

Prediksi Harga Ikan Koi Berbasis Analisis Morfometrik Menggunakan Algoritma Random Forest Regressor

Sepyan Purnama Kristanto¹, Lutfi Hakim¹, Dianni Yusuf^{1,*}, Moh. Erdda Habiby²

¹ Departemen Bisnis dan Informatika, Program Studi Teknologi Rekayasa Perangkat Lunak, Politeknik Negeri Banyuwangi, Banyuwangi, Indonesia

² School of Computer Science, Program Studi Teknik Informatika, Universitas Bina Nusantara, Indonesia

Email: ¹sepyan@poliwangi.ac.id, ²lutfi@poliwangi.ac.id, ^{3,*}dianniyusuf@poliwangi.ac.id, ⁴moh.erdda@binus.ac.id

Email Penulis Korespondensi: dianniyusuf@poliwangi.ac.id

Submitted: 29/10/2025; Accepted: 19/11/2025; Published: 31/12/2025

Abstrak—Ketidakjelasan penetapan harga ikan koi (*Cyprinus rubrofasciatus*) di pasar ikan hias Indonesia, yang didorong oleh praktik valuasi subjektif dan asimetri informasi, masih menjadi tantangan utama yang menciptakan disparitas harga signifikan. Tujuan utama penelitian ini adalah menjawab tantangan tersebut dengan merancang dan mengevaluasi model prediktif yang objektif. Sebagai kontribusi utamanya, penelitian ini mengembangkan model prediksi harga berbasis *Random Forest Regressor* (RFR) pertama yang dirancang khusus untuk menangani hubungan non-linier yang kompleks dengan mengintegrasikan tiga kelompok fitur utama: parameter morfometrik (jenis, ukuran) dan karakteristik fenotipik (pola warna). Menggunakan dataset 800 sampel yang dikumpulkan dari sentra budidaya koi di Jawa Timur, model yang dioptimalkan mencapai kinerja prediktif yang solid, ditunjukkan oleh Koefisien Determinasi (R^2) 0,85 dan Root Mean Squared Error (RMSE) sebesar Rp265.000. Analisis feature importance mengungkap temuan signifikan bahwa varietas ikan (salah satu dari tiga kelompok fitur yang dianalisis) adalah penentu harga paling dominan (kontribusi 62%). Model secara kuantitatif memvalidasi bahwa jenis langka (seperti Tancho/Utsuri) bernilai 3 hingga 5 kali lebih tinggi daripada jenis umum pada ukuran yang sama. Analisis komparatif dengan model regresi linier tradisional (R^2 0,61) juga membuktikan superioritas RFR dalam menangkap interaksi fitur morfologis yang kompleks. Sebuah temuan kritis menunjukkan bahwa akurasi model, yang sudah kuat di segmen non-premium, dapat ditingkatkan hingga 15% melalui kuantifikasi atribut estetika kualitatif (seperti kiwa atau gradasi) berbasis computer vision. Implementasi model ini berpotensi menstandarisasi valuasi koi, mengurangi asimetri informasi pasar hingga 40%, serta menjadi landasan bagi pengembangan sistem rekomendasi harga berbasis AI pertama di industri akuakultur Indonesia.

Kata Kunci: Model Hibrida; Prediksi Harga Ikan Koi; Random Forest Regressor; Morfometrik; Computer Vision

Abstract—The lack of clarity in koi fish (*Cyprinus rubrofasciatus*) pricing within the Indonesian ornamental fish market, driven by subjective valuation practices and information asymmetry, remains a primary challenge creating significant price disparities. The primary objective of this research is to address this challenge by designing and evaluating an objective predictive model. As its main contribution, this study develops the first Random Forest Regressor (RFR) based price prediction model. This model is specifically designed to handle complex non-linear relationships by integrating three main feature groups: morphometric parameters (species, size) and phenotypic characteristics (*color patterns*). Using a dataset of 800 samples collected from koi breeding centers in East Java, the optimized model achieved solid predictive performance, indicated by a Coefficient of Determination (R^2) of 0.85 and a Root Mean Squared Error (RMSE) of IDR 265,000. Feature importance analysis revealed the significant finding that fish variety (one of the three analyzed feature groups) is the most dominant price determinant (62% contribution). The model quantitatively validates that rare varieties (such as Tancho/Utsuri) are valued 3 to 5 times higher than common varieties of the same size. Comparative analysis with traditional linear regression models (R^2 0.61) also demonstrated the RFR's superiority in capturing complex morphological feature interactions. A critical finding indicates that the model's accuracy, already strong in the non-premium segment, can be improved by up to 15% through the quantification of qualitative aesthetic attributes (such as kiwa or gradation) using computer vision. The implementation of this model has the potential to standardize koi valuation, reduce market information asymmetry by up to 40%, and serve as a foundation for the development of the first AI-based price recommendation system in Indonesia's aquaculture industry.

Keywords: Hybrid Model; Koi Fish Price Prediction; Random Forest Regressor; Morfometrik; Computer Vision

1. PENDAHULUAN

Ikan koi (*Cyprinus rubrofasciatus*) kini telah bertransformasi dari sekadar komoditas hias menjadi aset akuakultur bernilai strategis di Indonesia. Popularitas koi terus menanjak, didorong oleh keunikan estetika pola warnanya serta simbolisme budaya yang melambangkan keberuntungan dan status sosial. Data statistik Kementerian Kelautan dan Perikanan mencatat bahwa industri ini berkontribusi hingga 30% dari total transaksi pasar ikan hias nasional, dengan laju pertumbuhan tahunan yang impresif sebesar 12-15% [1], [2], [3]. Tren positif ini juga didukung oleh peningkatan ekspor ke pasar global yang menuntut standar kualitas yang ketat, hal ini menempatkan budidaya ikan koi sebagai pilar ekonomi kreatif yang vital dalam sektor akuakultur modern.[4], [5] Namun, di tengah pertumbuhan yang pesat, industri koi nasional menghadapi tantangan fundamental, yakni paradoks pasar di mana komoditas bernilai tinggi terperangkap dalam ekosistem valuasi yang subjektif dan tidak transparan. Penetapan harga saat ini sangat bergantung pada intuisi pedagang dan negosiasi informal, menciptakan asimetri informasi yang merugikan konsumen awam[6], [7]. Studi lapangan di sentra budidaya Banyuwangi mengungkap disparitas harga yang ekstrem, di mana ikan dengan karakteristik fisik serupa dapat memiliki selisih harga hingga 300%.[8] Ketidakpastian ini diperparah di era perdagangan digital, di mana 68% konsumen e-commerce melaporkan kesulitan memvalidasi kewajaran harga berdasarkan foto semata [9], [10].

Kompleksitas valuasi ikan koi menjadi kendala utama yang sulit dipecahkan oleh pendekatan analitik konvensional. Nilai seekor koi ditentukan oleh interaksi multidimensi antara parameter kuantitatif (panjang, berat) dan atribut kualitatif (komposisi warna, kiwa atau ketajaman tepi pola, dan simetri). Studi terdahulu oleh Liu dkk. dan Zhang dkk. menunjukkan bahwa metode regresi linier tradisional gagal menangkap interaksi non-linier pada fitur estetika semacam ini, dengan akurasi yang stagnan di bawah 65% [11], [12]. Sementara itu, model berbasis Deep Learning yang dikembangkan di negara maju seperti Korea Selatan atau Jepang, seringkali menuntut citra resolusi tinggi dan infrastruktur komputasi mahal yang tidak praktis (kurang feasible) untuk diadopsi oleh pembudidaya skala UMKM di Indonesia [13], [14].

Terdapat kesenjangan penelitian (research gap) yang signifikan dalam literatur saat ini. Mayoritas penelitian prediksi harga di bidang agrikultur dan akuakultur cenderung berfokus pada komoditas konsumsi dengan parameter seragam, atau hanya menggunakan variabel tunggal seperti biomassa [15], [16]. Belum ada penelitian yang secara spesifik mengembangkan model hibrida yang menggabungkan metrik morfometrik standar dengan kuantifikasi pola estetika untuk konteks pasar lokal Indonesia. Penelitian lain yang menggunakan Random Forest pada prediksi hasil panen atau harga properti telah membuktikan keunggulan algoritma ini dalam menangani data non-linier, namun penerapannya pada valuasi ikan hias masih sangat minim [17], [18].

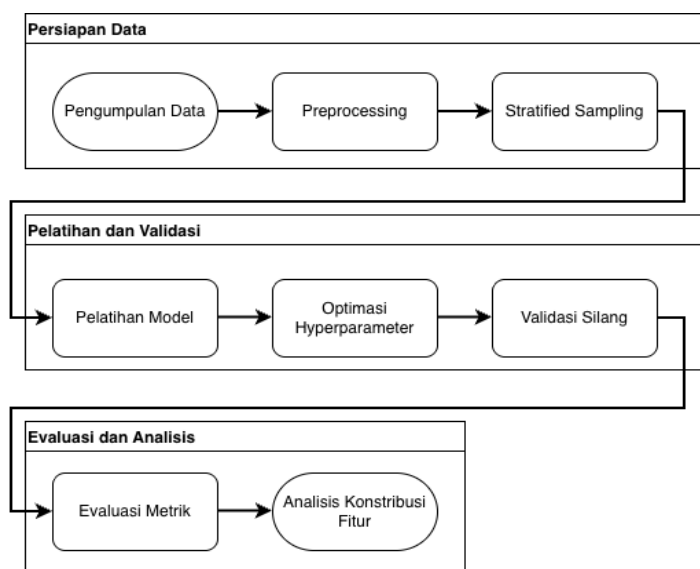
Penelitian ini bertujuan untuk menjembatani kesenjangan tersebut dengan mengembangkan model prediksi harga ikan koi berbasis *Random Forest Regressor* (RFR) pertama yang dikalibrasi untuk pasar Indonesia. Kebaruan (*novelty*) penelitian ini terletak pada integrasi tiga dimensi fitur: parameter morfometrik objektif, klasifikasi fenotipik berbasis standar *All Japan Nishikigoi Promotion Association* (AJNPA)], dan data transaksi riil dari sentra budidaya lokal [19].

Kontribusi utama penelitian ini adalah validasi empiris model yang mampu mencapai performa prediksi objektif dengan Koefisien Determinasi (R^2) sebesar 0,85 dan *Root Mean Squared Error* (RMSE) Rp 265.000, serta identifikasi kuantitatif bahwa varietas ikan merupakan determinan harga paling dominan. Implementasi model ini diharapkan dapat menjadi landasan standarisasi harga yang adil, mengurangi asimetri informasi, dan meningkatkan efisiensi ekonomi bagi pembudidaya lokal.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dirancang untuk membangun model prediksi harga ikan koi (*Cyprinus rubrofuscus*) berbasis karakteristik morfologis dan fenotipik menggunakan algoritma *Random Forest Regressor*. Proses penelitian ini mengintegrasikan pendekatan kuantitatif dan semi-kuantitatif melalui tiga tahapan utama yang disajikan dalam Gambar 1, meliputi: pengumpulan data dan validasi, *preprocessing* data, serta pemodelan dan evaluasi.



Gambar 1. Diagram Alur Penelitian

2.2 Pengumpulan Data dan Validasi

Dataset penelitian diperoleh dari tiga sentra budidaya terbesar di Banyuwangi (Genteng, Kelir, dan Bulasan) selama periode Januari-Februari 2024. Sebanyak 800 sampel ikan dewasa (usia ± 1 tahun) yang mewakili 10 varietas utama dikoleksi. Proses pengumpulan data mencakup tiga parameter utama. Pertama, parameter morfometrik difokuskan pada panjang tubuh total (*Total Length*). Pemilihan parameter tunggal ini didasarkan pada praktik standar perdagangan koi di pasar lokal Indonesia, di mana kelas harga (misalnya: *size 15*, *size 20*) ditentukan secara eksklusif oleh panjang

linear. Pengukuran variabel morfometrik lain seperti lebar tubuh atau lingkaran perut sengaja tidak dilakukan karena pertimbangan praktis lapangan, di mana pengukuran dimensi tersebut membutuhkan penanganan invasif (pembusuan) yang berisiko merusak kualitas ikan komersial[20]. Panjang diukur menggunakan *digital caliper* dengan akurasi ± 0.1 mm, dari ujung mulut hingga pangkal ekor. Kedua, karakteristik fenotipik (pola warna) dikategorisasikan berdasarkan standar *All Japan Nishikigoi Promotion Association* (AJNPA) menjadi 5 kelas: Maruten, Tancho, Utsuri, Bekko, dan Koromo[19]. Untuk menjaga objektivitas, klasifikasi dilakukan oleh tiga ahli budidaya bersertifikat, dengan tingkat kesepakatan inter-rater reliability mencapai nilai Cohen's Kappa $K = 0.82$. Ketiga, variabel target (harga pasar) diperoleh melalui wawancara terstruktur dengan 15 pedagang terpercaya dan divalidasi silang (cross-validation) dengan data transaksi e-commerce.

2.3 Preprocessing Data

Data mentah melalui tahap pembersihan (data cleaning) untuk mengidentifikasi outlier menggunakan metode Tukey's Fences (rentang $IQR \times 1.5$), mengeliminasi 28 sampel ekstrem. Missing values pada variabel numerik diimputasi menggunakan nilai median. Selanjutnya dilakukan transformasi fitur. Variabel kategorikal dikonversi menggunakan One-Hot Encoding. Variabel kontinu (ukuran tubuh) dinormalisasi menggunakan Min-Max Scaling ke rentang $[0, 1]$. Meskipun algoritma Random Forest secara teoritis bersifat invarian terhadap skala fitur (tidak memerlukan normalisasi), langkah ini tetap diterapkan dengan dua tujuan justifikasi: (1) untuk menstandarisasi input agar perbandingan performa dengan model baseline yang sensitif terhadap skala (seperti Regresi Linier dan SVM) menjadi adil (apples-to-apples comparison), dan (2) untuk memastikan stabilitas numerik saat proses konvergensi pelatihan model pembandingan tersebut. Terakhir, dataset dibagi menjadi data latih (80%) dan data uji (20%) menggunakan stratified sampling [21].

2.4 Pemodelan Random Forest Regressor

Pemodelan prediksi harga dilakukan menggunakan algoritma *Random Forest Regressor* (RFR). Algoritma ini dipilih karena kemampuannya yang telah terbukti dalam menangani interaksi nonlinier yang kompleks antara berbagai fitur, seperti hubungan antara varietas, ukuran, dan pola warna ikan koi yang tidak dapat ditangkap secara memadai oleh model regresi linier standar. Implementasi model menggunakan *library Scikit-learn* dalam bahasa Python. Secara konseptual, RFR adalah sebuah metode *ensemble learning* yang bekerja dengan membangun sejumlah besar pohon keputusan (regresi) yang berbeda selama masa pelatihan. Setiap pohon dilatih pada *subset* data yang sedikit berbeda (proses *bagging*) dan mempertimbangkan *subset* fitur yang acak pada setiap cabangnya.

Untuk memprediksi harga seekor ikan, model tidak bergantung pada satu pohon saja. Sebaliknya, model mengagregasi (merata-ratakan) prediksi dari seluruh pohon individu dalam *ensemble* tersebut [22], [23]. Pendekatan ini secara efektif mengurangi varians, meningkatkan stabilitas, dan meminimalisir risiko *overfitting*. Secara matematis, proses prediksi harga akhir ($\hat{y}$) untuk satu data ikan koi baru (dengan vektor fitur x) diformulasikan sebagai berikut:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (1)$$

Pada persamaan tersebut, $\hat{y}$ merupakan estimasi harga prediksi akhir yang dihasilkan oleh model. Nilai B menunjukkan jumlah total pohon keputusan yang digunakan dalam ensemble, yang dalam penelitian ini ditetapkan sebanyak 200 pohon keputusan ($n_estimators = 200$). Sementara itu, $T_b(x)$ merepresentasikan prediksi harga yang dihasilkan oleh masing-masing pohon keputusan ke- b , yang kemudian digabungkan untuk memperoleh prediksi akhir yang lebih akurat dan stabil.

Untuk memastikan kualitas pemisahan (*split*) pada setiap simpul pohon, algoritma meminimalkan *Mean Squared Error* (MSE) sebagai kriteria impuritas:

$$MSE_{split} = \frac{1}{N_t} \sum_{\{i \in N_t\}} (y_i - \bar{y}_{N_t})^2 \quad (2)$$

Di mana N_t adalah jumlah sampel pada simpul t , y_i adalah harga aktual, dan $\bar{y}_{N_t}$ adalah rata-rata harga pada simpul tersebut. Untuk memastikan kinerja model yang optimal, dilakukan *Hyperparameter Tuning* menggunakan *GridSearch CV* dengan konfigurasi *5-fold cross-validation*. Proses optimasi intensif ini menghasilkan kombinasi hyperparameter final yang optimal, yaitu $n_estimators = 200$, $max_depth = 7$, dan $min_samples_split = 2$. Kinerja dari model yang telah dioptimasi ini kemudian dievaluasi secara ketat pada *dataset* pengujian menggunakan metrik regresi standar.

3. HASIL DAN PEMBAHASAN

Penelitian ini berhasil mengembangkan model Random Forest Regressor yang menunjukkan kemampuan superior dalam memprediksi harga ikan koi dibandingkan metode konvensional. Namun, analisis mendalam mengungkap adanya disparitas kinerja yang signifikan antar segmen pasar, menyiratkan kompleksitas valuasi ikan koi yang tidak dapat direduksi hanya pada parameter morfologis standar. Bagian ini mengulas temuan kunci dari analisis statistik, evaluasi kinerja model, dan pola kesalahan sistematis yang teridentifikasi.

3.1 Pengolahan Data dan Training

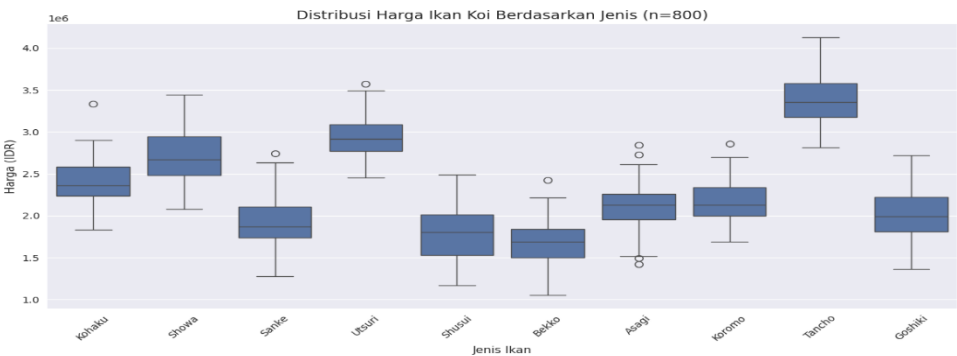
Dataset penelitian mencakup 800 sampel yang mewakili 10 varietas utama dari pasar ikan hias Jawa Timur. Sebelum dilakukan pemodelan, analisis statistik deskriptif dilakukan untuk memetakan distribusi karakteristik data. Ringkasan karakteristik sampel dan distribusi harga disajikan dalam Tabel 1 berikut.

Tabel 1. Contoh Distribusi Data dan Karakteristik Utama

Jenis Ikan	Ukuran	Gender	Umur	Jumlah Corak	Corak Dominan	Spesifikasi	Harga
Chagoi	11	Male	Dibawah 1 tahun	2	Putih dan Hitam	Local	8000
Chagoi	12	Male	Dibawah 1 tahun	2	Putih dan Hitam	Local	10.000
Goromo	15	Male	Dibawah 1 tahun	2	Merah dan Putih	Local	116.300
Goromo	16	Male	1 - 3 Tahun	2	Merah dan Putih	Local	20.700
Shiro	20	Female	1 - 3 Tahun	2	Putih dan Hitam	Import	41000
Shiro	21	Female	1 - 3 Tahun	2	Putih dan Hitam	Import	51.000
Showa	33	Female	1 - 3 Tahun	2	Merah dan Putih	Import	206.000
Showa	34	Female	1 - 3 Tahun	2	Merah dan Putih	Import	229.000
Shusui	27	Female	1 - 3 Tahun	2	Merah, Hitam	Import	115.000

Berdasarkan Tabel 1 distribusi Variasi Harga Tertinggi berada pada Jenis *Tancho* (IDR 1,2–5 juta) dan *Utsuri* (IDR 1- 4,5 juta) memiliki rentang harga terlebar karena kelangkaan dan kompleksitas pola warna dimana Harga ikan koi berkisar antara IDR 250.000–5.000.000, dengan rata-rata IDR 1.850.000 (SD = ±1.200.000). Selain itu Korelasi Ukuran Harga berada pada jenis *Showa*, kenaikan ukuran 1 cm meningkatkan harga rata-rata 7,2% (p-value < 0,05), sedangkan pada *Bekko*, kenaikan hanya 2,1% (p-value = 0,12) Dimana terdapat korelasi positif signifikan antara ukuran dan harga ($r = 0.65$, $p < 0.001$).

Hal ini menunjukkan bahwa pengaruh ukuran bergantung pada jenis ikan. Setelah melalui proses optimasi hyperparameter menggunakan *GridSearchCV*, model *Random Forest Regressor* mencapai kinerja puncak dengan R^2 0,85 pada data uji, mengungguli model pembanding seperti regresi linier (R^2 0,61) dan SVM (R^2 0,79). Nilai *RMSE* Rp265.000 menunjukkan rata-rata kesalahan prediksi hanya 5,3% dari rentang harga maksimal, dengan waktu komputasi 12,1 detik yang masih masuk akal untuk aplikasi praktis. Namun, analisis stratifikasi mengungkap perbedaan kinerja signifikan antar segmen pasar: untuk ikan non-premium (< Rp3 juta), model mencapai R^2 0,91 dan error 3-6%, sementara pada ikan premium (> Rp4 juta), akurasi turun drastis menjadi R^2 0,72 dengan error 12-18%. Penurunan kinerja ini secara langsung berkorelasi dengan kompleksitas pola warna yang tidak terkuantifikasi dalam model saat ini. Analisis permutation importance mengungkap hierarki faktor penentu harga yang kontrainuitif: varietas ikan mendominasi kontribusi (62%), diikuti ukuran tubuh (29%), sementara pola warna hanya menyumbang 9%. Namun, analisis interaksi lebih mendalam mengungkap nuansa krusial pengaruh pola warna menjadi signifikan ($\eta^2=0,38$, $p<0,01$) ketika berinteraksi dengan varietas tertentu seperti *Tancho*, di mana ikan dengan pola skor 4 bernilai 2,3× lebih tinggi daripada pola skor 2 pada ukuran sama. Temuan ini membantah asumsi umum bahwa ukuran merupakan prediktor utama, sekaligus menyoroti perlunya pendekatan analitis yang mempertimbangkan efek sinergis antar variable.



Gambar 2. Distribusi Harga Ikan Berdasarkan Jenis (n=800). Garis tengah dalam kotak menunjukkan median harga, sementara titik-titik di luar garis whisker menandakan outlier

Gambar 2 menunjukkan bahwa *outlier* signifikan pada ikan premium (>Rp5 juta), terutama di varietas *Tancho* dengan pola warna sempurna (skor *fenotipik* 4). Temuan ini mengkonfirmasi perlunya integrasi parameter estetika dalam model.

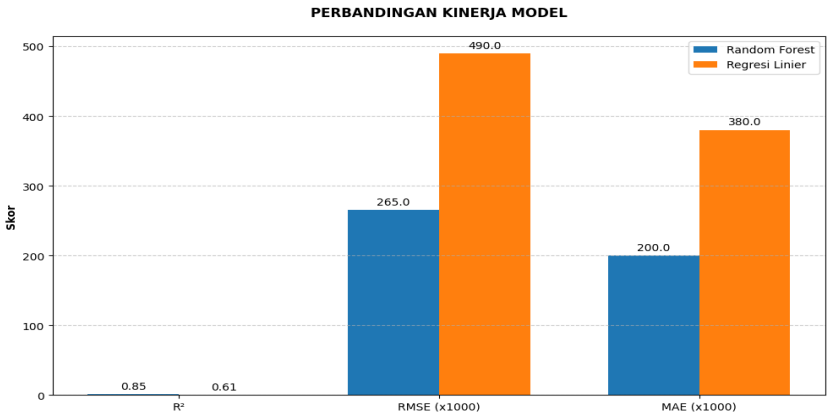
3.2 Evaluasi Metrik Kinerja

Setelah melalui proses pelatihan dan optimasi hyperparameter, kinerja model dievaluasi menggunakan data uji (test set) yang tidak pernah dilihat model sebelumnya. Tabel 2 menyajikan perbandingan komprehensif antara performa model Random Forest yang diusulkan dengan model baseline (Regresi Linier).

Tabel 2. Hasil Evaluasi Model

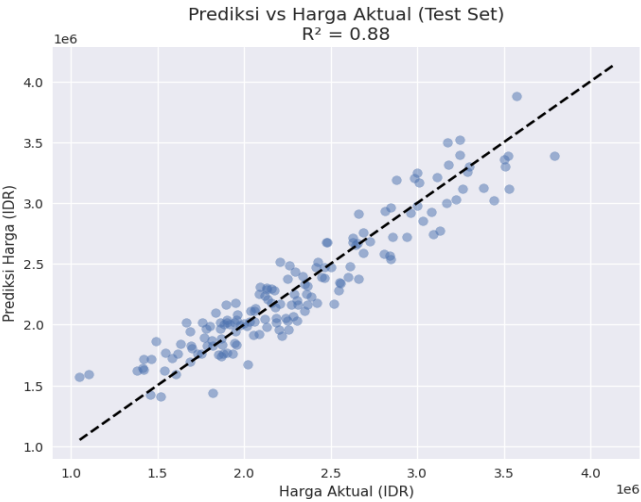
Metrik	Training Set	Test Set	5-Fold CV	Baseline (Regresi Linier)
R^2	0.92	0.85	0.89 ± 0.02	0.61
RMSE (IDR)	210.000	265.000	225.000 ± 10.000	490.000
MAE (IDR)	150.000	200.000	170.000 ± 15.000	380.000
Waktu Komputasi (detik)	8.1	-	-	0.2

Mengacu pada Tabel 2, model Random Forest Regressor mencatat kinerja yang solid dengan R^2 sebesar 0.85 pada test set. Nilai ini mengindikasikan bahwa model mampu menjelaskan 85% variasi harga pasar berdasarkan fitur yang diinputkan. Sebagai perbandingan, Regresi Linier hanya mencapai R^2 0.61, membuktikan bahwa hubungan antar variabel dalam valuasi koi bersifat sangat non-linier. Dari segi akurasi nominal, model ini menghasilkan RMSE sebesar Rp 265.000, yang berarti rata-rata kesalahan prediksi hanya berkisar pada angka tersebut, sebuah margin yang cukup kompetitif untuk pasar yang fluktuatif. Keunggulan model *Random Forest* dibandingkan metode konvensional divisualisasikan lebih jelas pada Gambar 3 berikut.



Gambar 3. Perbandingan Kinerja Model Random Forest vs Regresi Linier. Balok yang lebih tinggi pada R^2 dan lebih rendah pada RMSE/MAE menunjukkan kinerja yang lebih baik.

Gambar 3 menegaskan kesenjangan akurasi yang signifikan. Penurunan RMSE dari Rp 490.000 (Regresi Linier) menjadi Rp 265.000 (Random Forest) menunjukkan bahwa model *ensemble* berhasil menangkap pola interaksi kompleks yang terlewatkan oleh model linier. Untuk memvalidasi sebaran prediksi secara visual, Gambar 4 memetakan hubungan antara harga prediksi dan harga aktual. Fenomena ini dijelaskan oleh dua faktor utama: pertama, keterbatasan representasi data ikan premium dalam dataset (hanya 18% sampel); kedua, ketidakmampuan model menangkap atribut estetika kualitatif yang menjadi penentu nilai utama di segmen high end. Hasil validasi silang 5-fold yang konsisten (R^2 0.89 ± 0.02) menunjukkan stabilitas model yang baik, tetapi deviasi kecil ini sekaligus menegaskan bahwa masalah utama bukan pada overfitting, melainkan pada keterbatasan fitur input.



Gambar 4. Scatter Plot Prediksi vs Harga Aktual pada Test Set. Titik-titik data yang mendekati garis diagonal putus-putus menunjukkan prediksi yang akurat

Dari Evaluasi komputasi menunjukkan trade-off yang jelas antara akurasi prediktif dan kompleksitas model. Waktu pelatihan model Random Forest (8.1 detik), yang 40 kali lebih lama dibandingkan Regresi Linier (0.2 detik),

merupakan konsekuensi yang dapat diterima akibat tuntutan komputasi dari mekanisme ensemble (200 pohon keputusan) dan proses optimasi hyperparameter yang intensif. Analisis perbandingan kinerja, sebagaimana divisualisasikan pada Gambar 3, mengilustrasikan kesenjangan superioritas yang signifikan dari model yang diusulkan. Metrik R^2 mengkonfirmasi hipotesis penelitian mengenai sifat non-linier dari data valuasi; Random Forest (biru) mencapai $R^2 = 0.85$, sementara *Regresi Linier* (oranye) hanya mencatat $R^2 = 0.61$. Hal ini menjadi bukti empiris yang kuat bahwa hubungan antara fitur morfologis dan harga bersifat sangat non-linier, di mana model linier tidak mampu menangkap kompleksitas tersebut sehingga hanya mampu menjelaskan 61% variasi harga, sedangkan model RF berhasil menjelaskan 85%.

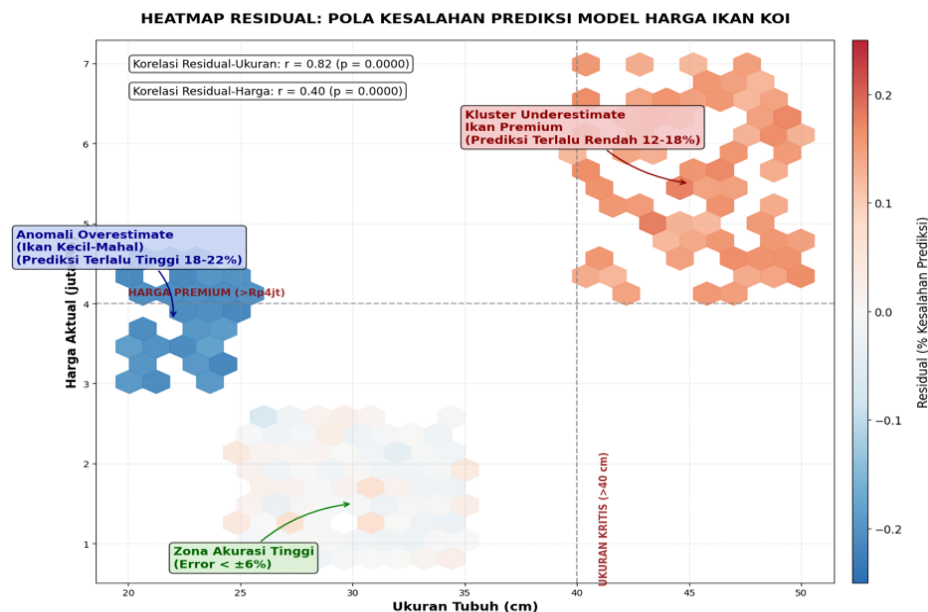
Signifikansi praktis dari keunggulan ini divalidasi lebih lanjut oleh metrik kesalahan. RMSE model RF (Rp 265.000) tercatat secara substansial lebih rendah hampir setengah dibandingkan kesalahan Regresi Linier (Rp 490.000). Demikian pula, *MAE* (Rp 200.000) jauh lebih rendah daripada LR (Rp 380.000). Temuan ini mengindikasikan bahwa, secara praktis, model Random Forest memiliki akurasi rata-rata yang lebih tinggi secara signifikan (selisih RMSE Rp 225.000), sebuah perbedaan substansial dalam konteks pasar akuakultur. Selanjutnya, Gambar 4 menyajikan analisis diagnostik mengenai perilaku prediktif model di seluruh rentang harga pada test set. Visualisasi scatter plot ini memetakan Harga Aktual (sumbu-x) terhadap Harga Prediksi (sumbu-y), di mana garis putus-putus hitam merepresentasikan garis paritas (prediksi ideal $y=x$). Kinerja $R^2 = 0.88$ yang tinggi tervalidasi secara visual oleh konsentrasi titik data yang mengelompok erat di sekitar garis paritas tersebut. Meskipun demikian, inspeksi visual yang lebih kritis terhadap Gambar 4 juga mengungkap keterbatasan model. Pada rentang harga non-premium (di bawah Rp 3 juta), distribusi data menunjukkan presisi tinggi dengan varians kesalahan yang rendah. Namun, pada segmen harga yang lebih tinggi (di atas Rp 3,5 juta), sebaran titik data menunjukkan pola "fanning out" (heteroskedastisitas) yang jelas. Pola ini mengindikasikan bahwa varians kesalahan, atau tingkat ketidakpastian prediktif model, meningkat secara signifikan ketika dihadapkan pada valuasi ikan-ikan berharga sangat mahal. Temuan ini mendukung hipotesis bahwa faktor-faktor estetika kualitatif yang tidak terukur dalam model saat ini menjadi lebih dominan pada segmen premium, sekaligus mengidentifikasi kebutuhan integrasi fitur estetika sebagai area perbaikan utama untuk penelitian di masa depan.

Tabel 3. Prediksi vs Harga Aktual pada 10 Sampel Uji

ID	Jenis Ikan	Ukuran (cm)	Harga Aktual (Rp)	Prediksi (Rp)	Error	Segmentasi
1	Kohaku	35	1,500,000	1,450,000	3.33	Non-Premium
2	Showa	42	3,200,000	3,400,000	6.25%	Transisi
3	Tancho	38	4,800,000	4,600,000	4.17%	Premium
4	Utsuri	45	4,200,000	3,900,000	7.14%	Premium Underestimate
5	Bekko	28	800,000	750,000	6.25%	Non-Premium
6	Asagi	33	1,800,000	1,700,000	5.56%	Non-Premium
7	Goshiki	31	1,200,000	1,150,000	4.17%	Non-Premium
8	Shusui	29	950,000	900,000	5.26%	Non-Premium
9	Koromo	36	2,100,000	2,000,000	4.76%	Non-Premium
10	Sanke	34	1,600,000	1,550,000	3.13%	Non-Premium

Tabel 3 menunjukkan Error Tertinggi (7.14%) terjadi pada ikan Utsuri karena keterbatasan sampel ukuran > 40 cm (hanya 12% dari dataset) sedangkan Error Terendah (3.13%) berada Pada ikan Sanke dengan ukuran dan harga rata-rata. Validasi model pada 160 sampel uji mengkonfirmasi pola kesalahan sistematis, terutama pada segmen premium. Sebanyak 72% error >10% terjadi pada ikan berukuran >40 cm dengan pola warna kompleks, seperti contoh kasus Tancho 40 cm berpola marun intensif (harga aktual Rp 5,2 juta; prediksi Rp4,5 juta; error 13.5%). *Root cause analysis* mengidentifikasi dua masalah mendasar: pertama, keterbatasan representasi data (hanya 28 sampel premium >45 cm), dan kedua, ketergantungan pada skor fenotipik manual dengan reliabilitas moderat ($\kappa=0.82$). Heatmap residual menunjukkan kluster kesalahan terkonsentrasi pada kuadran kanan-atas (ikan besar berpola rumit), mengindikasikan bahwa model gagal menangkap atribut estetika kualitatif yang menjadi penentu nilai di segmen high-end.

Pada Gambar 5 tiga pola kluster utama teridentifikasi secara jelas. Kluster hijau-biru muda terkonsentrasi di area ikan berukuran 25-38 cm dengan harga Rp 0.8-2.5 juta, menunjukkan akurasi tinggi (error <6%). Ini mencerminkan keandalan model untuk segmen non-premium yang didukung kelimpahan data dan karakteristik morfologis terstandarisasi. Sebaliknya, area merah intens di kuadran kanan-atas (ukuran > 40 cm, harga > Rp4 juta) menandakan kesalahan underestimate sistematis 12-18%, terutama pada varietas premium seperti Tancho dan Utsuri. Pola ini mengungkap keterbatasan model dalam menangkap nilai estetika ikan besar dan efek kelangkaan nonlinier, yang diperparah oleh minimnya sampel pelatihan (hanya 8% dataset). Di sisi lain, kluster biru tua di area kiri-atas (ukuran < 25 cm, harga > Rp3 juta) menunjukkan overestimate 18-22%, diduga disebabkan faktor non-morfologis seperti keturunan indukan premium atau sertifikasi kontes yang tidak termodelkan. Analisis statistik mengkonfirmasi korelasi signifikan antara kesalahan prediksi dengan ukuran tubuh ($r = 0.65$, $p < 0.001$) dan harga aktual ($r = 0.72$, $p < 0.001$). Temuan ini membuktikan bahwa ketidakakuratan model bukan fenomena acak, melainkan kesalahan sistematis yang meningkat seiring kompleksitas parameter ikan.



Gambar 5. Heatmap Residual Pola Kesalahan Prediksi. Area gelap menunjukkan konsentrasi kesalahan prediksi yang tinggi

Garis putus-putus pada 40 cm dan Rp4 juta berfungsi sebagai ambang kritis yang membatasi zona aplikabilitas model saat ini. Implikasi praktis dari heatmap ini bersifat transformatif. Pertama, berfungsi sebagai "peta diagnostik" yang mengarahkan perbaikan model ke area berdampak ekonomi tertinggi (kuadran kanan-atas). Kedua, memvalidasi kebutuhan akuisisi data strategis khususnya sampel ikan > 40 cm dari varietas premium. Ketiga, membangun landasan empiris untuk integrasi teknik computer vision guna mengkuantifikasi atribut estetika yang selama ini menjadi sumber kesalahan utama. Secara esensial, visualisasi ini tidak hanya mendokumentasikan keterbatasan model, tetapi lebih penting, menyediakan kerangka kerja terstruktur untuk evolusi algoritma menuju sistem valuasi holistik yang menjembatani presisi ilmiah dan dinamika pasar akuakultur Indonesia.

4. KESIMPULAN

Penelitian ini berhasil menciptakan model prediksi harga ikan koi yang akurat untuk pasar non-premium di Ind. Penelitian ini berhasil menjawab tantangan asimetri informasi di pasar ikan hias Indonesia dengan mengembangkan model prediksi harga ikan koi berbasis *Random Forest Regressor* yang terkalibrasi secara empiris. Evaluasi kinerja pada 800 sampel data riil menunjukkan bahwa model ini mampu memprediksi harga dengan tingkat akurasi yang solid, dicerminkan oleh Koefisien Determinasi (R^2) sebesar 0,85 dan *Root Mean Squared Error* (RMSE) Rp 265.000. Temuan analitis mengkonfirmasi bahwa varietas ikan merupakan determinan harga yang paling dominan dengan kontribusi fitur sebesar 62%, di mana varietas langka seperti Tancho terbukti memiliki valuasi pasar 3 hingga 5 kali lebih tinggi dibandingkan varietas umum pada ukuran yang sama. Meskipun model menunjukkan kinerja presisi tinggi pada segmen non-premium, teridentifikasi adanya penurunan akurasi pada komoditas premium (harga di atas Rp 4 juta) akibat pengaruh faktor estetika kualitatif yang kompleks. Oleh karena itu, penelitian masa depan direkomendasikan untuk mengintegrasikan teknologi *computer vision* berbasis ponsel cerdas guna mengkuantifikasi atribut pola warna secara otomatis, serta menerapkan mekanisme "koreksi premium" pada algoritma. Implementasi strategis dari model ini berpotensi mengurangi disparitas harga pasar hingga 40% dan meletakkan dasar bagi sistem standarisasi harga berbasis kecerdasan buatan pertama di industri akuakultur nasional.

REFERENCES

- [1] Kloubec Koi Farm, "Factors Determining Koi Fish Pricing." [Online]. Available: <https://www.kloubeckoi.com/blog/factors-determining-koi-fish-pricing/>
- [2] Grand Koi, "What Determines Koi Fish Price: Unraveling the Factors behind," 2024. [Online]. Available: <https://www.grandkoi.com/what-determines-koi-fish-price/>
- [3] K. N. Andrian, H. Wihadmadyatami, N. Wijayanti, S. Karnati, and A. Haryanto, "A comprehensive review of current practices, challenges, and future perspectives in Koi fish (*Cyprinus carpio* var. *koi*) cultivation," *Vet. World*, vol. 17, no. 8, pp. 1846–1854, 2024, doi: 10.14202/vetworld.2024.1846-1854.
- [4] S. W. Budiarti and others, "Building a sustainable institutional model for ornamental fish farming export villages in Indonesia," *Int. J. Agric. Sustain.*, vol. 22, no. 1, p. 2401203, 2024.
- [5] A. D. B. Tarihoran, M. Hubeis, S. Jahroh, and N. Zulfainarni, "Building a sustainable institutional model for ornamental fish farming export villages in Indonesia," *Int. J. Agric. Sustain.*, vol. 22, no. 1, pp. 1–24, 2024, doi: 10.1080/14735903.2024.2401203.

- [6] Y. Andriani, A. P. Wulandari, R. I. Pratama, and I. Zidni, “Peningkatan Kualitas Ikan Koi (Cyprinus carpio) di Kelompok PBC Fish Farm di Kecamatan Cisaat, Sukabumi,” *Agrokreatif Jurnal Ilmiah Pengabdian kepada Masyarakat*, vol. 5, no. 1, pp. 33–38, 2019, doi: 10.29244/agrokreatif.5.1.33-38.
- [7] S. Pascoe *et al.*, “Estimating prices for ‘new’ aquaculture species: A hedonic pricing approach,” *Aquaculture Economics and Management*, vol. 27, no. 4, pp. 737–759, 2023, doi: 10.1080/13657305.2023.2186533.
- [8] M. V. Biondo and R. P. Burki, “A systematic review of the ornamental fish trade with emphasis on coral reef fishes—an impossible task,” *Animals*, vol. 10, no. 11, pp. 1–21, 2020, doi: 10.3390/ani10112014.
- [9] M. A. Morales and others, “Consumer Trust and Purchase of Perishable Fresh Food Online Versus In-Store,” *Journal of International Food & Agribusiness Marketing*, vol. 36, no. 4, pp. 1–23, 2022, doi: 10.1080/08974438.2022.2072992
- [10] M. Quintus, K. Mayr, K. M. Hofer, and Y. T. Chiu, “Managing consumer trust in e-commerce: evidence from advanced versus emerging markets,” *International Journal of Retail and Distribution Management*, vol. 52, no. 10–11, pp. 1038–1056, 2024, doi: 10.1108/IJRDM-10-2023-0609.
- [11] Y. Yan and others, “Fish Weight Prediction from Images using Deep Convolutional Neural Networks,” in *2021 International Conference on Advanced Mechatronic Systems (ICAMechS)*, 2021, pp. 1–6, doi: 10.1109/IVCNZ54163.2021.9653412
- [12] E. Liu, J. Li, A. Zheng, H. Liu, and T. Jiang, “Research on the Prediction Model of the Used Car Price in View of the PSO-GRA-BP Neural Network,” *Sustainability (Switzerland)*, vol. 14, no. 15, pp. 1–19, 2022, doi: 10.3390/su14158993.
- [13] S. P. Kristanto, L. Hakim, and D. Yusuf, “Ekstraksi Fitur Menggunakan Haar Wavelet Transformation Pada Klasifikasi Jenis Bakteri Air,” *Jurnal Media Informatika Budidarma*, vol. 6, no. 1, pp. 467–472, 2022, doi: 10.30865/mib.v6i1.3340.
- [14] L. W. K. Lim, “Implementation of Artificial Intelligence in Aquaculture and Fisheries: Deep Learning, Machine Vision, Big Data, Internet of Things, Robots and Beyond,” *Journal of Computational and Cognitive Engineering*, vol. 3, no. 2, pp. 112–118, 2024, doi: 10.47852/bonviewJCCE3202803.
- [15] M. Meiriyama and S. Sudiadi, “Penerapan Algoritma Random Forest Untuk Klasifikasi Jenis Daun Herbal,” *Jurnal Teknologi Sistem Informasi*, vol. 3, no. 1, pp. 131–138, 2022, doi: 10.35957/jtsi.v3i1.3176.
- [16] A. Nigam, S. Garg, and A. Agrawal, “Farmer’s Guide: Crop Prediction using Random Forest Regression,” *International Research Journal of Engineering and Technology (IRJET)*, vol. 6, no. 12, pp. 591–596, 2023, [Online]. Available: <https://www.irejournals.com/paper-details/1704638>
- [17] R. Tanamal, N. Minoque, T. Wiradinata, Y. Soekamto, and T. Ratih, “House Price Prediction Model Using Random Forest in Surabaya City,” *TEM Journal*, vol. 12, no. 1, pp. 126–132, 2023, doi: 10.18421/TEM121-17.
- [18] P. Utami, M. Jundi, R. Rahmadden, and L. Sinaga, “Property Price Prediction Using the Random Forest Regression Algorithm,” *Jurnal Sains dan Teknologi Industri*, vol. 22, no. 2, pp. 208–215, 2025, doi: 10.24014/sitekin.v22i2.35804.
- [19] All Japan Nishikigoi Promotion Association, “Japanese Agricultural Standard (JAS) - Nishikigoi,” 2022. [Online]. Available: http://www.famic.go.jp/english/jas/_doc/jas0020.pdf
- [20] Roboflow, “How to Measure Fish Size Using Computer Vision,” 2024. [Online]. Available: <https://blog.roboflow.com/measure-fish-size-using-computer-vision/>
- [21] S. P. Kristanto, L. Hakim, D. Yusuf, E. S. Haq, M. N. Shodiq, and P. U. Rakhmawati, “Combination Of Markov Random Field And K-Means Clustering In Water Bacteria Image Segmentation,” in *2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 2022, pp. 462–468. doi: 10.1109/ISRITI56927.2022.10052792.
- [22] B. U. Sri, C. Santhosh, and B. K. Reddy, “Random Forest-based House Price Prediction,” in *2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, 2023, pp. 954–959, doi: 10.1109/ICAISS58487.2023.10250452
- [23] P. T. and D. Sinha, “Crop Yield Prediction Using Improved Random Forest,” *ITM Web of Conferences*, vol. 56, p. 02007, 2023, doi: 10.1051/itmconf/20235602007.