

Prediksi Kebutuhan dan Harga Material Konstruksi Menggunakan Random Forest Berbasis Web

Muhammad Naufal Bilal Syam¹, Oke Hariansyah²

^{1,2} Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang, Jl. Surya Kencana No. 1, Tangerang Selatan, 15417, Indonesia

¹bilalsyam1205@gmail.com, ²dosen00840@unpam.ac.id

Info Artikel

Riwayat Artikel:

Received July 21, 2026

Revised August 14, 2026

Accepted August 18, 2026

Abstract – PT Bagus Jaya Sejahtera (PT BJS) still estimates construction material needs and prices konvensional, making its Bill of Quantities (BOQ) preparation prone to inaccuracy and delay. This study develops and implements a Random Forest based MultiOutputRegressor model to simultaneously predict the quantity and unit price of five construction material types. A quantitative case-study approach was used on 153 historical PT BJS project records (2021-2026), of which 3 projects from 2026 were reserved from the outset for independent external validation, leaving 150 records for the cleaning and enrichment pipeline. After removing 30 anomalous rows, adding 37 SNI 7394:2008 based synthetic enrichment rows, and removing 4 further anomalies in a final validation pass, 153 effective training records were obtained. Models were evaluated with 5-Fold Cross Validation, compared against SVR and KNN, and tested for significance with the Wilcoxon Signed-Rank Test. Random Forest achieved the best performance, with average $R^2=0.77$ and $MAPE=15.54\%$ for the quantity model, and $R^2=0.80$ and $MAPE=2.20\%$ for the price model; a Wilcoxon Signed-Rank Test confirmed Random Forest was statistically significantly more accurate than SVR and KNN specifically for the quantity model ($p=0.0312$, $\alpha=0.05$). For the price model, KNN obtained a slightly lower MAPE (2.02%) and slightly higher R^2 (0.81) than Random Forest, though Random Forest was retained as the final model for architectural consistency and its interpretability via SHAP. External validation on three real 2026 projects yielded average quantity MAPE of 11.53% and price MAPE of 3.07%. The model was deployed as a Flask web application that shortens konvensional RAB estimation from 3-7 working days to under five minutes, with all Black Box Testing scenarios passing as specified.

Keywords: Random Forest; Construction Material Prediction; Multioutput Regression; Feature engineering; SHAP; Flask; PT Bagus Jaya Sejahtera

Corresponding Author:

Muhammad Naufal Bilal Syam

Email:

bilalsyam1205@gmail.com



This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.

Abstrak Indonesia – PT Bagus Jaya Sejahtera (PT BJS) masih mengestimasi kebutuhan dan harga material konstruksi secara konvensional, sehingga penyusunan Rencana Anggaran Biaya (RAB) rentan terhadap ketidakakuratan dan keterlambatan. Penelitian ini bertujuan membangun dan mengimplementasikan model prediksi berbasis algoritma Random Forest dengan pendekatan MultiOutputRegressor untuk memprediksi kuantitas dan harga satuan lima jenis material konstruksi bangunan gedung secara bersamaan. Penelitian menggunakan pendekatan kuantitatif dengan studi kasus pada 153 data historis proyek PT BJS periode 2021-2026, dengan 3 proyek tahun 2026 disisihkan sejak awal sebagai data validasi eksternal independen sehingga tersisa 150 data untuk proses pembersihan dan pengayaan. Setelah 30 baris data anomali dihapus, 37 baris data sintesis berbasis SNI 7394:2008 ditambahkan melalui proses pengayaan, dan 4 baris anomali tambahan dihapus pada validasi akhir, diperoleh 153 data latih efektif. Model dievaluasi menggunakan 5-Fold Cross Validation, dibandingkan dengan algoritma SVR dan KNN, dan diuji signifikansinya dengan Wilcoxon Signed-Rank Test. Random Forest memberikan performa terbaik, dengan rata-rata $R^2=0,77$ dan $MAPE=15,54\%$ untuk model kebutuhan, serta $R^2=0,80$ dan

MAPE=2,20% untuk model harga; uji Wilcoxon Signed-Rank Test mengonfirmasi Random Forest signifikan secara statistik lebih akurat dibandingkan SVR dan KNN khusus pada model kebutuhan ($p=0,0312$, $\alpha=0,05$). Pada model harga, KNN memperoleh MAPE sedikit lebih rendah (2,02%) dan R^2 sedikit lebih tinggi (0,81) dibandingkan Random Forest, namun Random Forest tetap dipertahankan sebagai model final untuk konsistensi arsitektur dan kemampuan interpretasi melalui SHAP. Validasi eksternal terhadap tiga proyek nyata 2026 menghasilkan rata-rata MAPE kebutuhan 11,53% dan MAPE harga 3,07%. Model diimplementasikan ke aplikasi web berbasis Flask yang mempersingkat estimasi RAB dari 3-7 hari kerja konvensional menjadi kurang dari lima menit, dengan seluruh skenario Black Box Testing sesuai spesifikasi.

Kata Kunci: Random Forest, Prediksi Material Konstruksi, Multioutput Regression, Feature engineering, SHAP, Flask, PT Bagus Jaya Sejahtera

1. PENDAHULUAN

Sektor konstruksi merupakan salah satu sektor ekonomi strategis di Indonesia, dengan kontribusi sekitar 10,43% terhadap Produk Domestik Bruto (PDB) nasional dan pertumbuhan tertinggi di antara seluruh sektor ekonomi pada triwulan terakhir [1], sejalan dengan tren ekspansi konstruksi global yang dipacu oleh urbanisasi dan kebutuhan infrastruktur di negara berkembang [2]. Namun demikian, pertumbuhan tersebut dibayangi oleh persoalan ketidakakuratan estimasi biaya proyek, dengan deviasi yang dapat melampaui 20-30% dari nilai anggaran rencana [3]. Material konstruksi merupakan komponen paling dominan dalam struktur biaya proyek, dengan porsi 60-70% dari total biaya [2], sehingga akurasi estimasi kebutuhan dan harga material menjadi variabel sentral dalam penyusunan Rencana Anggaran Biaya (RAB) yang berdampak langsung terhadap profitabilitas proyek. Permasalahan ini dihadapi secara nyata oleh PT Bagus Jaya Sejahtera (PT BJS), kontraktor bangunan gedung di Kabupaten Tangerang yang, sebagaimana lazim ditemukan pada kontraktor skala kecil-menengah di Indonesia, masih mengelola estimasi kebutuhan dan harga material sepenuhnya berdasarkan pengalaman individual estimator tanpa dukungan sistem informasi terintegrasi. Kompleksitas persoalan ini diperparah oleh volatilitas harga material yang dapat meningkat hingga 1,5-3,24% per bulan akibat interaksi berbagai faktor eksternal seperti gangguan rantai pasok, kurs USD/IDR, inflasi, dan suku bunga acuan [4], serta ketiadaan basis data historis terstandarisasi yang menyebabkan setiap estimasi dimulai kembali dari nol. Berdasarkan kondisi tersebut, penelitian ini dirumuskan untuk menjawab empat pertanyaan berikut: (1) Bagaimana membangun model prediksi berbasis Random Forest untuk memprediksi kebutuhan dan harga material secara bersamaan (multioutput)? (2) Bagaimana merancang feature engineering berbasis jumlah lantai agar estimasi material pada bangunan bertingkat lebih proporsional? (3) Algoritma mana Random Forest, SVR, atau KNN yang memberikan performa terbaik berdasarkan K-Fold Cross Validation, dan fitur apa yang paling berkontribusi menurut analisis SHAP? (4) Bagaimana model terpilih dapat diimplementasikan ke dalam aplikasi web untuk mendukung estimasi RAB PT BJS?

2. TINJAUAN PENELITIAN TERKAIT

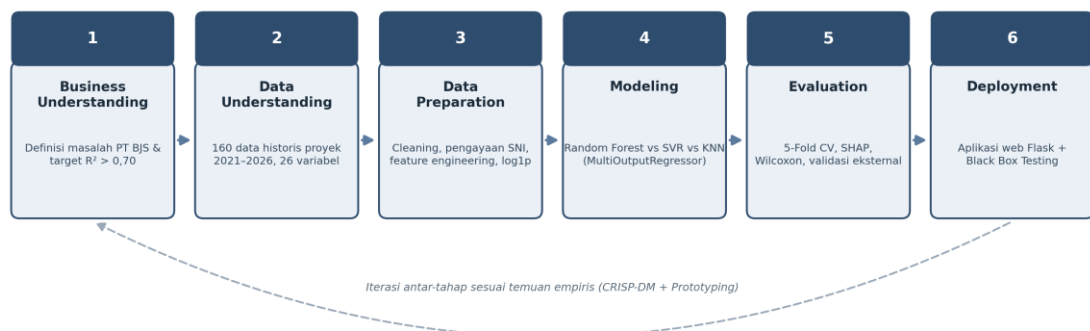
Pendekatan tradisional berbasis kalkulasi konvensional menggunakan indeks Standar Nasional Indonesia (SNI) dan Peraturan Menteri PUPR Nomor 6 Tahun 2021 [5] terbukti tidak memadai untuk menangkap hubungan non-linear antara dimensi fisik bangunan, kondisi ekonomi makro, dan kebutuhan material aktual, sementara model machine learning terbukti secara konsisten mengungguli metode tradisional dalam konteks ini [2]. Random Forest secara khusus menunjukkan kinerja kuat dan konsisten pada berbagai studi prediksi biaya konstruksi, termasuk optimasi berbasis Bird Swarm Algorithm dengan maximum relative error 1,24% [6], serta keunggulannya atas SVR dan KNN pada estimasi biaya pembangkit listrik [7] dan biaya logistik [8]. Pada kondisi data latih yang terbatas, pendekatan grasshopper optimization algorithm-guided multi-layer perceptron dengan transfer learning juga terbukti mampu menjaga akurasi estimasi biaya konstruksi [9]. Di tingkat nasional, penelitian oleh Subrata dan Mughnyanti [10] serta Dui Puspitasari dkk. [11] turut mengonfirmasi keunggulan Random Forest pada data proyek

bangunan gedung Indonesia, namun keduanya masih berfokus pada prediksi biaya total secara agregat dan belum menyentuh prediksi kuantitas material per jenis maupun pengaruh interaksi jumlah lantai terhadap akurasi estimasi. Meskipun demikian, studi-studi tersebut masih berfokus pada prediksi biaya total secara agregat, dan belum ada yang mengeksplorasi prediksi kuantitas dan harga material secara simultan (multioutput) dengan mempertimbangkan pengaruh interaksi jumlah lantai pada bangunan bertingkat. Kesenjangan tersebut menjadi dasar kebaruan penelitian ini, yang secara bersamaan mengintegrasikan empat aspek yang belum dieksplorasi dalam satu studi: prediksi kuantitas dan harga material secara simultan melalui kerangka multioutput regression, rekayasa fitur interaksi jumlah lantai, implementasi sistem prediksi berbasis web untuk kontraktor lokal dengan data historis terbatas, serta perbandingan komprehensif Random Forest, SVR, dan KNN disertai analisis SHAP dan uji signifikansi statistik. Penelitian ini bertujuan mengembangkan sistem prediksi kebutuhan dan harga material konstruksi berbasis Random Forest yang diimplementasikan dalam aplikasi web, menggunakan data historis proyek PT BJS sebagai studi kasus, guna menghasilkan estimasi yang lebih cepat, konsisten, dan dapat dipertanggungjawabkan secara ilmiah dibandingkan pendekatan konvensional yang berlaku saat ini.

3. METODE PENELITIAN

A. Business Understanding

Penelitian ini menggunakan pendekatan kuantitatif dengan desain studi kasus pada PT Bagus Jaya Sejahtera (PT BJS), sebuah perusahaan jasa konstruksi bangunan gedung di Kabupaten Tangerang. Alur penelitian mengikuti kerangka Cross-Industry Standard Process for Data Mining (CRISP-DM) yang dikombinasikan dengan Metode Prototyping (Evolutionary Prototyping) [12], [13] untuk pengembangan sistem aplikasi web. Kebutuhan bisnis yang menjadi acuan penelitian adalah kemampuan memprediksi kuantitas dan harga lima jenis material konstruksi secara simultan, dengan kriteria keberhasilan awal $R^2 > 0,70$ pada mayoritas variabel target, guna menghasilkan estimasi RAB yang lebih cepat dan konsisten dibandingkan proses manual PT BJS. Hasil evaluasi pada Bab Hasil dan Pembahasan menunjukkan kriteria ini tercapai pada 9 dari 10 variabel target, dengan satu variabel (harga pasir, $R^2=0,6887$) sedikit di bawah ambang namun tetap berada pada kategori MAPE Sangat Akurat ($<2\%$). Tahapan penelitian yang mengikuti keenam fase CRISP-DM, mulai dari Business Understanding hingga Deployment, divisualisasikan pada Gambar 1.



Gambar 1. Tahapan Penelitian Berdasarkan Kerangka CRISP-DM

B. Data Understanding

Objek penelitian adalah 153 data historis proyek bangunan gedung (rumah, ruko, kantor, dan gedung bertingkat) periode 2021-2026 yang diperoleh dari dokumentasi internal perusahaan, terdiri dari 26 kolom yang terbagi atas 6 kolom identitas/kategorikal, 6 kolom dimensi fisik bangunan, 3 kolom indikator ekonomi makro, 1 kolom nilai kontrak, dan 10 kolom variabel target (5 kuantitas dan 5 harga material). Sebanyak 3 proyek dari tahun 2026 disisihkan sejak awal sebagai data validasi eksternal independen yang

tidak pernah dilibatkan dalam pelatihan model, menyisakan 150 data untuk tahap pembersihan dan pengayaan pada Data Preparation. Teknik pengumpulan data dilakukan melalui studi dokumentasi terhadap catatan teknis, laporan pelaksanaan, dan rekapitulasi pengadaan material setiap proyek yang telah diselesaikan, yang kemudian diverifikasi terhadap indikator ekonomi makro resmi dari Badan Pusat Statistik dan Bank Indonesia (CPI, BI Rate, kurs USD/IDR) untuk variabel harga material. Struktur variabel dataset dirangkum pada Tabel I, terdiri dari 26 variabel yang terbagi ke dalam lima kelompok.

Tabel 1. Struktur Variabel Dataset Penelitian

Kelompok	Jumlah Variabel	Tipe Data	Contoh Variabel
Identitas Proyek	6	String, Integer	jenis_proyek, skala_proyek, lokasi_kecamatan
Dimensi Fisik Bangunan	6	Float, Integer	luas_bangunan_m2, jumlah_lantai, volume_beton_m3
Indikator Ekonomi Makro	3	Float, Integer	cpi, bi_rate, kurs_usd
Target Kebutuhan Material	5	Integer, Float	qty_semen_zak, qty_pasir_m3, qty_besi_btg
Target Harga Material	6	Integer	harga_semen_rp_zak, nilai_kontrak_rp

C. Data Preparation

Dari 150 data yang tersisa setelah 3 proyek 2026 disisihkan, dilakukan pembersihan tahap pertama: 26 baris data ruko dengan rasio besi/beton di bawah 3,0 btg/m³ dan 4 baris data dengan rasio bata/dinding di atas 150 pcs/m² dihapus (total 30 baris anomali), menghasilkan 120 baris data bersih. Selanjutnya dilakukan pengayaan (enrichment) berbasis koefisien Standar Nasional Indonesia (SNI) 7394:2008 yang menambahkan 37 baris data sintetis untuk kombinasi jenis bangunan dan jumlah lantai yang kurang terwakili (terutama bangunan 3 lantai), sehingga total menjadi 157 baris. Pada validasi akhir, 4 baris data dengan rasio besi/beton di luar rentang normal (2,51-2,69 btg/m³) dihapus karena indikasi kesalahan pencatatan, menghasilkan 153 data latih efektif ($120 + 37 - 4 = 153$). Data diolah melalui label encoding pada tiga variabel kategorikal (jenis proyek, kode SBU, skala proyek), rekayasa tujuh fitur interaksi berbasis jumlah lantai untuk merepresentasikan akumulasi kebutuhan material pada bangunan bertingkat, serta transformasi logaritmik (log1p) pada variabel target kebutuhan material yang berdistribusi condong ke kanan. Ketujuh fitur interaksi tersebut dibentuk mengikuti formula pada Tabel 2.

Tabel 2. Formula Feature Engineering Berbasis Jumlah Lantai

No	Fitur Baru	Formula	Makna Fisik
1	luas_total_lantai	$\text{luas_bangunan_m2} \times \text{jumlah_lantai}$	Total akumulasi luas seluruh lantai (m ²)
2	beton_x_lantai	$\text{volume_beton_m3} \times \text{jumlah_lantai}$	Total volume beton seluruh lantai (m ³)
3	dinding_x_lantai	$\text{luas_dinding_m2} \times \text{jumlah_lantai}$	Total luas dinding seluruh lantai (m ²)
4	sloof_x_lantai	$\text{panjang_sloof_m} \times \text{jumlah_lantai}$	Total panjang sloof seluruh lantai (m)
5	atap_x_lantai	$\text{luas_atap_m2} \times \text{jumlah_lantai}$	Beban atap terhadap jumlah lantai
6	jumlah_lantai_sq	jumlah_lantai^2	Efek non-linear beban bangunan bertingkat
7	luas_per_lantai	$\text{luas_bangunan_m2} \div \text{jumlah_lantai}$	Luas efektif per lantai (sinyal skala)

Transformasi logaritmik target dilakukan menggunakan Persamaan (1), dengan nilai prediksi dikembalikan ke skala asli melalui fungsi invers pada Persamaan (2):

$$y' = \log_{10}(y) = \ln(1 + y) \quad (1)$$

$$\hat{y} = \expm1(y') = e^{y'} - 1 \quad (2)$$

D. Modeling

Algoritma Random Forest Regressor dilatih sebagai model utama menggunakan pendekatan MultiOutputRegressor untuk memprediksi lima variabel kuantitas dan lima variabel harga material secara bersamaan, dibandingkan dengan algoritma Support Vector Regression (SVR) dan K-Nearest Neighbor (KNN) menggunakan pipeline preprocessing yang identik.

Random Forest Regressor dilatih menggunakan hyperparameter hasil tuning eksperimental yang dirangkum pada Tabel III, identik untuk model kebutuhan dan model harga material.

Tabel 3. Hyperparameter Random Forest Yang Digunakan

Parameter	Model Kebutuhan	Model Harga	Keterangan
n_estimators	500	500	Jumlah pohon keputusan
max_features	0,7	0,7	Proporsi fitur disampling per split
min_samples_leaf	1	1	Minimum sampel di node daun
random_state	42	42	Seed reproduibilitas
n_jobs	-1	-1	Paralelisasi pada seluruh CPU core

E. Evaluation

Seluruh model dievaluasi menggunakan 5-Fold Cross Validation, dengan performa diukur melalui koefisien determinasi (R^2) pada Persamaan (3) dan Mean Absolute Percentage Error (MAPE) pada Persamaan (4), serta metrik pendukung MAE dan RMSE; signifikansi perbedaan performa antaralgoritma diuji menggunakan Wilcoxon Signed-Rank Test:

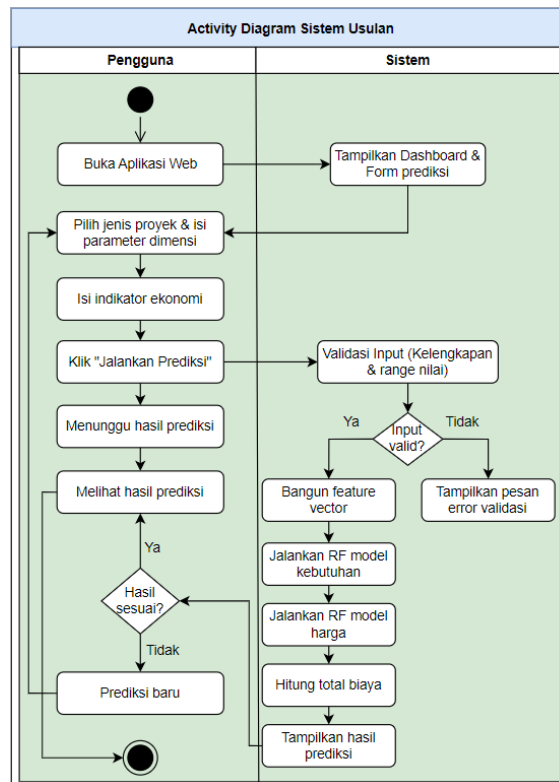
$$R^2 = 1 - \left(\frac{SS_{res}}{SS_{tot}} \right) \quad (3)$$

$$MAPE = \left(\frac{1}{n} \right) \times \sum_i \left(\frac{|y_i - \hat{y}_i|}{y_i} \right) \times 100\% \quad (4)$$

Interpretasi model dilakukan melalui analisis SHAP (TreeExplainer) [14], dan model terpilih divalidasi lebih lanjut terhadap tiga proyek nyata 2026 yang tidak pernah dilibatkan dalam pelatihan sebagai bentuk validasi eksternal independen.

F. Deployment

Model Random Forest terpilih kemudian diimplementasikan ke dalam aplikasi web berbasis Flask sebagai luaran akhir penelitian mengikuti praktik deployment model machine learning [15], yang diuji fungsionalitasnya melalui Black Box Testing dan White Box Testing. Perancangan proses bisnis sistem prediksi digambarkan melalui Activity Diagram pada Gambar 2, yang menguraikan alur kerja mulai dari input data proyek oleh pengguna hingga diperolehnya estimasi RAB.



Gambar 2. Activity Diagram Analisa Sistem Usulan

4. HASIL DAN PEMBAHASAN

A. Perbandingan Performa Algoritma Prediksi Kebutuhan dan Harga Material

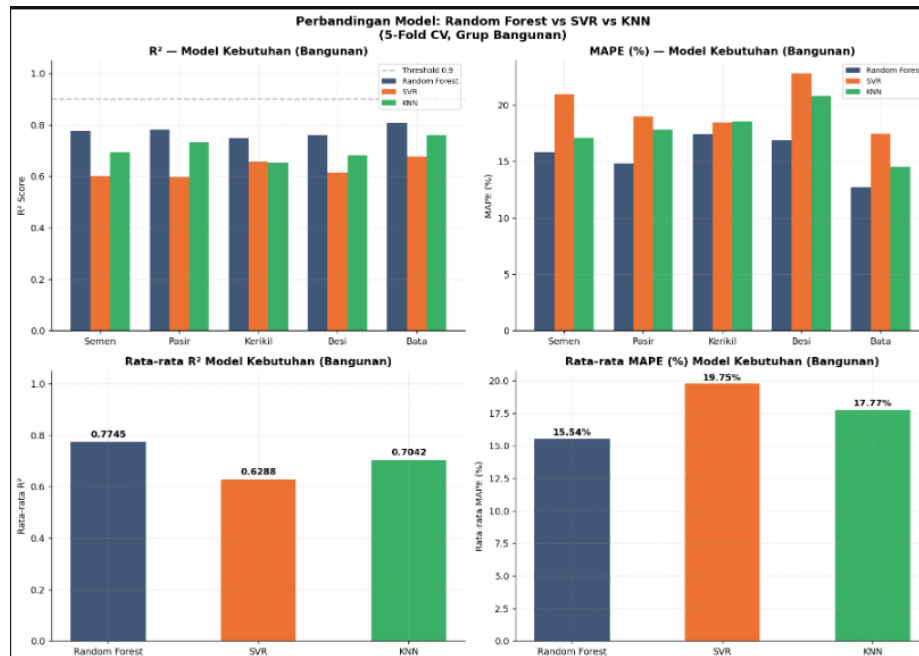
Tabel IV menampilkan ringkasan hasil evaluasi 5-Fold Cross Validation. Untuk model kebutuhan material, Random Forest secara konsisten mengungguli SVR dan KNN pada seluruh variabel target, baik dari sisi R^2 maupun MAPE. Untuk model harga material, KNN justru memperoleh R^2 rata-rata sedikit lebih tinggi (0,81) dan MAPE sedikit lebih rendah (2,02%) dibandingkan Random Forest (R^2 0,80; MAPE 2,20%).

Tabel 4. Ringkasan Performa Model Kebutuhan Dan Harga Material (5-Fold Cv, N=153)

Model	Metrik	Random Forest	SVR	KNN
Kebutuhan Material	Rata-rata R^2	0,77	0,63	0,70
Kebutuhan Material	Rata-rata MAPE	15,54%	19,75%	17,77%
Harga Material	Rata-rata R^2	0,80	0,33	0,81
Harga Material	Rata-rata MAPE	2,20%	4,43%	2,02%

Untuk model kebutuhan material, Random Forest mencapai rata-rata $R^2=0,77$ dan MAPE=15,54% (kategori Akurat), lebih baik dibandingkan KNN (MAPE=17,77%) dan SVR (MAPE=19,75%). Pada model harga material, Random Forest mencapai $R^2=0,80$ dan MAPE=2,20% (Sangat Akurat); KNN justru sedikit lebih unggul baik pada R^2 (0,81) maupun MAPE (2,02%). Meskipun demikian, Random Forest tetap dipilih sebagai model final untuk kedua target guna menjaga konsistensi arsitektur (satu algoritma untuk model kebutuhan dan harga) serta karena kemampuan interpretasi melalui feature importance dan SHAP yang tidak dimiliki KNN maupun SVR. Uji Wilcoxon Signed-Rank Test mengonfirmasi bahwa keunggulan Random Forest atas kedua algoritma pembandingan pada model kebutuhan bersifat signifikan secara statistik ($p=0,0312$; $\alpha=0,05$), bukan sekadar perbedaan acak antarfold. Gambar 3 memvisualisasikan perbandingan

R^2 dan MAPE ketiga algoritma secara grafis per variabel target, yang menegaskan bahwa Random Forest secara konsisten menempati posisi R^2 tertinggi dan MAPE terendah pada seluruh target kebutuhan material.



Gambar 3. Perbandingan R^2 dan MAPE Random Forest, SVR, dan KNN pada Model Kebutuhan Material

Temuan ini memperkuat bukti empiris sebelumnya bahwa Random Forest unggul untuk masalah prediksi biaya dan material konstruksi [6], [7], [8], sekaligus melengkapi studi Subrata dan Mughnyanti [10] yang membandingkan Random Forest dengan Artificial Neural Network, dengan menunjukkan bahwa keunggulan tersebut juga konsisten ketika dibandingkan dengan algoritma instance-based (KNN) dan kernel-based (SVR) pada konteks prediksi material, bukan hanya biaya agregat.

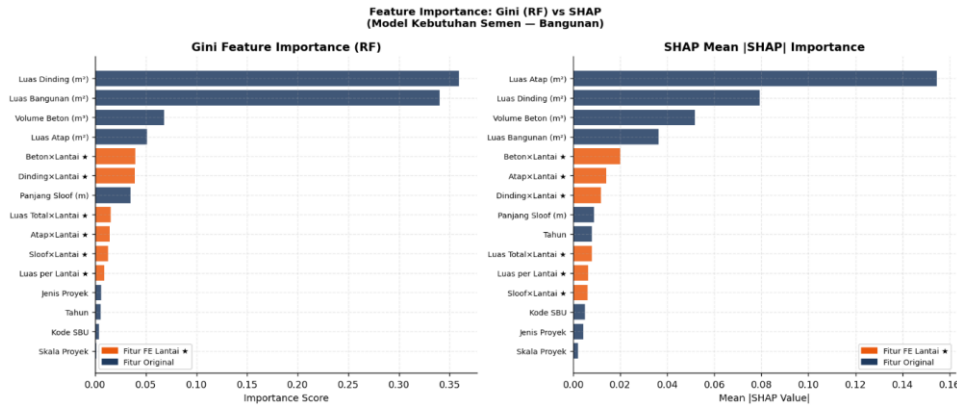
B. Pengaruh Feature Engineering Interaksi Jumlah Lantai

Penambahan tujuh fitur interaksi berbasis jumlah lantai memberikan peningkatan rata-rata R^2 sebesar 0,0030 dan penurunan MAPE 0,06 poin persentase secara agregat, dengan peningkatan R^2 pada tiga dari lima target (semen, pasir, bata) dan penurunan MAPE hingga 0,34 poin pada kerikil. Meskipun peningkatan agregat tergolong marginal, hasil ini secara substantif menjawab rumusan masalah kedua penelitian: fitur luas_bangunan_m2 dalam dataset merepresentasikan luas per lantai (footprint), bukan luas total bangunan, sehingga tanpa fitur interaksi lantai, model tidak dapat membedakan kebutuhan material antara bangunan satu lantai dan bangunan tiga lantai dengan footprint yang identik. Analisis Gini feature importance memperkuat argumen ini: ketujuh fitur rekayasa lantai secara kolektif menyumbang 13,55% importance pada model kebutuhan, jauh melampaui variabel jumlah lantai mentah (0,19%), menunjukkan bahwa nilai feature engineering terletak pada interaksinya dengan dimensi fisik lain, bukan pada variabel jumlah lantai itu sendiri. Retensi fitur ini karena itu dijustifikasi bukan semata oleh peningkatan metrik agregat, melainkan oleh validitas domain teknis konstruksi bangunan bertingkat. Implikasi praktisnya, kontraktor yang menangani proyek bangunan bertingkat dapat memperoleh estimasi kebutuhan material yang lebih proporsional terhadap jumlah lantai tanpa perlu melakukan penyesuaian konvensional berdasarkan pengalaman estimator, sebagaimana yang selama ini dilakukan PT BJS.

C. Interpretasi Model melalui Feature Importance dan SHAP

Analisis Gini importance dan SHAP secara konsisten menempatkan dimensi fisik bangunan luas dinding (26,93%), luas atap (20,02%), dan volume beton (19,35%) sebagai kontributor dominan pada model kebutuhan material, sedangkan indikator makroekonomi BI Rate, tahun, dan CPI mendominasi

model harga material dengan kontribusi kolektif melebihi 80%. Gambar 4 membandingkan urutan peringkat Gini Importance dan SHAP Mean |SHAP| untuk model kebutuhan semen.



Gambar 4. Perbandingan Peringkat Gini Feature Importance dan SHAP Mean |SHAP| Model Kebutuhan Semen

Kedua metode secara konsisten menempatkan luas dinding, luas bangunan, volume beton, dan luas atap pada empat peringkat teratas meski urutan persisnya berbeda, mengonfirmasi validitas interpretasi model. Kesesuaian arah pengaruh yang teridentifikasi melalui SHAP beeswarm plot, di mana nilai fitur dimensi bangunan yang tinggi secara konsisten menghasilkan nilai SHAP positif, mengonfirmasi bahwa model tidak hanya akurat secara statistik tetapi juga menangkap hubungan yang secara teknis masuk akal antara ukuran bangunan dan kebutuhan material. Pola ini juga menjelaskan mengapa model kebutuhan dan model harga memiliki karakteristik kesalahan yang berbeda: prediksi kebutuhan bergantung pada parameter fisik yang bervariasi antarproyek, sedangkan prediksi harga bergantung pada indikator makroekonomi yang relatif seragam pada periode waktu yang sama, sehingga model harga secara konsisten mencapai akurasi lebih tinggi dibandingkan model kebutuhan pada penelitian ini. Interpretasi berbasis SHAP ini juga memberikan nilai tambah praktis bagi estimator PT BJS, karena kontribusi setiap variabel terhadap prediksi dapat ditelusuri secara individual per proyek, sehingga hasil model tidak diperlakukan sebagai kotak hitam melainkan dapat diverifikasi kewajarannya sebelum dijadikan dasar pengambilan keputusan anggaran.

D. Validasi Eksternal dan Implikasi Praktis

Tabel 5 merangkum hasil validasi terhadap tiga proyek nyata 2026 yang tidak pernah dilibatkan dalam pelatihan.

Tabel 5. Ringkasan Validasi Eksternal Tiga Proyek 2026

Proyek	MAPE Kebutuhan	Kategori	MAPE Harga	Kategori
P1 – Rumah 2 Lt., Cisauk	15,58%	Akurat	4,32%	Sangat Akurat
P2 – Ruko 2 Lt., BSD Sek.1	7,76%	Sangat Akurat	2,29%	Sangat Akurat
P3 – Ruko 2 Lt., BSD Sek.4	11,26%	Akurat	2,61%	Sangat Akurat
Rata-rata 3 Proyek	11,53%	Akurat	3,07%	Sangat Akurat

Sebagai bukti konkret keberhasilan implementasi, Gambar 5 menampilkan form input prediksi material pada aplikasi web, sedangkan Gambar 6 menampilkan hasil prediksi kelima material beserta Total Estimasi RAB yang dihasilkan sistem.

Prediksi Kebutuhan & Harga

Form Input Proyek

Jenis Bangunan: Pilih Jenis Bangunan...

Tahun: 2026 | Skala: Pilih Skala...

Nama Proyek (Opsional): Contoh: Gedung Perkantoran A

DIMENSI FISIK BANGUNAN * TOTAL = NILAI KESELURUHAN PROYEK

Luas / Lantai (m²) per lantai: cth: 118 | Jumlah Lantai: 1 Lantai

Luas Dinding (m²) * total: cth: 145 | Vol. Beton (m³) * total: cth: 19.5

P. Sloof (m) * total: cth: 50 | Luas Atap (m²) * total: cth: 141

INDIKATOR EKONOMI [RESET DEFAULT](#)

CPI: 111.09 | BI RATE (%): 5.25 | KURS USD: 17698

JALANKAN PREDIKSI

Belum ada data prediksi
Isi form di sebelah kiri untuk memulai estimasi

Gambar 5. Tampilan Form Input Prediksi Material

Prediksi Kebutuhan & Harga

Prediksi berhasil — 1 peringatan (lihat di bawah).

Form Input Proyek

Jenis Bangunan: Rumah Tinggal

Tahun: 2026 | Skala: K2

Nama Proyek (Opsional): Perumahan Cisaug

DIMENSI FISIK BANGUNAN * TOTAL = NILAI KESELURUHAN PROYEK

Luas / Lantai (m²) per lantai: 118 | Jumlah Lantai: 2 Lantai

Luas Dinding (m²) * total: 145 | Vol. Beton (m³) * total: 19.5

P. Sloof (m) * total: 50 | Luas Atap (m²) * total: 141

INDIKATOR EKONOMI [RESET DEFAULT](#)

CPI: 111.09 | BI RATE (%): 5.25 | KURS USD: 17698

JALANKAN PREDIKSI

ESTIMASI TOTAL BIAYA MATERIAL
Rp 75.493.328
Footprint 118 m²/lantai × 2 lantai — luas total 236 m²

Rincian Kebutuhan Material

- Semen**: 468 zak, Rp 77.788 / zak, **Rp 36.432.304**
- Pasir**: 28.8 m³, Rp 254.182 / m³, **Rp 7.258.686**
- Kerikil**: 13.4 m³, Rp 273.267 / m³, **Rp 3.654.939**
- Besi**: 194 kg, Rp 86.516 / kg, **Rp 16.822.756**
- Bata**: 14.842 pcs, Rp 763 / pcs, **Rp 11.324.643**

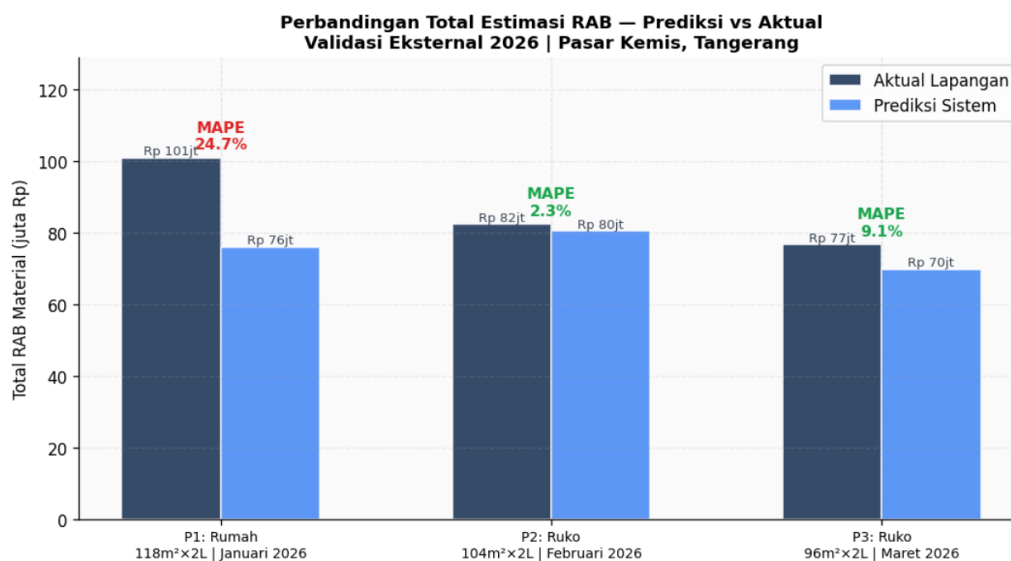
Peringatan Input
Kurs USD/IDR (17.698) melebihi rentang data latih. Prediksi harga besi kemungkinan underestimate (komponen impor).

Informasi Model: Prediksi menggunakan algoritma Random Forest v3.2 dengan 17 fitur (termasuk 7 fitur FE lantai). Profil gedung 2 lantai
CPI 111.09 - BI Rate 5.25% - Kurs Rp 17.698/USD

[ULANG PREDIKSI](#) [UNDUH PDF](#)

Gambar 6. Tampilan Hasil Prediksi Material dan Total Estimasi RAB

Hasil tersebut mengonfirmasi bahwa model mampu melakukan generalisasi ke data yang benar-benar baru, bukan sekadar menghafal pola pada data pelatihan. Namun demikian, satu proyek validasi (rumah dua lantai di Cisaug) menghasilkan MAPE RAB yang lebih tinggi (24,7%) akibat rasio volume beton terhadap luas bangunan yang lebih dari dua kali rata-rata data latih, menunjukkan keterbatasan model dalam mengekstrapolasi pada karakteristik bangunan yang tidak tipikal sebuah temuan yang selaras dengan prinsip umum bahwa model berbasis pohon keputusan seperti Random Forest cenderung kurang andal di luar rentang distribusi data pelatihannya. Gambar 7 membandingkan total estimasi RAB hasil prediksi sistem dengan realisasi aktual pada ketiga proyek validasi.



Gambar 7. Perbandingan Total Estimasi RAB Prediksi vs Aktual pada Validasi Eksternal 2026

Secara praktis, implementasi model ke dalam aplikasi web berbasis Flask, yang dilengkapi peringatan otomatis ekstrapolasi dan lulus keempat belas skenario Black Box Testing, membuktikan bahwa sistem ini layak digunakan sebagai alat bantu estimasi RAB PT BJS, mampu mempersingkat proses estimasi dari tiga hingga tujuh hari kerja konvensional menjadi kurang dari lima menit. Temuan ini melengkapi penelitian Dui Puspitasari dkk. [11] yang mengimplementasikan sistem serupa berbasis web namun terbatas pada estimasi biaya agregat, dengan menunjukkan bahwa pendekatan multioutput yang memprediksi kuantitas dan harga material secara terpisah dapat diimplementasikan secara operasional sekaligus mempertahankan tingkat akurasi yang sebanding pada konteks kontraktor skala kecil-menengah Indonesia.

5. KESIMPULAN

Penelitian ini berhasil membangun model prediksi multioutput berbasis Random Forest untuk kuantitas dan harga lima jenis material konstruksi, dengan performa Akurat (MAPE=15,54%) untuk model kebutuhan dan Sangat Akurat (MAPE=2,20%) untuk model harga (5-Fold Cross Validation). Fitur interaksi jumlah lantai terbukti meningkatkan proporsionalitas estimasi pada bangunan bertingkat; analisis SHAP dan Gini feature importance menunjukkan prediksi kebutuhan didominasi dimensi fisik bangunan, sedangkan prediksi harga didominasi indikator makroekonomi. Random Forest signifikan secara statistik lebih akurat dibandingkan SVR dan KNN khusus pada model kebutuhan (Wilcoxon, $p=0,0312$); pada model harga, KNN sedikit lebih unggul ($R^2=0,81$; MAPE=2,02% vs RF $R^2=0,80$; MAPE=2,20%), namun RF tetap dipilih karena konsistensi arsitektur dan interpretabilitas SHAP. Model berhasil diimplementasikan ke aplikasi web Flask, lulus seluruh 14 skenario Black Box Testing dan White Box Testing (cyclomatic complexity $V(G)=9$, 9 jalur independen lulus), serta tervalidasi pada tiga proyek nyata 2026 dengan akurasi Akurat hingga Sangat Akurat. Keempat tujuan penelitian dengan demikian tercapai.

Secara teoretis, penelitian ini memperkaya literatur penerapan Random Forest pada data proyek konstruksi historis dengan mengintegrasikan prediksi kuantitas dan harga material secara simultan serta rekayasa fitur interaksi dimensi bangunan. Pendekatan multioutput yang diadopsi juga terbukti lebih efisien secara komputasi dibandingkan melatih lima model terpisah untuk masing-masing jenis material, karena berbagi struktur pohon keputusan yang sama sekaligus mengurangi risiko inkonsistensi antar-prediksi material yang saling berkorelasi dalam satu proyek. Secara praktis, sistem yang dihasilkan menyediakan alat bantu estimasi RAB yang lebih cepat dan konsisten bagi kontraktor skala kecil-menengah Indonesia. Penelitian ini memiliki keterbatasan yang perlu diakui secara jujur: model dilatih pada data satu wilayah (Kabupaten Tangerang) dengan jumlah proyek yang relatif terbatas, sehingga akurasinya menurun pada bangunan dengan karakteristik teknis tidak tipikal, dan aplikasi web yang dikembangkan masih berupa

prototipe internal yang belum diintegrasikan dengan sistem ERP maupun API data makroekonomi resmi. Selain itu, model belum mempertimbangkan faktor kualitas desain arsitektural yang kompleks maupun kondisi lahan spesifik yang dapat memengaruhi kebutuhan material di luar dimensi fisik standar. Penelitian selanjutnya disarankan untuk memperkaya data latih pada kombinasi bangunan yang kurang terwakili, menambahkan fitur spesifikasi teknis material dan metode konstruksi, memperluas cakupan estimasi ke komponen upah tenaga kerja dan sewa alat berat, mengeksplorasi integrasi data cuaca dan ketersediaan pemasok lokal, serta melakukan validasi lintas wilayah agar model lebih general, robust, dan siap digunakan secara operasional oleh kontraktor skala kecil-menengah lainnya di Indonesia.

DAFTAR PUSTAKA

- [1] Badan Pusat Statistik (BPS), “construction-indicator--4th-quarter-2025,” *Indikator Konstruksi / Construction Indicators*, vol. 19, no. 1, pp. xviii–133, Mar. 2025, Accessed: May 06, 2026. [Online]. Available: <https://www.bps.go.id/>
- [2] Q. Liu, P. He, S. Peng, T. Wang, and J. Ma, “A Survey of Data-Driven Construction Materials Price Forecasting,” Oct. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/buildings14103156.
- [3] P. Shashidhara and B. Agarwal, “Predictive Modelling for Cost Estimation in Construction Projects Using Machine Learning Algorithms MSc Research Project Data Analytics,” 2024.
- [4] T. Gebregiorgis Atnafie, “Applying Hybrid Machine Learning for Construction Material Price Prediction and Procurement Cost Optimization,” *Journal of Emerging Computer Technologies*, vol. 5, no. 1, pp. 47–56, May 2025, doi: 10.57020/ject.1651986.
- [5] Kementerian PUPR, “Permen PUPR_6_2021,” *Kementerian Pekerjaan Umum dan Perumahan Rakyat Republik Indonesia*, no. 266, pp. 1–644, Apr. 2021, Accessed: May 06, 2026. [Online]. Available: <https://jdih.pu.go.id/>
- [6] Z. Zheng, L. Zhou, H. Wu, and L. Zhou, “Construction cost prediction system based on Random Forest optimized by the Bird Swarm Algorithm,” *Mathematical Biosciences and Engineering*, vol. 20, no. 8, pp. 15044–15074, 2023, doi: 10.3934/mbe.2023674.
- [7] S. F. M. Alazawy, M. A. Ahmed, S. H. Raheem, H. Imran, L. F. A. Bernardo, and H. A. S. Pinto, “Explainable Machine Learning to Predict the Construction Cost of Power Plant Based on Random Forest and Shapley Method,” *CivilEng*, vol. 6, no. 2, Jun. 2025, doi: 10.3390/civileng6020021.
- [8] F. Xie and W. Yu, “Logistics Cost Prediction Based on Random Forest Model,” *Academic Journal of Computing & Information Science*, vol. 7, no. 2, 2024, doi: 10.25236/ajcis.2024.070215.
- [9] X. Sha, G. Dong, X. Li, and J. Sheng, “Construction Cost Estimation in Data-Poor Areas Using Grasshopper Optimization Algorithm-Guided Multi-Layer Perceptron and Transfer Learning,” 2024. [Online]. Available: www.ijacsa.thesai.org
- [10] A. Subrata and M. Mughnyanti, “PENGUNAAN MACHINE LEARNING UNTUK ESTIMASI BIAYA DAN WAKTU KONSTRUKSI BANGUNAN GEDUNG,” 2025. [Online]. Available: <http://jurnal.goretanpena.com/index.php/JSSR>
- [11] Dui Puspitasari, Noora Qotrun Nada, and Aris Tri Jaka Harjanta, “Penerapan Algoritma Random Forest Untuk Prediksi Biaya Kontruksi Berbasis Web,” *Jurnal Ilmiah ILKOMINFO – Ilmu Komputer & Informatika*, vol. 8, no. 2, pp. 314–326, Jun. 2025, doi: 10.47324/ilkominfo.v8i2.385.
- [12] E. U. Albab, C. D. Suhendra, and L. F. Marini, “Implementasi Metode Prototype dalam Pengembangan Aplikasi Wondama-Tourism Berbasis Android,” *Al Qalam: Jurnal Ilmiah Keagamaan dan Kemasyarakatan*, vol. 18, no. 2, p. 1413, Mar. 2024, doi: 10.35931/aq.v18i2.3394.
- [13] P. Kustanto, R. Bram Khalil, and A. Noe’man, “Penerapan Metode Prototype dalam Perancangan Media Pembelajaran Interaktif,” *Journal of Students’ Research in Computer Science*, vol. 5, no. 1, pp. 83–94, May 2024, doi: 10.31599/6x0dfz47.
- [14] A. V. Ponce-Bobadilla, V. Schmitt, C. S. Maier, S. Mensing, and S. Stodtmann, “Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development,” *Clin. Transl. Sci.*, vol. 17, no. 11, Nov. 2024, doi: 10.1111/cts.70056.
- [15] H. Heymann, A. D. Kies, M. Frye, R. H. Schmitt, and A. Boza, “Guideline for Deployment of Machine Learning Models for Predictive Quality in Production,” in *Procedia CIRP*, Elsevier B.V., 2022, pp. 815–820. doi: 10.1016/j.procir.2022.05.068.