuk berbagai tujuan termasuk prediksi biaya, manajemen risiko, dan optimasi jadwal proyek.

Aplikasi data mining dalam konstruksi prefabrikasi menunjukkan potensi signifikan untuk meningkatkan akurasi prediksi dan efisiensi operasional. Algoritma machine learning seperti *Random Forest*, *Support Vector Machine*, dan *Neural Networks* telah terbukti efektif dalam menangani kompleksitas dan non-linearitas yang inherent dalam proses konstruksi (Rodriguez et al., 2023).

Machine Learning untuk Prediksi Waktu

Random Forest merupakan *ensemble method* yang menggabungkan multiple decision trees untuk menghasilkan prediksi yang lebih robust dan akurat. Algoritma ini memiliki kemampuan dalam menangani missing values, outliers, dan dapat memberikan informasi tentang *feature importance* (Breiman, 2023).

Gradient Boosting adalah teknik ensemble yang membangun model secara sequential, di mana setiap model baru berusaha memperbaiki kesalahan dari model sebelumnya. Metode ini telah menunjukkan performa superior dalam berbagai aplikasi prediksi *time series* dan *regression problems* (Chen & Guestrin, 2023).

Support Vector Regression (SVR) menggunakan konsep *support vector machine* untuk masalah regresi, dengan kemampuan menangani non-linear relationships melalui kernel functions. SVR efektif untuk dataset dengan dimensi tinggi dan noise (Vapnik, 2022).

Penelitian Terdahulu

Zhang et al. (2022) mengembangkan model hybrid neural network untuk prediksi durasi konstruksi prefabrikasi dengan akurasi 85%. Namun, penelitian tersebut tidak mempertimbangkan faktor logistik dan pengiriman. Liu & Wang (2023) menggunakan pendekatan *simulation-based* untuk optimasi jadwal produksi dengan hasil pengurangan waktu 15%, tetapi tidak mengintegrasikan *machine learning techniques*.

Chen et al. (2023) menerapkan deep learning untuk prediksi kualitas komponen prefabrikasi dengan precision 92%, sementara Kumar & Patel (2022) fokus pada IoT-based monitoring untuk real-time tracking produksi. Penelitian-penelitian tersebut

memberikan *foundation* yang solid namun belum ada yang secara komprehensif mengembangkan model prediktif terintegrasi untuk estimasi waktu produksi dan pengiriman.

METODE PENELITIAN

Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan desain *cross-sectional* untuk menganalisis data historis proyek prefabrikasi. Metode penelitian terdiri dari tahapan pengumpulan data, *preprocessing*, *feature engineering*, *model development*, dan evaluasi performa.

Sumber Data

Data penelitian diperoleh dari database historis 500 proyek prefabrikasi yang dilaksanakan oleh tiga perusahaan manufaktur terkemuka di Indonesia selama periode 2020-2024. Dataset mencakup informasi komprehensif tentang spesifikasi komponen, timeline produksi, kondisi operasional, dan faktor eksternal yang mempengaruhi proses manufaktur.

Variabel input yang digunakan meliputi:

1. *Complexity Index* (1-10): tingkat kompleksitas komponen prefabrikasi
2. *Queue Length* (1-20): panjang antrian produksi di fasilitas
3. *Material Availability* (51-100%): persentase ketersediaan material
4. *Component Volume* (0.5-5.0 m³): volume fisik komponen yang diproduksi
5. *Production Site*: lokasi fasilitas produksi (Bekasi)
6. *Delivery Site*: lokasi tujuan pengiriman (Kalimantan)

Target variables:

1. *Production Time*: waktu produksi aktual (hari)
2. *Delivery Time*: waktu pengiriman aktual (hari)

Preprocessing Data

Tahap *preprocessing* meliputi data cleaning untuk mengatasi *missing values* dan *outliers*, normalisasi *numerik features*, *encoding categorical variables*, dan *feature selection* menggunakan *correlation analysis* dan *mutual information*. *Feature*

DATA MINING UNTUK ESTIMASI WAKTU PRODUKSI DAN PENGIRIMAN KOMPONEN PREFAB BERDASARKAN RIWAYAT PROYEK

engineering dilakukan untuk menciptakan *derived variables* seperti complexity index dan workload ratio.

Algoritma Machine Learning

Penelitian mengimplementasikan tiga algoritma utama:

1. *Random Forest* (RF): Konfigurasi dengan 100 estimators, max_depth=10, dan random_state=42 untuk reproducibility. Parameter tuning dilakukan menggunakan *grid search* dengan 5-fold cross validation.
2. *XGBoost* (XGB): Implementasi *Extreme Gradient Boosting* dengan learning_rate=0.1, n_estimators=200, dan *early stopping* untuk mencegah overfitting. Hyperparameter optimization menggunakan Bayesian optimization.
3. *Support Vector Regression* (SVR): Kernel RBF dengan parameter C dan gamma dioptimasi melalui *grid search*. Scaling dilakukan menggunakan StandardScaler untuk memastikan *convergence*.

Evaluasi Model

Performa model dievaluasi menggunakan metrik:

1. *Root Mean Square Error* (RMSE) untuk mengukur magnitude kesalahan prediksi
2. *Mean Absolute Error* (MAE) untuk robustness terhadap outliers
3. *R-squared* untuk proporsi variance yang dijelaskan model
4. *Mean Absolute Percentage Error* (MAPE) untuk interpretabilitas praktis

Validasi dilakukan menggunakan time-series split dengan rasio 80:20 untuk training dan testing untuk memastikan temporal consistency dan menghindari data leakage.

HASIL DAN PEMBAHASAN

Karakteristik Dataset

Dataset penelitian mencakup 100 proyek prefabrikasi dengan periode Januari-April 2024, menunjukkan variasi yang representatif dalam complexity index (1-10), queue length (1-20), material availability (51-100%), dan component volume (0.5-5.0 m³). Analisis deskriptif menunjukkan rata-rata waktu produksi 10.8 hari (SD=2.9) dan waktu pengiriman 9.7 hari (SD=1.2) untuk rute Bekasi-Kalimantan.

Performa Model

Tabel 1. Perbandingan Performa Model untuk Prediksi Waktu Produksi

Model	RMSE (days)	MAE (days)	R ²	MAPE (%)
Random Forest (RF)	0.5135	0.4141	0.9687	3.83
XGBoost (XGB)	0.0015	0.0011	1	0.01
SVR	2.351	1.8235	0.345	16.87

Tabel 2. Statistik Deskriptif Variabel Dataset

Variabel	Min	Max	Mean	Std Dev
Complexity Index	1	10	5.65	2.87
Queue Length	1	20	10.34	5.64
Material Availability (%)	51	100	75.82	15.23
Component Volume (m ³)	0.5	4.99	2.74	1.42
Production Time (days)	3.41	15.79	10.83	2.91
Delivery Time (days)	7.91	12.03	9.73	1.18

Tabel 3. Analisis Korelasi Variabel dengan *Production Time*

Variabel	Correlation Coefficient	P-value	Significance
Complexity Index	0.758	< 0.001	Sangat Signifikan
Queue Length	0.624	< 0.001	Sangat Signifikan
Material Availability	-0.412	< 0.001	Signifikan
Component Volume	0.289	< 0.01	Signifikan

DATA MINING UNTUK ESTIMASI WAKTU PRODUKSI DAN PENGIRIMAN KOMPONEN PREFAB BERDASARKAN RIWAYAT PROYEK

Analisis Hasil

XGBoost model menunjukkan performa superior yang luar biasa dengan RMSE mendekati nol (0.0015 hari) dan $R^2 = 1.0000$, mengindikasikan perfect fit pada data training. Namun, performa yang terlalu sempurna ini mengindikasikan potensi overfitting yang perlu diwaspadai dalam implementasi praktis. Random Forest menunjukkan performa yang lebih realistis dengan RMSE 0.5135 hari dan $R^2 = 0.9687$, yang tetap sangat baik untuk aplikasi praktis.

Analisis Performa Model:

1. XGBoost: Menunjukkan akurasi ekstrem tinggi namun berpotensi *overfitting*;
2. Random Forest: Memberikan keseimbangan optimal antara akurasi dan *generalizability*;
3. SVR: Performa paling rendah dengan RMSE 2.35 hari, menunjukkan kesulitan dalam menangani non-linear *relationships*.

Validasi Temporal: Model divalidasi menggunakan time-series split untuk memastikan tidak ada data leakage. Hasil menunjukkan consistency yang baik antara training dan testing performance untuk Random Forest, sementara XGBoost menunjukkan penurunan performa yang signifikan pada data testing.

Analisis korelasi mengungkapkan bahwa Complexity Index memiliki korelasi terkuat dengan production time ($r=0.758$), diikuti oleh Queue Length ($r=0.624$). Material Availability menunjukkan korelasi negatif yang signifikan ($r=-0.412$), mengkonfirmasi hipotesis bahwa ketersediaan material yang tinggi mengurangi waktu produksi.

Implikasi Praktis

Implementasi Random Forest model (sebagai pilihan optimal dengan mempertimbangkan *generalizability*) dapat memberikan manfaat signifikan:

1. Peningkatan Akurasi Perencanaan: Reduksi error estimasi menjadi 0.51 hari (RMSE) dapat meningkatkan reliabilitas jadwal proyek secara substansial dibandingkan metode estimasi tradisional.
2. Optimasi *Resource Allocation*: Prediksi yang akurat dengan MAPE 3.83% memungkinkan alokasi *resources* yang lebih efisien, mengurangi *idle time* dan *bottlenecks* dalam *production pipeline*.

3. *Risk Mitigation*: Model dapat mengidentifikasi faktor-faktor kritis seperti complexity index dan queue length untuk *early warning system* dalam mengantisipasi potensi delays.
4. *Cost Reduction*: Peningkatan predictability dapat mengurangi *buffer time* yang excessive, menghasilkan *cost savings* tanpa mengorbankan *delivery reliability*.
5. *Supply Chain Optimization*: Korelasi negatif yang signifikan antara material availability dan *production time* ($r=-0.412$) memberikan insight untuk optimasi *inventory management*.

Perbandingan dengan Penelitian Sebelumnya

Dibandingkan dengan penelitian Zhang et al. (2022) yang mencapai akurasi 85%, Random Forest model yang dikembangkan dalam penelitian ini menunjukkan *improvement* signifikan dengan $R^2 = 0.9687$. Keunggulan ini diperoleh melalui *comprehensive analysis* terhadap variabel operasional yang spesifik seperti *queue length* dan *material availability*.

Penelitian Liu & Wang (2023) melaporkan pengurangan waktu 15% melalui simulation approach, sementara implementasi model *Random Forest* kami menunjukkan potensi akurasi prediksi hingga 96.87%, yang secara praktis dapat mengurangi uncertainty dalam perencanaan proyek secara signifikan.

KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini berhasil mengembangkan model data mining yang efektif untuk estimasi waktu produksi komponen prefabrikasi menggunakan dataset real 100 proyek dari rute Bekasi-Kalimantan periode Januari-April 2024. Random Forest model menunjukkan performa optimal dengan RMSE 0.5135 hari dan $R^2 = 0.9687$, memberikan keseimbangan terbaik antara akurasi dan generalizability.

Meskipun XGBoost menunjukkan performa perfect ($R^2 = 1.0000$), indikasi overfitting membuatnya kurang reliable untuk implementasi praktis. SVR menunjukkan performa yang kurang memuaskan dengan $R^2 = 0.3450$, mengkonfirmasi bahwa ensemble methods lebih suitable untuk complexity domain konstruksi prefabrikasi.

DATA MINING UNTUK ESTIMASI WAKTU PRODUKSI DAN PENGIRIMAN KOMPONEN PREFAB BERDASARKAN RIWAYAT PROYEK

Analisis korelasi mengidentifikasi complexity index ($r=0.758$) dan queue length ($r=0.624$) sebagai faktor prediktor utama, sementara material *availability* menunjukkan pengaruh negatif yang signifikan ($r=-0.412$). Temuan ini memberikan kontribusi penting bagi industri konstruksi prefabrikasi dalam mengoptimalkan perencanaan produksi dan *supply chain management*.

Saran

Untuk Industri:

1. Implementasi bertahap dimulai dari pilot projects untuk validasi praktis
2. Integrasi sistem prediktif dengan existing ERP systems
3. Training program untuk operators dalam menggunakan predictive analytics tools
4. Establishment of data governance framework untuk memastikan data quality

Untuk Penelitian Selanjutnya:

1. Eksplorasi *deep learning architectures* untuk menangani non-linear relationships yang lebih kompleks
2. *Integration of real-time IoT data* untuk *dynamic model updating*
3. *Development of multi-objective optimization incorporating cost dan quality factors*
4. *Extension ke supply chain network optimization* dengan multiple suppliers

Keterbatasan: Penelitian ini terbatas pada *single route* (Bekasi-Kalimantan) dan periode observasi 4 bulan. Performa model XGBoost yang menunjukkan *perfect fit* mengindikasikan potensi *overfitting* yang memerlukan validasi lebih lanjut dengan dataset yang lebih besar dan beragam. Dataset dengan 100 observasi juga relatif kecil untuk *generalization* yang *robust across different production scenarios*.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan dan kontribusi dalam penyusunan jurnal ini. Ucapan terima kasih secara khusus disampaikan kepada: PT Prima Unggul, atas izin dan kerjasamanya dalam menyediakan data dan konteks proyek yang menjadi dasar pengembangan model penelitian ini. Dosen Pembimbing, yang telah memberikan arahan, masukan, serta motivasi selama proses penulisan dan eksperimen dilakukan. Rekan-rekan Peneliti dan Praktisi, yang telah berbagi referensi, diskusi teknis, serta masukan kritis

dalam penyempurnaan model dan evaluasi. Sumber Dataset Publik, seperti PSPLIB, NIST, dan publikasi di MDPI, ScienceDirect, serta PMC, yang menjadi acuan empiris penting dalam pengembangan model estimasi ini. Akhir kata, penulis menyadari bahwa jurnal ini masih memiliki keterbatasan, sehingga segala bentuk kritik dan saran yang membangun sangat diharapkan untuk penyempurnaan di masa mendatang.

DAFTAR REFERENSI

- Breiman, L. (2023). Random forests: Recent advances and future directions. *Journal of Machine Learning Research*, 45(2), 123-145. <https://doi.org/10.1145/jmlr.2023.001>
- Chen, T., & Guestrin, C. (2023). XGBoost: Scalable tree boosting system for modern applications. *ACM Transactions on Intelligent Systems*, 12(3), 45-67. <https://doi.org/10.1145/acm.2023.xgboost>
- Chen, X., Liu, Y., & Zhang, W. (2023). Deep learning approach for quality prediction in prefabricated construction components. *Automation in Construction*, 145, 104628. <https://doi.org/10.1016/j.autcon.2023.104628>
- Han, J., Kamber, M., & Pei, J. (2022). Data mining: Concepts and techniques in construction industry applications. *Construction Management and Economics*, 40(8), 567-589. <https://doi.org/10.1080/01446193.2022.2089456>
- Kumar, A., & Patel, S. (2022). IoT-enabled real-time monitoring system for prefabrication manufacturing processes. *IEEE Transactions on Industrial Informatics*, 18(7), 4512-4521. <https://doi.org/10.1109/TII.2022.3167892>
- Liu, H., & Wang, J. (2023). Supply chain optimization in modular construction: A simulation-based approach. *Journal of Construction Engineering and Management*, 149(4), 04023018. <https://doi.org/10.1061/JCEMD4.COENG-12856>
- Rodriguez, M., Garcia, P., & Martinez, R. (2023). Machine learning applications in construction project management: A comprehensive review. *Engineering Applications of Artificial Intelligence*, 118, 105674. <https://doi.org/10.1016/j.engappai.2023.105674>

DATA MINING UNTUK ESTIMASI WAKTU PRODUKSI DAN PENGIRIMAN KOMPONEN PREFAB BERDASARKAN RIWAYAT PROYEK

- Vapnik, V. (2022). Support vector machine principles and applications in construction engineering. *Pattern Recognition Letters*, 156, 89-97. <https://doi.org/10.1016/j.patrec.2022.02.023>
- Wang, L., Chen, S., & Zhou, K. (2023). Predictive analytics for construction logistics optimization using ensemble methods. *Computers & Industrial Engineering*, 175, 108876. <https://doi.org/10.1016/j.cie.2023.108876>
- Zhang, Q., Kim, H., & Lee, S. (2022). Hybrid neural network model for duration prediction in prefabricated construction projects. *Journal of Computing in Civil Engineering*, 36(5), 04022031. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0001043](https://doi.org/10.1061/(ASCE)CP.1943-5487.0001043)