Blend Sains
Jurnal Teknik
https://jurnal.ilmubersama.com/index.php/blendsains

Artikel Penelitian

Model Klasifikasi Persepsi Kebutuhan Penyesuaian AHSP pada Proyek Infrastruktur Skala Kecil Berbasis Machine Learning

Andi Subhan Nur[*], *Muh. Ryan Pratama, Afrinaldy Andara, Ainun Jariah Hilal Anshary, Akbar Indrawan Saudi, Ellyni Dwi Fortuna, Michail Amin, Syamsul Fahri, Muh. Miftahulkhair*

Fakultas Teknik, Teknik Sipil, Universitas Sulawesi Barat, Majene, Indonesia

INFORMASI ARTIKEL

Diterima Redaksi: 04 Oktober 2026
Revisi Akhir: 31 Oktober 2026
Diterbitkan *Online*: 31 Oktober 2026

KATA KUNCI

AHSP
Biaya Konstruksi
Infrastruktur Skala Kecil
Machine Learning
Penyesuaian Lokal

KORESPONDENSI (*)

Phone: +62 813-4165-0300
E-mail: andi.subhannur@unsulbar.ac.id

A B S T R A K

Analisis Harga Satuan Pekerjaan (AHSP) merupakan dasar penting dalam penyusunan biaya konstruksi, sementara proyek infrastruktur skala kecil dapat berlangsung pada kondisi akses, topografi, logistik material, tenaga kerja, metode pelaksanaan, dan peralatan yang beragam. Penelitian ini bertujuan menganalisis persepsi kebutuhan penyesuaian AHSP berdasarkan kondisi lokal proyek serta mengeksplorasi pemanfaatan *machine learning* sebagai alat *screening* awal. Sebanyak 83 respons kuesioner dianalisis menggunakan Q2–Q17 sebagai prediktor dan Q28 sebagai target klasifikasi tiga kelas. Data dibagi secara *stratified* menjadi 58 data latih dan 25 data uji. Representasi fitur kompak menghasilkan 56 fitur dengan rasio p/n 0,966. Logistic Regression, Decision Tree, dan Random Forest dibandingkan menggunakan *repeated stratified 5-fold cross-validation* sebanyak 10 pengulangan. Cronbach's alpha Q22–Q27 sebesar 0,8649. Distribusi target terdiri atas kelas Tinggi 50 responden (60,24%), Sedang 22 (26,51%), dan Rendah 11 (13,25%). Decision Tree memperoleh rerata F1-makro tertinggi pada *cross-validation* sebesar 0,3661 ± 0,1252, sedangkan pada data uji Logistic Regression menghasilkan F1-makro tertinggi 0,5349 dan *balanced accuracy* 0,5460. Random Forest memperoleh akurasi tertinggi 0,6800. Q17 memiliki *permutation importance* tertinggi, sedangkan Q4 menunjukkan sinyal prediktif paling konsisten. Pasangan batu, rabat beton, dan galian tanah menjadi tiga prioritas awal observasi. Model diposisikan sebagai *screening* eksploratif, bukan dasar langsung penetapan koefisien AHSP lokal.

PENDAHULUAN

Estimasi biaya konstruksi dipengaruhi oleh tenaga kerja, bahan, peralatan, metode pelaksanaan, lokasi, dan kondisi logistik. Dalam kerangka nasional, AHSP menempatkan tenaga kerja, bahan, dan peralatan sebagai komponen utama harga satuan serta memungkinkan perhitungan teknis dan analisis produktivitas ketika kebutuhan pekerjaan belum terakomodasi secara memadai [1], [2]. Pedoman *engineer's estimate* juga menekankan perlunya penyesuaian terhadap kondisi proyek, ketersediaan data, dan ketidakpastian biaya [3].

Penelitian terdahulu menunjukkan bahwa lokasi dan logistik merupakan faktor penting dalam biaya dan pelaksanaan konstruksi. Mudita dkk. mengidentifikasi akses menuju proyek, topografi, alat angkut, stok pemasok, dan hubungan pemasok-kontraktor sebagai faktor waktu tunggu material [4] . Optimasi sumber dan distribusi material juga berkaitan langsung dengan biaya pengadaan [5].

Produktivitas tenaga kerja juga tidak seragam antar proyek. Meta-analisis Jian dkk. menunjukkan peran pengalaman dan keterampilan pekerja, supervisi, komunikasi, serta cuaca terhadap produktivitas [6]. Studi keterlambatan selanjutnya menegaskan peran cuaca, topografi, akses material, serta keterbatasan tenaga terampil [7], [8].

https://doi.org/10.56211/blendsains.v5i2.2095 Attribution-ShareAlike 4.0 International Some rights reserved

Secara normatif, SE Direktur Jenderal Bina Konstruksi Nomor 47/SE/Dk/2026 menyediakan acuan penyusunan AHSP, analisis produktivitas, faktor bahan, serta tata cara pengumpulan Harga Satuan Pokok [1]. Lampiran teknisnya menekankan verifikasi, validasi, konsistensi, kewajaran data, serta kondisi lokasi dalam penyusunan biaya. Kerangka tersebut menunjukkan bahwa penyesuaian berbasis kondisi aktual merupakan proses teknis yang harus dibuktikan dengan data lapangan dan tidak dimaksudkan sebagai penggantian sepihak terhadap AHSP nasional [1], [2].

Kuesioner skala *Likert* memungkinkan persepsi responden dikonversi menjadi data kuantitatif [9]. Skala lima poin umum digunakan, tetapi jumlah kategori respons dan kecenderungan memilih kategori tengah dapat memengaruhi bias interpretasi [10]. Konsistensi internal dapat dievaluasi menggunakan *Cronbach's alpha*, namun nilai alpha tidak dengan sendirinya membuktikan unidimensionalitas konstruk [11]. Oleh karena itu, reliabilitas dalam penelitian ini digunakan sebagai pemeriksaan konsistensi butir persepsi, bukan sebagai bukti tunggal validitas konstruk.

Penelitian mengenai estimasi biaya konstruksi berbasis *machine learning* telah berkembang dalam beberapa tahun terakhir. Kajian sistematis menunjukkan bahwa berbagai algoritma *machine learning* telah digunakan untuk mendukung estimasi dan prediksi biaya proyek konstruksi [12] . Studi komparatif juga telah mengevaluasi model linier dan *ensemble* dalam prediksi biaya konstruksi [13], sedangkan pendekatan yang menggabungkan *machine learning* dengan *explainable artificial intelligence* dikembangkan untuk meningkatkan transparansi dan keandalan prediksi biaya [14]. Penelitian Subrata dkk. [15] selanjutnya menerapkan Random Forest Regressor dan Artificial Neural Network (ANN) untuk mengestimasi biaya dan waktu konstruksi bangunan gedung, dengan Random Forest menunjukkan performa yang lebih baik dibandingkan ANN pada kedua target estimasi. Meskipun demikian, sebagian besar penelitian tersebut berfokus pada prediksi biaya, durasi, atau nilai akhir proyek. Pemanfaatan *machine learning* untuk mengidentifikasi pola persepsi kebutuhan penyesuaian AHSP berdasarkan karakteristik responden dan kondisi operasional proyek lokal masih terbatas. Di sisi lain, data mengenai kondisi akses, topografi, logistik material, tenaga kerja, metode pelaksanaan, dan penggunaan peralatan berpotensi memberikan informasi awal mengenai konteks proyek yang memerlukan evaluasi lebih lanjut terhadap kesesuaian AHSP. Oleh karena itu, *research gap* penelitian ini terletak pada integrasi karakteristik operasional proyek skala kecil, persepsi kebutuhan penyesuaian AHSP, evaluasi klasifikasi yang mengendalikan risiko *target leakage* dan dimensionalitas, serta pemeringkatan pekerjaan yang perlu diprioritaskan untuk observasi produktivitas lapangan.

Penelitian ini tidak dimaksudkan untuk menentukan koefisien AHSP lokal secara langsung. *Machine learning* digunakan sebagai alat penapisan *(screening)* eksploratif untuk mengidentifikasi pola yang berkaitan dengan persepsi kebutuhan penyesuaian AHSP dan membantu menentukan pekerjaan yang perlu diprioritaskan untuk pengukuran produktivitas aktual. Penelitian bertujuan untuk: (1) menggambarkan karakteristik responden, wilayah, proyek, dan hambatan pelaksanaan; (2) menganalisis tingkat persepsi kebutuhan penyesuaian AHSP serta konsistensi internal instrumen; dan (3) membandingkan performa model klasifikasi, mengevaluasi sensitivitas model, serta menentukan prioritas item pekerjaan untuk observasi dan kalibrasi lapangan.

METODE PENELITIAN

Desain, Sampel, dan Instrumen

Penelitian ini menggunakan pendekatan kuantitatif deskriptif-prediktif dengan desain survei potong lintang (*cross-sectional survey*). Pengambilan sampel dilakukan menggunakan teknik *nonprobability purposive sampling* dengan mempertimbangkan keterlibatan responden dalam kegiatan perencanaan, pelaksanaan, pengawasan, maupun pengelolaan proyek infrastruktur lokal. Dataset akhir terdiri atas 83 respons valid, yang meliputi 53 respons (63,86%) melalui pengisian mandiri dan 30 respons (36,14%) melalui wawancara oleh peneliti atau surveyor. Responden terdiri atas konsultan pengawas, kontraktor, unsur dinas teknis/PU/PPK/PPTK, pelaksana lapangan, konsultan perencana, serta pemangku kepentingan lain yang memiliki pengalaman dalam proyek infrastruktur lokal. Unit analisis dalam penelitian ini adalah responden dan persepsinya berdasarkan pengalaman dalam proyek, bukan proyek yang diidentifikasi secara unik. Dalam penelitian ini, istilah proyek infrastruktur skala kecil digunakan secara operasional untuk menggambarkan pekerjaan infrastruktur lokal yang dilaporkan responden melalui instrumen penelitian dan memiliki variasi nilai proyek, metode pelaksanaan, penggunaan tenaga kerja, peralatan, serta kondisi akses sebagaimana dikategorikan dalam kuesioner. Istilah tersebut tidak dimaksudkan sebagai klasifikasi regulatif tunggal berdasarkan batas nilai kontrak tertentu.

 https://doi.org/10.56211/blendsains.v5i2.2095

Instrumen penelitian terdiri atas 31 pertanyaan utama yang dikelompokkan berdasarkan fungsi analisis. Q2–Q17 digunakan untuk menggambarkan karakteristik responden dan kondisi operasional proyek, meliputi peran dan wilayah kerja, pengalaman proyek, jenis dan nilai proyek, kondisi lokasi, topografi, kondisi jalan, jarak angkut manual, jenis tanah, faktor wilayah, metode pelaksanaan, sistem dan asal tenaga kerja, penggunaan peralatan, serta kendala lapangan. Q18–Q19 digunakan secara terpisah untuk menyusun pemeringkatan item pekerjaan yang diprioritaskan untuk observasi produktivitas lapangan. Q20–Q21 dianalisis secara deskriptif untuk menggambarkan persepsi kesesuaian AHSP serta faktor yang menurut responden perlu dipertimbangkan dalam penyesuaian. Q22–Q27 digunakan untuk mengevaluasi konsistensi internal konstruk persepsi melalui *Cronbach's alpha*, sedangkan Q28 digunakan secara eksklusif sebagai target klasifikasi. Q29–Q31 digunakan sebagai informasi pendukung dan tidak dimasukkan sebagai prediktor model.

Variabel Q3 yang merepresentasikan wilayah kerja dinormalisasi ke dalam enam kabupaten di Provinsi Sulawesi Barat melalui penyesuaian padanan nama kecamatan dan kabupaten berdasarkan data mentah. Seluruh 83 respons pada Q3 berhasil dipetakan ke enam kabupaten tersebut. Dalam proses *encoding*, pipeline tetap menyediakan kategori teknis *fallback* "luar/belum terpetakan" untuk mengantisipasi nilai yang tidak dikenali pada proses transformasi. Kategori tersebut tidak merepresentasikan tambahan wilayah responden dalam sampel penelitian. Pemisahan fungsi antarbutir instrumen dilakukan sejak tahap analisis untuk mengurangi risiko penggunaan variabel yang secara konseptual terlalu dekat dengan target klasifikasi dan meminimalkan risiko *target leakage*.

Pra-pemrosesan dan Reliabilitas

Pra-pemrosesan data mencakup pemeriksaan kelengkapan dan konsistensi isian, penghapusan informasi identitas, normalisasi kategori, serta pemisahan jawaban multirespons. Tahap tersebut dilakukan agar data yang digunakan dalam analisis memiliki format yang konsisten dan dapat ditransformasikan sesuai dengan karakteristik masing-masing variabel. Prinsip pemeriksaan, verifikasi, dan konsistensi data juga sejalan dengan ketentuan penyusunan AHSP yang menekankan pentingnya validitas dan kewajaran data [1].

Cronbach's alpha dihitung hanya menggunakan Q22–Q27 karena keenam butir tersebut digunakan untuk mengevaluasi konsistensi internal konstruk persepsi. Q28 tidak dimasukkan dalam pengujian reliabilitas karena berfungsi sebagai target klasifikasi. Target Q28 yang semula menggunakan lima tingkat respons direkode menjadi tiga kelas, yaitu Rendah yang merupakan gabungan kategori sangat rendah dan rendah, Sedang, serta Tinggi yang merupakan gabungan kategori tinggi dan sangat tinggi. Seluruh 83 respons memiliki nilai Q28 yang valid. Pemisahan antara butir konstruk dan target dilakukan untuk menjaga batas konseptual antara pengukuran persepsi dan variabel yang akan diprediksi.

Pemodelan Machine Learning

Pemodelan *machine learning* menggunakan Q2–Q17 sebagai variabel prediktor dan Q28 sebagai target klasifikasi. Q18–Q27 serta Q29–Q31 tidak digunakan sebagai prediktor karena memiliki fungsi analisis yang berbeda dan sebagian butir secara konseptual berdekatan dengan target persepsi kebutuhan penyesuaian AHSP. Pemisahan tersebut dilakukan untuk mengurangi risiko *target leakage*. Dengan demikian, model tidak dimaksudkan untuk menentukan koefisien AHSP lokal secara langsung, melainkan digunakan sebagai alat penapisan *(screening)* eksploratif untuk mengidentifikasi pola karakteristik responden dan kondisi proyek yang berkaitan dengan persepsi kebutuhan penyesuaian AHSP.

Dataset dibagi secara *stratified* dengan proporsi 70:30 dan *random_state* = 42 sehingga diperoleh 58 observasi sebagai data latih dan 25 observasi sebagai data uji *(holdout)*. Stratifikasi digunakan untuk mempertahankan proporsi kelas target pada kedua bagian data. Seluruh proses *encoding*, *scaling*, dan validasi dilakukan menggunakan data latih di dalam *pipeline*, sedangkan data uji *(holdout)* tidak digunakan dalam proses pemilihan model dan hanya digunakan satu kali untuk evaluasi akhir.

Transformasi variabel prediktor Q2–Q17 dilakukan berdasarkan karakteristik masing-masing variabel menggunakan kombinasi ordinal *encoding*, *capped one-hot encoding*, dan pengkodean jawaban multirespons. Variabel dengan kategori berurutan direpresentasikan melalui pengkodean ordinal, sedangkan variabel nominal ditransformasikan menjadi indikator kategori melalui *one-hot encoding* dengan pembatasan jumlah kategori. Jawaban multirespons dikodekan sebagai fitur indikator biner sesuai pilihan jawaban. Karena satu kelompok pertanyaan dapat menghasilkan lebih dari satu fitur setelah transformasi, penerapan konfigurasi kompak pada 16 kelompok prediktor Q2–Q17 menghasilkan total 56 fitur dari 58 observasi data latih, dengan rasio p/n sebesar 0,966. Konfigurasi transformasi ditentukan tanpa memanfaatkan

nilai target maupun performa model untuk mengurangi risiko kebocoran informasi. Meskipun jumlah fitur telah dikendalikan agar tidak melebihi jumlah observasi data latih, dimensionalitas masih relatif tinggi terhadap ukuran sampel. Oleh karena itu, hasil pemodelan diperlakukan sebagai analisis eksploratif dan tidak ditafsirkan sebagai model prediktif final yang siap digeneralisasikan secara luas.

Tiga algoritma *machine learning* dibandingkan dalam penelitian ini, yaitu Logistic Regression, Decision Tree, dan Random Forest. Ketiga model menggunakan *class_weight = balanced* untuk mengurangi dominasi kelas dengan jumlah observasi terbesar selama proses pembelajaran. Kinerja model pada data latih dievaluasi menggunakan *repeated stratified 5-fold cross-validation* sebanyak 10 pengulangan sehingga menghasilkan 50 proses validasi. F1-makro digunakan sebagai metrik utama karena memberikan bobot yang sama terhadap setiap kelas dan tidak didominasi oleh kelas mayoritas, sedangkan *balanced accuracy* digunakan sebagai metrik pendukung. *Majority-class baseline* digunakan sebagai pembanding sederhana untuk menilai apakah model memberikan kemampuan diskriminasi antarkelas yang lebih baik dibandingkan strategi yang hanya memprediksi kelas mayoritas. Model referensi untuk analisis lanjutan ditentukan berdasarkan rerata F1-makro pada *repeated cross-validation*. Setelah proses pemilihan tersebut, data uji *(holdout)* digunakan untuk mengevaluasi kemampuan generalisasi model melalui akurasi, *precision weighted*, *recall weighted*, F1-weighted, F1-makro, dan *balanced accuracy*.

Analisis uji ketahanan *(robustness)* dilakukan dengan mengeluarkan Q2–Q4 dari kelompok prediktor untuk mengevaluasi sensitivitas model terhadap karakteristik responden yang mencakup peran, wilayah, dan pengalaman proyek. Performa model setelah penghapusan Q2–Q4 dibandingkan dengan model penuh berdasarkan perubahan akurasi, F1-makro, dan *balanced accuracy* pada *repeated cross-validation*. Perubahan performa tersebut digunakan sebagai indikator sensitivitas prediktif dan tidak ditafsirkan sebagai bukti hubungan sebab-akibat.

Interpretasi kontribusi variabel dilakukan menggunakan dua pendekatan yang saling melengkapi, yaitu *grouped permutation importance* dan *leave-one-feature-out* (LOFO). *Grouped permutation importance* digunakan untuk mengevaluasi perubahan F1-makro pada data uji *(holdout)* ketika seluruh fitur turunan dari suatu kelompok pertanyaan diacak secara bersama. Pengelompokan dilakukan karena satu pertanyaan dapat menghasilkan lebih dari satu fitur setelah proses *encoding*. Sementara itu, LOFO dilakukan pada data latih menggunakan *repeated cross-validation* dengan mengeluarkan satu kelompok pertanyaan secara bergantian dan membandingkan performanya terhadap model penuh. Perubahan F1-makro dan *balanced accuracy* digunakan untuk mengidentifikasi kelompok variabel yang memberikan sinyal prediktif relatif lebih konsisten. Hasil kedua pendekatan tersebut tidak digunakan untuk menyimpulkan hubungan kausal.

Analisis pemeringkatan Q18–Q19 dilakukan secara terpisah dari pemodelan klasifikasi. Q18–Q19 tidak digunakan sebagai prediktor maupun target dalam model *machine learning*, tetapi digunakan untuk menentukan item pekerjaan yang perlu diprioritaskan dalam observasi produktivitas lapangan. Skor prioritas dihitung menggunakan persamaan $S = F_{RAB} + 2F_P$, dengan F_{RAB} sebagai frekuensi suatu item pekerjaan dilaporkan terdapat dalam RAB responden dan F_P sebagai frekuensi item tersebut dipilih secara eksplisit sebagai pekerjaan yang dinilai memerlukan koefisien lokal. Bobot 2 pada F_P digunakan secara heuristik untuk memberikan penekanan lebih besar pada pemilihan eksplisit dibandingkan sekadar kemunculan pekerjaan dalam RAB. Untuk mengevaluasi kestabilan pemeringkatan terhadap pemilihan bobot, dilakukan uji sensitivitas menggunakan bobot 1, 2, dan 3. Pemeringkatan tersebut digunakan untuk menentukan prioritas observasi lapangan dan tidak dimaksudkan sebagai dasar langsung penetapan koefisien AHSP lokal.

HASIL DAN PEMBAHASAN

Karakteristik Responden, Wilayah, dan Kondisi Proyek

Kelompok responden terbesar adalah konsultan pengawas 21 orang (25,30%), kontraktor 13 orang (15,66%), unsur dinas teknis/PU/PPK/PPTK 11 orang (13,25%), pelaksana lapangan 8 orang (9,64%), dan konsultan perencana 7 orang (8,43%). Pengalaman paling banyak berada pada kelompok 1-3 tahun sebanyak 20 orang (24,10%), diikuti 4-6 tahun 17 orang (20,48%), 7-10 tahun 14 orang (16,87%), dan lebih dari 10 tahun 13 orang (15,66%). Keragaman responden memberikan sudut pandang perencanaan, pelaksanaan, pengawasan, dan pengelolaan proyek.

 https://doi.org/10.56211/blendsains.v5i2.2095

Karakteristik Wilayah Berdasarkan Kabupaten di Provinsi Sulawesi Barat

Karakteristik lokasi kerja disajikan pada tingkat kabupaten agar variasi wilayah tidak terpecah oleh banyak nama kecamatan. Seluruh 83 respons (100,00%) berasal dari enam kabupaten di Provinsi Sulawesi Barat. Mamuju memiliki frekuensi terbesar 23 responden (27,71%), diikuti Polewali Mandar 18 (21,69%), Pasangkayu 12 (14,46%), Mamasa 11 (13,25%), Mamuju Tengah 10 (12,05%), dan Majene 9 (10,84%). Dengan demikian, seluruh kabupaten di Sulawesi Barat terwakili dalam sampel. Distribusi ini hanya menggambarkan karakteristik sampel karena teknik sampling bersifat *nonprobability* dan jumlah responden antar kabupaten tidak sama.

Tabel 1. Sebaran Responden Berdasarkan Wilayah (Kabupaten)

Aspek	Kategori/Temuan	n (%)
Kabupaten	Mamuju	23 (27,71)
Kabupaten	Polewali Mandar	18 (21,69)
Kabupaten	Pasangkayu	12 (14,46)
Kabupaten	Mamasa	11 (13,25)
Kabupaten	Mamuju Tengah	10 (12,05)
Kabupaten	Majene	9 (10,84)
Total Q3	**Enam kabupaten Sulawesi Barat**	**83 (100,00)**

Karakteristik Kondisi Proyek dan Pelaksanaan

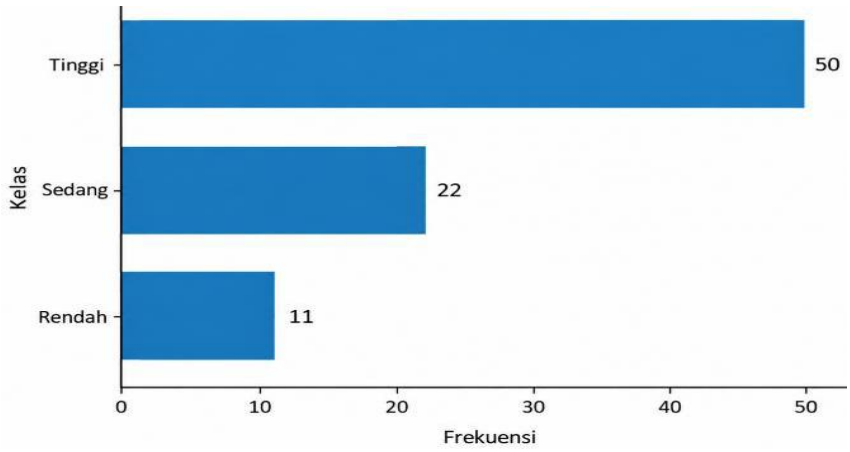
Pada pertanyaan multirespons, tanah biasa ditemukan pada 67 respons, tanah berbatu 35, dan tanah pasir 25. Faktor wilayah yang paling sering dilaporkan adalah cuaca/ hujan 39 respons, jarak sumber material 38, akses lokasi sulit 31, topografi/kemiringan 19, material tidak dapat diturunkan langsung pada titik pekerjaan 18, dan jenis tanah 16. Alat yang paling umum digunakan adalah sekop 70 respons, gerobak dorong 66, molen/concrete mixer 65, dan dump truck 58. Kendala pelaksanaan didominasi cuaca/hujan 43 respons, material jauh dari lokasi 38, material terlambat 37, dan keterbatasan tukang terampil 28. Pola ini konsisten dengan literatur mengenai akses, logistik, cuaca, material, alat, serta keterampilan tenaga kerja [4], [8].

Tabel 2. Karakteristik Kondisi Proyek dan Pelaksanaan

Aspek	Kategori/Temuan	n (%)
Topografi	Datar/Dataran	42 (50,60)
Kondisi jalan	Sedang	47 (56,63)
Kondisi jalan	Rusak/sulit dilalui	13 (15,66)
Jarak angkut	>100 meter	23 (27,71)
Metode	Kontraktor/penyedia	31 (37,35)
Metode	Swakelola	28 (33,73)
Sistem kerja	Harian	30 (36,14)
Tenaga dominan	Tenaga kerja lokal	49 (59,04)

Kesesuaian AHSP, Tingkat Kebutuhan, dan Reliabilitas

Penilaian terhadap AHSP standar menunjukkan 29 responden (34,94%) memilih kurang sesuai dan 7 responden (8,43%) memilih tidak sesuai, sedangkan 28 responden (33,73%) menilai cukup sesuai, 18 (21,69%) sudah sesuai, dan 1 (1,20%) tidak tahu. Pada Q21 yang bersifat multirespons, faktor kondisi lokasi/ faktor koreksi wilayah dipilih 47 responden (56,63%), mobilisasi bahan dan jarak angkut 41 (49,40%), produktivitas tenaga kerja 35 (42,17%), koefisien alat 23 (27,71%), koefisien bahan 21 (25,30%), dan waktu pelaksanaan 18 (21,69%). Temuan tersebut menegaskan bahwa kebutuhan penyesuaian lebih berkaitan dengan kondisi aktual dan produktivitas daripada semata-mata perubahan harga material [1], [2].



Gambar 1. Distribusi kelas kebutuhan penyesuaian AHSP

Pada skala asli lima tingkat, 28 responden (33,73%) menyatakan kebutuhan sangat tinggi, 22 (26,51%) tinggi, 22 (26,51%) sedang, 8 (9,64%) rendah, dan 3 (3,61%) sangat rendah. Setelah direkode, kelas Tinggi berjumlah 50 responden (60,24%), Sedang 22 (26,51%), dan Rendah 11 (13,25%). Ketidakseimbangan ini menjadi alasan penggunaan F1-makro dan *balanced accuracy* selain akurasi dalam evaluasi model.

Cronbach's alpha untuk enam butir Q22-Q27 adalah 0,8649 dan menunjukkan konsistensi internal yang tinggi pada sampel penelitian. Q28 tidak dimasukkan dalam perhitungan reliabilitas karena digunakan sebagai target klasifikasi. Pemisahan ini menjaga batas antara konstruk persepsi dan target klasifikasi serta sejalan dengan prinsip bahwa alpha digunakan untuk mengevaluasi konsistensi internal item yang mengukur konstruk yang sama [11].

Evaluasi Model Klasifikasi dan* Uji ketahanan *(robustness)

Tabel 3. Kinerja model pada *repeated stratified 5-fold cross-validation × 10* (data latih)

Model	*Accuracy*	F1-Makro	*Balanced Accuracy*
Decision Tree	0,5120 ± 0,1132	**0,3661 ± 0,1252**	**0,3954 ± 0,1439**
Logistic Regression	0,4870 ± 0,1091	0,3499 ± 0,1171	0,3768 ± 0,1293
Random Forest	0,5742 ± 0,0777	0,2974 ± 0,0974	0,3462 ± 0,0864
Majority Baseline	**0,6045 ± 0,0262**	0,2511 ± 0,0068	0,3333 ± 0,0000

Pada *repeated stratified 5-fold cross-validation* × 10 yang dilakukan pada 58 data latih, *majority-class baseline* memperoleh akurasi tertinggi sebesar 0,6045 ± 0,0262. Di antara model *machine learning*, Random Forest menghasilkan rerata akurasi tertinggi sebesar 0,5742 ± 0,0777. Namun, Decision Tree memperoleh F1-makro tertinggi sebesar 0,3661 ± 0,1252 dan *balanced accuracy* tertinggi sebesar 0,3954 ± 0,1439. Karena distribusi target tidak seimbang, F1-makro digunakan sebagai metrik utama sehingga Decision Tree ditetapkan sebagai model referensi untuk analisis lanjutan. Meskipun demikian, nilai *majority baseline accuracy* yang lebih tinggi serta simpangan baku yang relatif besar menunjukkan bahwa kemampuan model masih terbatas dan bervariasi antar pembagian data. Oleh karena itu, hasil pemodelan tetap diposisikan sebagai penapisan *(screening)* eksploratif.

Tabel 4. Kinerja Model pada Data uji *(holdout)*

Model	Accuracy	Precision Weighted	Recall Weighted	F1-Weighted	F1-Makro	Balanced Accuracy
Majority Baseline	0,6000	0,3600	0,6000	0,4500	0,2500	0,3333
Logistic Regression	0,6400	**0,6777**	0,6400	**0,6550**	**0,5349**	**0,5460**
Decision Tree	0,4800	0,5422	0,4800	0,5043	0,4008	0,4063
Random Forest	**0,6800**	0,6691	**0,6800**	0,6087	0,5110	0,4921

 https://doi.org/10.56211/blendsains.v5i2.2095

Evaluasi pada 25 observasi data uji *(holdout)* menunjukkan bahwa Random Forest memperoleh akurasi tertinggi sebesar 0,6800. Namun, Logistic Regression memperoleh F1-makro tertinggi sebesar 0,5349 dan *balanced accuracy* tertinggi sebesar 0,5460. Karena F1-makro dan *balanced accuracy* digunakan untuk mempertimbangkan ketidakseimbangan kelas, Logistic Regression menunjukkan kemampuan diskriminasi antarkelas terbaik pada data uji *(holdout)*. Perbedaan peringkat kinerja model antara *repeated cross-validation* dan evaluasi *holdout* dapat disebabkan oleh ukuran sampel yang relatif terbatas serta perbedaan karakteristik algoritma dalam mempelajari pola data. Dengan jumlah data latih sebanyak 58 observasi dan 56 fitur hasil transformasi, *Decision Tree* berpotensi mengalami *overfittin*g karena pembentukan aturan pemisahan yang spesifik terhadap karakteristik sampel pelatihan. Sebaliknya, *Logistic Regression* memiliki struktur model linier yang lebih sederhana sehingga berpotensi menghasilkan generalisasi yang lebih stabil pada sebagian pembagian data. Namun, keunggulan *Logistic Regression* pada 25 observasi *holdout* belum cukup untuk membuktikan bahwa algoritma tersebut secara konsisten lebih baik. Variasi distribusi kelas dan komposisi sampel dapat memengaruhi hasil evaluasi. Oleh karena itu, perbedaan kinerja tersebut menunjukkan bahwa hasil pemodelan masih bersifat eksploratif dan belum dapat digunakan untuk menetapkan satu algoritma sebagai model terbaik secara umum.

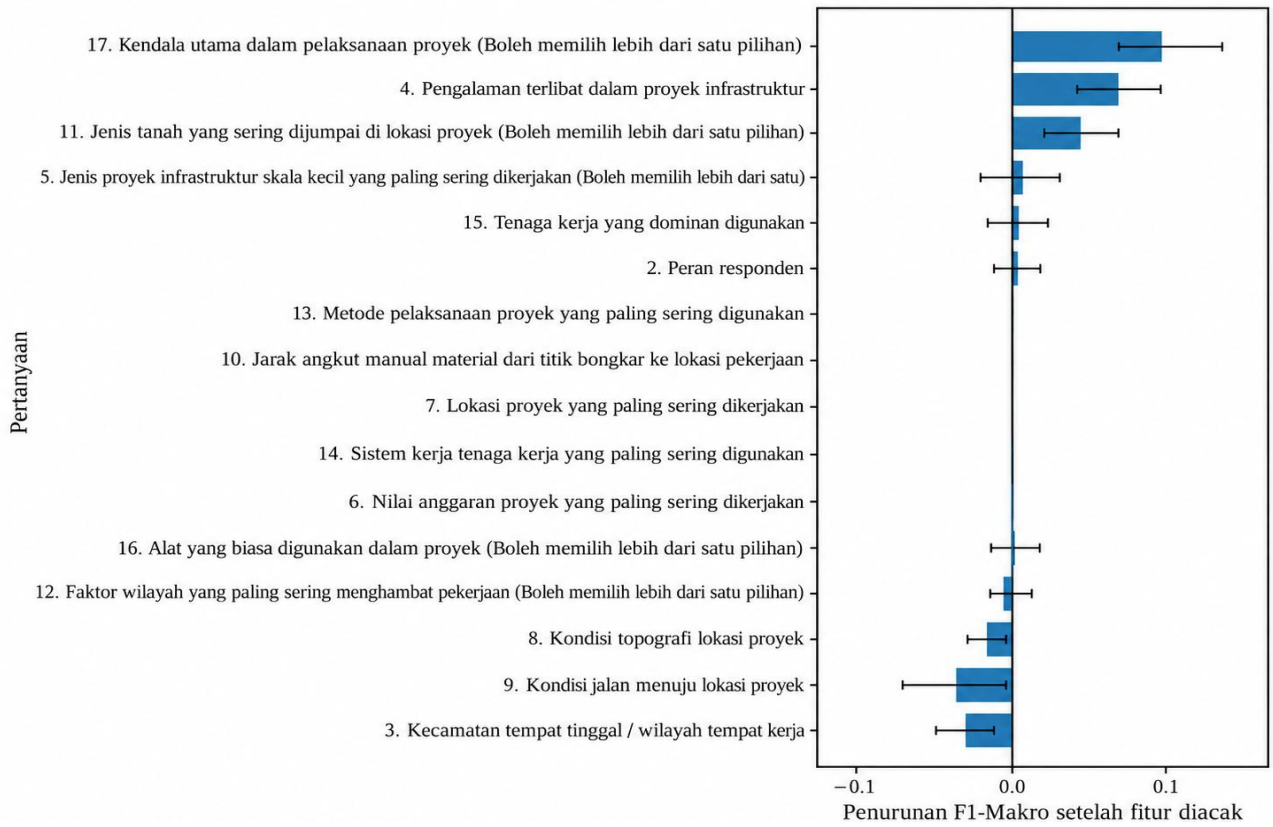
Tabel 5. Confusion Matrix Decision Tree pada Data Uji

Aktual / Prediksi	Rendah	Sedang	Tinggi
Rendah	**1**	**2**	**0**
Sedang	**2**	**2**	**3**
Tinggi	**1**	**5**	**9**

Evaluasi per kelas Decision Tree pada data uji menunjukkan F1 sebesar 0,2857 untuk kelas Rendah, 0,2500 untuk kelas Sedang, dan 0,6667 untuk kelas Tinggi. *Confusion matrix* menunjukkan bahwa 12 dari 25 observasi diklasifikasikan dengan benar. Kemampuan klasifikasi relatif lebih baik pada kelas Tinggi, sedangkan kelas Rendah dan Sedang masih menunjukkan kemampuan diskriminasi yang terbatas.

Uji ketahanan *(robustness)* dengan menghapus Q2–Q4 menurunkan jumlah fitur dari 56 menjadi 44. Pada Decision Tree, rerata *accuracy repeated cross-validation* menurun dari 0,5120 menjadi 0,4979 atau sebesar 0,0141. F1-makro menurun dari 0,3661 menjadi 0,3582 atau sebesar 0,0079, sedangkan *balanced accuracy* menurun dari 0,3954 menjadi 0,3849 atau sebesar 0,0105. Penurunan tersebut relatif kecil, sehingga menunjukkan bahwa Q2–Q4 memberikan tambahan informasi prediktif, tetapi performa model tidak sepenuhnya bergantung pada karakteristik peran, wilayah, dan pengalaman responden. Perubahan performa ditafsirkan sebagai sensitivitas prediktif dan bukan sebagai bukti hubungan sebab-akibat.

Interpretasi Faktor Prediktif dan Sensitivitas Model



Gambar 2. *Grouped Permutation Importance* Model Decision Tree

Grouped permutation importance pada data uji *(holdout)* menunjukkan penurunan F1-makro terbesar ketika Q17, yaitu kendala utama pelaksanaan proyek, diacak sebesar 0,1054 ± 0,0461. Posisi berikutnya ditempati Q4, yaitu pengalaman proyek, sebesar 0,0877 ± 0,0509 dan Q11, yaitu jenis tanah, sebesar 0,0550 ± 0,0439. Beberapa kelompok fitur lainnya menghasilkan nilai mendekati nol atau negatif. Nilai nol maupun negatif tidak ditafsirkan sebagai pengaruh negatif secara kausal, tetapi menunjukkan bahwa pengacakan kelompok fitur tersebut pada data uji *(holdout)* tidak selalu menyebabkan penurunan performa model. Mengingat data uji *(holdout)* hanya terdiri atas 25 observasi, hasil *permutation importance* ditafsirkan secara hati-hati dan dibaca bersama dengan analisis *leave-one-feature-out* berbasis *repeated cross-validation*.

Tabel 6. Sensitivitas LOFO lima peringkat teratas

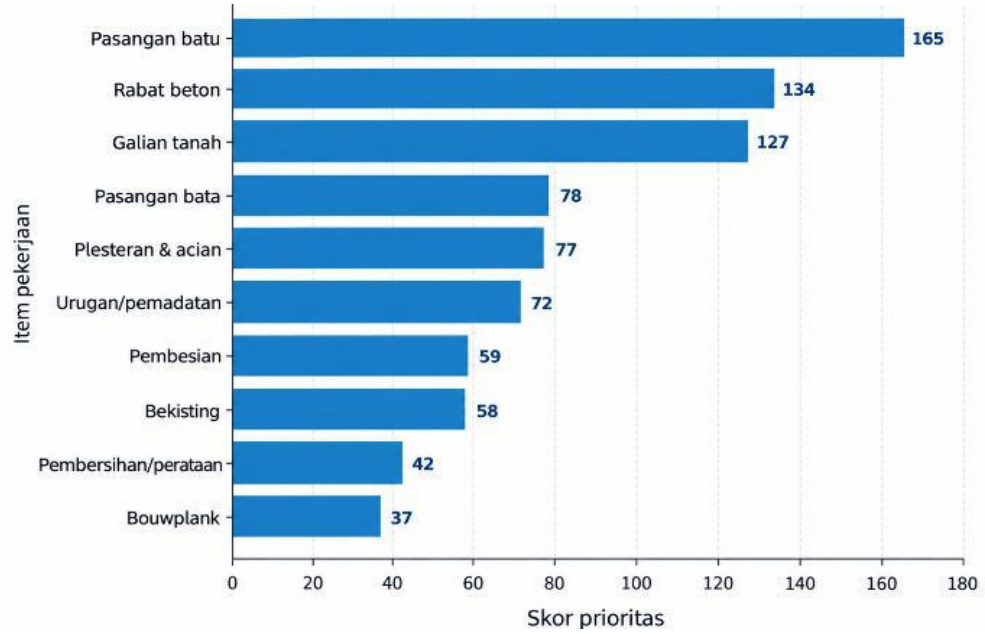
Peringkat	Fitur yang Dihapus	Penurunan F1-Makro	Penurunan Balanced Accuracy
1	Q4 – Pengalaman proyek	**0,0297**	**0,0313**
2	Q14 – Sistem kerja tenaga kerja	0,0063	0,0052
3	Q8 – Kondisi topografi	0,0043	0,0122
4	Q13 – Metode pelaksanaan	-0,0007	0,0011
5	Q16 – Penggunaan alat	-0,0020	-0,0035

Analisis *leave-one-feature-out* (LOFO) berbasis *repeated cross-validation* menunjukkan bahwa penghapusan Q4, yaitu pengalaman proyek, menghasilkan penurunan performa terbesar dibandingkan kelompok fitur lainnya. F1-makro menurun dari 0,3661 menjadi 0,3364 atau sebesar 0,0297, sedangkan *balanced accuracy* menurun dari 0,3954 menjadi 0,3641 atau sebesar 0,0313. Penghapusan Q14, yaitu sistem kerja tenaga kerja, menghasilkan penurunan F1-makro sebesar 0,0063 dan *balanced accuracy* sebesar 0,0052, sementara penghapusan Q8, yaitu kondisi topografi, menurunkan F1-makro sebesar 0,0043 dan *balanced accuracy* sebesar 0,0122. Sebaliknya, penghapusan Q13 dan Q16 menghasilkan perubahan F1-makro yang sedikit negatif, yang menunjukkan bahwa pengeluaran kedua kelompok fitur tersebut tidak menurunkan performa model dan dapat mengindikasikan adanya redundansi atau *derau (noise)* pada konfigurasi model. Jika dibandingkan dengan hasil *grouped permutation importance*, Q4 menempati peringkat kedua pada data uji *(holdout)* dan peringkat pertama pada analisis LOFO. Oleh karena itu, pengalaman proyek dapat dipandang sebagai sinyal prediktif yang relatif paling konsisten pada kedua pendekatan. Sementara itu, Q17 menghasilkan *permutation importance* tertinggi

 https://doi.org/10.56211/blendsains.v5i2.2095

pada data uji *(holdout)*, tetapi tidak menunjukkan dukungan yang sama pada LOFO, sehingga kontribusinya belum dapat dianggap stabil. Seluruh hasil tersebut menunjukkan kontribusi prediktif relatif dan tidak ditafsirkan sebagai hubungan sebab-akibat.

Prioritas Item Pekerjaan, Implikasi, dan Keterbatasan



Gambar 3. Sepuluh Prioritas Item Pekerjaan untuk Observasi Koefisien Lokal

Pemeringkatan Q18-Q19 menempatkan pasangan batu gunung/batu kosong/batu belah sebagai prioritas pertama dengan skor 165, diikuti rabat beton/pekerjaan beton sederhana 134 dan galian tanah 127. Prioritas berikutnya adalah pasangan bata merah/batu bata 78, plesteran dan acian 77, urugan/ timbunan/ pemadatan 72, pembesian/ baja tulangan 59, bekisting 58, pembersihan/ perataan lapangan 42, dan pemasangan bouwplank 37. Skor menggunakan S = F_RAB + 2F_P untuk memberikan bobot lebih besar pada pemilihan eksplisit di Q19. Uji sensitivitas menggunakan bobot 1, 2, dan 3 mempertahankan tiga peringkat teratas yang sama, sehingga prioritas utama relatif stabil terhadap pilihan bobot. Pemeringkatan ini berbeda fungsi dari *feature importance*: Q18-Q19 menentukan objek pekerjaan yang perlu lebih dahulu dikalibrasi, sedangkan model ML mengidentifikasi karakteristik responden/ proyek yang membantu klasifikasi persepsi kebutuhan penyesuaian.

Keterbatasan penelitian ini terutama berkaitan dengan ukuran sampel dan karakteristik data yang digunakan. Penelitian melibatkan 83 responden dengan distribusi kelas target 50:22:11, yang selanjutnya dibagi menjadi 58 observasi data latih dan 25 observasi data uji. Representasi fitur kompak menghasilkan 56 fitur pada data latih dengan rasio p/n sebesar 0,966. Meskipun jumlah fitur telah dikendalikan agar tidak melebihi jumlah observasi data latih, dimensionalitas masih relatif tinggi sehingga hasil pemodelan diposisikan sebagai analisis eksploratif dan interpretasi terhadap performa model perlu dilakukan secara proporsional. Teknik sampling yang digunakan bersifat *nonprobability purposive sampling* dengan unit analisis berupa responden berdasarkan pengalaman proyek, bukan proyek yang diidentifikasi secara unik. Dengan demikian, kemungkinan adanya beberapa responden yang merujuk pada pengalaman proyek yang sama belum dapat dievaluasi secara khusus.

Seluruh enam kabupaten di Provinsi Sulawesi Barat telah terwakili dalam sampel, tetapi jumlah responden antarwilayah tidak sama. Oleh karena itu, distribusi wilayah pada penelitian ini menggambarkan karakteristik sampel dan belum dimaksudkan sebagai representasi proporsional seluruh populasi pelaku proyek infrastruktur di Sulawesi Barat. Selain itu, prediktor dan target diperoleh dari instrumen kuesioner yang sama sehingga kemungkinan pengaruh common-method bias belum dapat sepenuhnya dikesampingkan. Perbedaan metode pengumpulan data, yaitu 53 respons melalui pengisian mandiri dan 30 respons melalui wawancara, juga dapat menimbulkan variasi pola respons yang belum dianalisis secara terpisah dalam penelitian ini.

Pada aspek pemodelan, performa klasifikasi untuk kelas Rendah dan terutama kelas Sedang masih lebih terbatas dibandingkan kelas Tinggi. Peringkat performa model juga berbeda antara *repeated cross-validation* dan evaluasi data uji *(holdout)*, sedangkan hasil *permutation importance* diperoleh dari data uji yang berjumlah relatif terbatas. Oleh karena itu, hasil evaluasi model dan interpretasi kontribusi fitur sebaiknya dibaca sebagai indikasi pola prediktif pada dataset penelitian ini, bukan sebagai bukti hubungan sebab-akibat. Penghapusan Q2–Q4 hanya menghasilkan penurunan performa yang relatif kecil, sehingga model tidak sepenuhnya bergantung pada informasi mengenai peran, wilayah, dan pengalaman responden. Meskipun demikian, validasi menggunakan sampel dan proyek independen masih diperlukan sebelum model dapat digunakan pada konteks yang lebih luas.

Target Q28 juga memiliki urutan alami dari Rendah, Sedang, hingga Tinggi, sementara pendekatan dalam penelitian ini memperlakukannya sebagai klasifikasi multikelas nominal. Pendekatan tersebut memadai untuk tujuan eksploratif penelitian ini, tetapi belum mempertimbangkan secara khusus hubungan ordinal antar kelas. Penelitian selanjutnya dapat memperluas jumlah sampel, meningkatkan representasi setiap kelas dan wilayah, menggunakan identitas proyek yang unik, melakukan validasi eksternal, serta membandingkan pendekatan klasifikasi multikelas dengan metode klasifikasi ordinal. Dengan mempertimbangkan keterbatasan tersebut, hasil penelitian ini lebih tepat digunakan sebagai dasar penapisan (*screening*) awal dan pengembangan penelitian lanjutan daripada sebagai dasar langsung untuk generalisasi atau penetapan koefisien AHSP lokal.

KESIMPULAN

Hasil penelitian terhadap 83 responden menunjukkan adanya heterogenitas kondisi pelaksanaan proyek infrastruktur skala kecil di enam kabupaten Provinsi Sulawesi Barat. Kondisi jalan kategori sedang dilaporkan oleh 56,63% responden, kondisi jalan rusak atau sulit dilalui sebesar 15,66%, dan jarak angkut manual lebih dari 100meter sebesar 27,71%. Tenaga kerja lokal menjadi tenaga kerja dominan pada 59,04% responden, sedangkan kendala pelaksanaan yang paling sering dilaporkan mencakup cuaca atau hujan, jarak sumber material yang jauh, keterlambatan material, serta keterbatasan tenaga kerja terampil. Temuan tersebut menunjukkan bahwa kondisi lokal proyek merupakan aspek penting yang perlu dipertimbangkan dalam evaluasi kesesuaian AHSP.

Persepsi kebutuhan penyesuaian AHSP didominasi oleh kelas Tinggi sebanyak 50 responden (60,24%), diikuti kelas Sedang 22 responden (26,51%) dan Rendah 11 responden (13,25%). Uji reliabilitas Q22–Q27 menghasilkan *Cronbach's alpha* sebesar 0,8649. Representasi fitur kompak menghasilkan 56 fitur dari 58 data latih dengan rasio p/n sebesar 0,966. Pada *repeated stratified 5-fold cross-validation*, Decision Tree memperoleh rerata F1-makro tertinggi sebesar 0,3661 ± 0,1252 dengan *balanced accuracy* sebesar 0,3954 ± 0,1439 sehingga digunakan sebagai model referensi. Pada data uji, Logistic Regression memperoleh F1-makro tertinggi sebesar 0,5349 dan *balanced accuracy* sebesar 0,5460, sedangkan Random Forest memperoleh akurasi tertinggi sebesar 0,6800. Perbedaan peringkat tersebut menunjukkan bahwa belum terdapat satu algoritma yang secara konsisten dominan.

Analisis kontribusi variabel menunjukkan bahwa Q17, yaitu kendala utama pelaksanaan proyek, menghasilkan *permutation importance* tertinggi pada data uji, sedangkan Q4, yaitu pengalaman proyek, menunjukkan sinyal prediktif yang relatif paling konsisten karena menempati peringkat pertama pada analisis LOFO dan peringkat kedua pada *permutation importance*. Pemeringkatan Q18–Q19 menempatkan pasangan batu, rabat beton atau pekerjaan beton sederhana, dan galian tanah sebagai tiga prioritas utama untuk observasi produktivitas lapangan. Dengan demikian, model dalam penelitian ini diposisikan sebagai alat penapisan eksploratif untuk mengidentifikasi pola persepsi kebutuhan penyesuaian AHSP dan menentukan prioritas kajian lanjutan, bukan sebagai dasar langsung penetapan koefisien AHSP lokal. Penetapan koefisien lokal tetap memerlukan pengukuran produktivitas aktual dan validasi teknis melalui observasi lapangan.

SARAN

Penelitian selanjutnya direkomendasikan berfokus pada pengukuran produktivitas aktual tenaga kerja, jam alat, konsumsi bahan, waktu produktif/ tidak produktif, jarak angkut atau lansir, kondisi akses, topografi, jenis tanah, cuaca, serta metode pelaksanaan pada item prioritas. Data tersebut diperlukan untuk membandingkan koefisien aktual dengan AHSP standar dan menyusun kandidat faktor koreksi lokal yang dapat diuji secara teknis sesuai SE DJBK No. 47/SE/Dk/2026 [1].

Observasi lapangan sebaiknya dimulai dari pasangan batu, rabat beton, galian tanah, pasangan bata, plesteran/ acian, urugan/ pemadatan, pembesian, bekisting, pembersihan/ perataan lapangan, dan pemasangan bouwplank. Sampel responden dan proyek unik perlu diperluas, khususnya untuk meningkatkan representasi kelas Rendah serta variasi lokasi, kemudian model divalidasi pada proyek independen sebelum digunakan sebagai sistem pendukung keputusan. Produktivitas dapat dinyatakan sebagai output per orang-jam dan koefisien aktual sebagai orang-hari per satuan output; jam alat dan konsumsi bahan dicatat sesuai satuan pekerjaan. Tahap lanjutan perlu membandingkan hasil observasi dengan koefisien AHSP, mengevaluasi stabilitas faktor koreksi, dan menguji apakah peningkatan jumlah data memperbaiki F1-makro serta *balanced accuracy* pada seluruh kelas.

DAFTAR PUSTAKA

[1] Kementerian Pekerjaan Umum Republik Indonesia, "Surat Edaran Direktur Jenderal Bina Konstruksi Nomor 47/SE/Dk/2026 tentang Tata Cara Penyusunan Perkiraan Biaya Pekerjaan Konstruksi Bidang Pekerjaan Umum, Lampiran I, II, dan VI," 2026.

[2] Kementerian Pekerjaan Umum dan Perumahan Rakyat Republik Indonesia, "Peraturan Menteri Pekerjaan Umum dan Perumahan Rakyat Republik Indonesia Nomor 8 Tahun 2023 tentang Pedoman Penyusunan Perkiraan Biaya Pekerjaan Konstruksi Bidang Pekerjaan Umum dan Perumahan Rakyat," 2023.

[3] Western Federal Lands Highway Division, "Estimating Handbook. U.S. Department of Transportation, Federal Highway Administration," Dec. 2025.

[4] P. K. Mudita, I. K. Sudarsana, and M. Nadiasa, "Analisis Faktor-Faktor yang Mempengaruhi Waktu Tunggu Pengadaan Material Konstruksi pada Proyek Gedung di Kabupaten Badung," *Jurnal Spektran*, vol. 4, no. 2, pp. 18–26, Jul. 2016, doi: 10.24843/SPEKTRAN.2016.v04.i02.p03.

[5] A. H. Nawir, A. K. Hadi, and S. Bachmid, "Optimasi Biaya Pengadaan Material Pekerjaan Gedung Puskesmas di Kabupaten Pohuwato Provinsi Gorontalo," *Jurnal Konstruksi: Teknik, Infrastruktur dan Sains*, vol. 1, no. 3, pp. 22–30, Mar. 2022.

[6] F. Jian, Q. Liu, C. Feng, Q. Hu, Q. Yu, and Q. Guo, "Critical Factors Affecting Construction Labor Productivity: A Systematic Review and Meta-Analysis," *Buildings*, vol. 15, no. 14, Jul. 2025, doi: 10.3390/buildings15142463.

[7] F. Elza and W. Hatmoko, "Analisis Faktor Penyebab Keterlambatan pada Proyek Konstruksi Pembangunan Tebing Sungai di Kota Sungai Penuh," *Jurnal Talenta Sipil*, vol. 9, no. 1, pp. 64-72, Feb. 2026, doi: 10.33087/talentasipil.v9i1.1071.

[8] M. Yazid, Y. P. W. Prasetyo, and Sunarko, "Identifikasi dan Analisis Faktor Dominan Penyebab Keterlambatan Proyek Konstruksi di Kecamatan Songgon, Banyuwangi," *Sultra Civil Engineering Journal*, vol. 6, no. 2, pp. 545–552, Oct. 2025, doi: 10.54297/sciej.v6i2.1234.

[9] M. Koo and S. W. Yang, "Likert-Type Scale," Encyclopedia, vol. 5, no. 1, art. no. 18, Feb. 2025, doi: 10.3390/encyclopedia5010018.

[10] I. Kusmaryono, D. Wijayanti, and H. R. Maharani, "Number of Response Options, Reliability, Validity, and Potential Bias in the Use of the Likert Scale Education and Social Science Research: A Literature Review," International Journal of Educational Methodology, vol. 8, no. 4, pp. 625–637, 2022, doi: 10.12973/ijem.8.4.625.

[11] M. Tavakol and R. Dennick, "Making sense of Cronbach's alpha," International Journal of Medical Education, vol. 2, pp. 53–55, Jun. 2011, doi: 10.5116/ijme.4dfb.8dfd.

[12] S. Tayefeh Hashemi, O. M. Ebadati, and H. Kaur, "Cost estimation and prediction in construction projects: a systematic review on machine learning techniques," *SN Applied Sciences*, vol. 2, art. no. 1703, 2020, doi: 10.1007/s42452-020-03497-1.

[13] L. Chen, S. S. Tiang, K. S. Chong, A. Sharma, T. Berghout, and W. H. Lim, "Predicting Construction Costs with Machine Learning: A Comparative Study on Ensemble and Linear Models," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 3, pp. 624–645, Jul. 2025, doi: 10.35882/jeeemi.v7i3.799.

[14] L. Chen *et al.*, "Transparent and reliable construction cost prediction using advanced machine learning and explainable AI," *Engineering Science and Technology, an International Journal*, vol. 70, pp. 1–27, Oct. 2025, doi: 10.1016/j.jestch.2025.102159.

[15] A. Subrata, Parman, and M. Mughnyanti, "Penggunaan Machine Learning untuk Estimasi Biaya dan Waktu Konstruksi Bangunan Gedung," *Journal of Science and Social Research*, vol. 8, no. 3, pp. 3791–3795, Aug. 2025, doi: 10.54314/jssr.v8i3.3322.