ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

IMPLEMENTASI *XGBOOST* DENGAN *HYPERPARAMETER TUNING OPTUNA* UNTUK ESTIMASI HARGA AKOMODASI DI BALI BERBASIS MOBILE

I Gede Igo Wisnu Wardana✉, I Wayan Candra Winetra,
Ida Bagus Adisimakrisna Peling, Ni Gusti Ayu Putu Harry Saptarini
Jurusan Teknologi Informasi, Politeknik Negeri Bali, Badung, Indonesia
Email: igowisnuwardana@gmail.com

DOI: https://doi.org/10.46880/jmika.Vol10No2.pp717-728

ABSTRACT

The recovering tourism industry in Bali has triggered a surge in accommodation demand, accompanied by daily rental price inconsistency and disparity due to subjective pricing mechanisms. Previous studies mostly predicted long-term property prices using standard models, leaving a gap in exploring optimized machine learning for dynamic daily rental characteristics and real-time mobile implementations. Therefore, this study contributes to developing an objective pricing model by implementing the Extreme Gradient Boosting (XGBoost) algorithm optimized with the Optuna framework. The modeling utilized location and facility variables acquired from an Online Travel Agent. The model's performance was evaluated statistically by R^2, RMSE, and MAE, and application acceptance was validated through a User Acceptance Test (UAT). The results showed the optimized XGBoost model achieved high accuracy with an R^2 of 0.797, RMSE of 0.3752, and MAE of 0.3007. The system's implementation into a mobile application recorded an acceptance rate of 89.2% (Very Good). Theoretically, this study enriches the data engineering literature for tourism property price estimation. Practically, the system is highly applicable as an objective pricing reference instrument to minimize price disparity for both property managers and tourists.

Keyword: Accommodation Price, Machine Learning, XGBoost, Hyperparameter Tuning, Optuna.

ABSTRAK

Industri pariwisata Bali yang kembali pulih memicu lonjakan permintaan akomodasi, namun hal ini turut menimbulkan permasalahan berupa ketidakpastian dan disparitas harga sewa harian akibat mekanisme penentuan harga yang subjektif. Penelitian terdahulu umumnya memprediksi harga properti jangka panjang menggunakan pemodelan standar, namun masih jarang mengeksplorasi optimasi machine learning untuk karakteristik dinamis harga sewa harian pariwisata serta belum adanya implementasi real-time berbasis mobile. Oleh karena itu, penelitian ini berkontribusi untuk mengembangkan pemodelan harga yang objektif dengan mengimplementasikan algoritma Extreme Gradient Boosting (XGBoost) yang dioptimasi menggunakan kerangka kerja Optuna. Pemodelan menggunakan variabel lokasi dan fasilitas yang diakuisisi dari Online Travel Agent. Kinerja model dievaluasi secara statistik dengan R^2, RMSE, dan MAE, dan tingkat penerimaan aplikasi divalidasi melalui User Acceptance Test (UAT). Hasil penelitian menunjukkan model XGBoost yang dioptimasi mencapai tingkat akurasi tinggi dengan R^2 sebesar 0,797, RMSE 0,3752, dan MAE 0,3007. Implementasi sistem ke dalam aplikasi mobile mencatatkan tingkat penerimaan sebesar 89,2% (Sangat Baik). Secara teoretis, penelitian ini memperkaya literatur rekayasa data untuk estimasi harga properti pariwisata. Secara praktis, sistem ini aplikatif menjadi instrumen acuan harga yang objektif untuk meminimalkan ketimpangan harga bagi pengelola properti maupun wisatawan.

Kata Kunci: Harga Akomodasi, Machine Learning, XGBoost, Hyperparameter Tuning, Optuna.

PENDAHULUAN

Pulau Bali sebagai destinasi pariwisata bertaraf internasional kini mengalami pemulihan ekonomi yang signifikan pasca-pandemi. Hal ini ditandai dengan lonjakan kunjungan wisatawan domestik maupun mancanegara yang berdampak langsung pada peningkatan permintaan akomodasi. Tingkat Penghunian Tempat Tidur (TPT) untuk akomodasi hotel non-bintang mencapai rata-rata 62,37% per tahun (Badan Pusat Statistik Provinsi Bali, 2025). Lonjakan permintaan yang signifikan terlihat pada masa puncak liburan, dimana tingkat penghunian pada bulan Juli dan Agustus masing-masing menembus angka 70,78% dan 72,36% (Badan Pusat Statistik Provinsi Bali, 2025).

Namun, di balik tingginya permintaan pasar tersebut, muncul sebuah permasalahan berupa

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ketidakpastian harga pada industri akomodasi. Selama ini, mekanisme penentuan harga sewa per malam sering kali dilakukan secara subjektif oleh pengelola tanpa adanya standar indikator yang terukur. Akibatnya, kerap terjadi disparitas harga yang merugikan keseimbangan pasar.

Salah satu algoritma yang terbukti handal dalam menangani kasus regresi pada data properti adalah *Extreme Gradient Boosting (XGBoost)*. Keunggulan algoritma ini diperkuat oleh penelitian yang dilakukan oleh (Fauzi et al., 2025), yang membandingkan performa berbagai algoritma untuk prediksi harga rumah. Dalam penelitian tersebut, disimpulkan bahwa XGBoost mampu menghasilkan tingkat akurasi terbaik dengan nilai *Root Mean Square Error (RMSE)* yang lebih rendah dibandingkan algoritma *Random Forest* dan *Linear Regression*, terutama pada dataset yang memiliki banyak fitur kategorikal seperti fasilitas hunian.

Meskipun algoritma XGBoost terbukti andal dalam regresi properti, tinjauan terhadap studi terdahulu mengungkapkan dua keterbatasan yang signifikan. Keterbatasan pertama bersifat kontekstual, dimana mayoritas pemodelan harga properti masih terpaku pada aset statis berjangka panjang, seperti estimasi harga jual rumah atau tanah. Pendekatan ini mengabaikan volatilitas dan karakteristik dinamis dari harga sewa akomodasi harian di kawasan pariwisata, yang sensitif terhadap kombinasi fitur spesifik seperti keberadaan fasilitas privat mencakup private pool, tipe pemandangan, dan geolokasi mikro. Keterbatasan kedua bersifat metodologis, penelitian berbasis machine learning sebelumnya umumnya mengandalkan parameter bawaan atau teknik pencarian konvensional seperti *Grid Search* dan *Random Search*. Pada dataset berdimensi tinggi hasil *web scraping*, metode konvensional tersebut memakan beban komputasi yang masif dan rentan mengalami *overfitting*.

Untuk menjembatani celah penelitian tersebut, penelitian ini mengusulkan kebaruan metodologis dan teoretis. Secara metodologis, kebaruan penelitian ini terletak pada integrasi pemrosesan fitur akomodasi pariwisata yang kompleks dengan algoritma XGBoost yang dioptimasi secara adaptif menggunakan kerangka kerja Optuna berbasis *Tree-structured Parzen Estimator (TPE)*. Pendekatan ini dikembangkan untuk menemukan kombinasi *hyperparameter* paling optimal dengan beban komputasi yang efisien guna menekan tingkat error pada data *scraping*. Sementara itu, kebaruan teoretis dan aplikatifnya diwujudkan melalui pergeseran fokus kajian data science ke ranah estimasi harga akomodasi harian pariwisata yang fluktuatif, serta implementasinya ke dalam arsitektur sistem real-time berbasis mobile.

Melalui kebaharuan ini, penelitian diharapkan tidak hanya menghasilkan luaran berupa rekayasa sistem perangkat lunak, tetapi juga memberikan kontribusi ilmiah berupa pengayaan literatur empiris terkait implementasi machine learning untuk menekan disparitas harga. Sistem yang dikembangkan diharapkan mampu menyajikan estimasi harga sewa yang akurat sebagai instrumen objektif bagi pemilik akomodasi, sekaligus menjadi acuan referensi yang adil bagi wisatawan.

TINJAUAN PUSTAKA

Akomodasi dan Karakteristik Properti Bali

Akomodasi dalam konteks pariwisata didefinisikan sebagai kesepakatan pemindahan hak guna sementara atas akomodasi dari pengelola kepada wisatawan (Megawati et al., 2024), di mana keputusan ini sangat ditentukan oleh persepsi terhadap harga, lokasi strategis, dan kualitas pelayanan. Dalam konteks pariwisata Bali, properti sewa memiliki karakteristik unik di mana harga sangat dipengaruhi oleh faktor lokasi, musim liburan, dan fasilitas penunjang gaya hidup seperti kolam renang pribadi atau akses internet

Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting atau *XGBoost* adalah algoritma machine learning berbasis *ensemble* yang menggunakan prinsip *Gradient Boosting Framework*. Algoritma ini bekerja dengan cara menggabungkan prediksi dari beberapa model sederhana (*weak learners*), biasanya berupa *decision trees*, untuk membentuk model yang lebih kuat (*strong learner*). Keunggulan utama *XGBoost* terletak pada kemampuannya melakukan komputasi paralel, penanganan *overfitting* melalui mekanisme regularisasi, dan efisiensi tinggi dalam menangani dataset berskala besar dengan tipe data campuran (*numerik dan kategorikal*).

Arsitektur dasar *XGBoost* merupakan sekumpulan *Classification and Regression Trees* (CART). Dalam arsitektur ini, prediksi akhir ($\hat{y}_i$) untuk sebuah input data diperoleh dengan menjumlahkan bobot prediksi dari seluruh pohon (K) yang ada dalam model *ensemble* (Chen & Guestrin, 2016), sebagaimana dirumuskan berikut:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F}$$

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Di mana f_k adalah fungsi independen dari setiap pohon, dan $\mathcal{F}$ adalah ruang fungsi dari seluruh pohon *CART*. Berbeda dengan pohon keputusan biasa yang hanya memberikan nilai kelas, setiap daun (*leaf*) pada *CART* dalam *XGBoost* memiliki bobot nilai (*score*) yang kontinu. Berdasarkan (Chen & Guestrin, 2016), langkah kerja *XGBoost* dalam membangun model mengikuti prinsip *Additive Training* dengan optimasi fungsi objektif sebagai berikut:

a. Fungsi Objektif Ter-regularisasi (*Regularized Objective Function*)

Kunci utama keunggulan *XGBoost* dibanding *Gradient Boosting* biasa terletak pada fungsi objektifnya. *XGBoost* tidak hanya meminimalkan *loss* (selisih error), tetapi juga menambahkan komponen regularisasi (Ω) untuk mengontrol kompleksitas model dan mencegah *overfitting*:

$$Obj(\Theta) = \sum_i l\,(\hat{y}_i, y_i) + \sum_k \Omega\,(f_k)$$

b. Pelatihan Adatif

XGBoost tidak melatih semua pohon secara bersamaan, melainkan satu per satu secara berurutan (*sequential*). Pada setiap langkah iterasi ke-t, algoritma menambahkan pohon baru yang paling mampu memperbaiki kesalahan dari pohon sebelumnya (Chen & Guestrin, 2016)

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i)$$

c. Aproksimasi Taylor dan Penghitungan Gradien

Untuk mempercepat proses optimasi, *XGBoost* menggunakan ekspansi *Taylor* orde kedua. Algoritma ini menghitung nilai Gradien (g_i, turunan pertama) dan Hessian (h_i, turunan kedua) dari fungsi loss untuk menentukan arah penurunan error yang paling cepat dan akurat.

d. Pencarian Split Terbaik (*Exact Greedy Algorithm*)

Dalam membangun setiap pohon, *XGBoost* mencari titik pemisahan terbaik di setiap simpul dengan menghitung *Gain* (keuntungan informasi). Algoritma akan mengevaluasi semua kemungkinan pemisahan fitur dan memilih pemisahan yang memberikan penurunan nilai loss terbesar setelah dikurangi biaya regularisasi. Algoritma *XGBoost* tidak menyelesaikan persamaan tersebut sekaligus, melainkan mengeksekusinya melalui pendekatan iteratif (*looping*) untuk membangun model secara bertahap (*Additive Training*). Untuk memberikan gambaran yang lebih konkret mengenai bagaimana teori tersebut diimplementasikan, Gambar berikut memvisualisasikan alur kerja algoritma *XGBoost* mulai dari inisialisasi data hingga terbentuknya model prediksi akhir.

Web Scraping

Web Scraping didefinisikan sebagai teknik ekstraksi data secara otomatis dari halaman *website* yang melibatkan proses pengunduhan dokumen *web* dan pengambilan elemen data spesifik seperti harga, nama produk, dan deskripsi berdasarkan struktur HTML DOM (Rizquina & Ratnasari, 2023). Dalam konteks penelitian ini, teknik ini digunakan sebagai metode akuisisi data sekunder yang efisien untuk mengumpulkan informasi dari situs *Online Travel Agent (OTA)*, yang kemudian dibersihkan dan distrukturisasi menjadi dataset latih (*training data*) untuk keperluan pemodelan *machine learning*.

Data Cleaning

Data Cleaning adalah proses mendeteksi dan mengoreksi atau menghapus catatan yang rusak, tidak akurat, atau tidak relevan dari sebuah set data (Han et al., 2011). Dalam konteks data properti, proses ini melibatkan penanganan:

1. Missing Values, Menangani data yang kosong dengan cara menghapus baris data tersebut atau melakukan imputasi/pengisian nilai menggunakan rata-rata atau median.
2. Noise Data, Menghilangkan karakter yang tidak diperlukan yang dapat mengganggu proses analisis numerik, seperti simbol mata uang ("Rp"), tanda baca, atau satuan ukuran yang tercampur dalam kolom angka.

***Encoding* Data Kategorikal**

Encoding data kategorikal adalah proses transformasi variabel yang memiliki nilai kualitatif (berupa teks, label, atau string) menjadi representasi numerik yang dapat dipahami oleh algoritma pembelajaran mesin (Han et al., 2011). Proses *encoding* berfungsi agar algoritma dapat memproses fitur-fitur non-numerik seperti "Tipe Properti" atau "Nama Fasilitas".

Hyperparameter* dan *Hyperparameter Tuning

Menurut Claesen dan De Moor (Claesen & De Moor, 2015), *model parameters* adalah konfigurasi internal yang nilainya diestimasi secara otomatis dari data selama proses pelatihan berlangsung (misalnya bobot atau weights pada regresi linear). Sebaliknya, *Hyperparameter* adalah variabel konfigurasi eksternal yang nilainya tidak dapat dipelajari dari data latih,

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

melainkan harus ditetapkan secara manual oleh peneliti sebelum proses pelatihan dimulai.

Hyperparameter Tuning atau optimasi *hyperparameter* adalah proses sistematis untuk mencari kombinasi nilai *hyperparameter* terbaik yang dapat menghasilkan kinerja model paling optimal. Proses ini pada dasarnya adalah masalah optimasi, di mana tujuannya adalah meminimalkan fungsi kerugian (*loss function*) atau memaksimalkan akurasi pada data validasi (Bergstra & Bengio, 2012).

Mobile Application Development (Flutter)

Aplikasi *mobile* didefinisikan sebagai perangkat lunak yang dirancang secara spesifik untuk beroperasi pada perangkat seluler pintar. Dalam pengembangan sistem ini, digunakan *Flutter*, sebuah *framework open-source* yang dikembangkan oleh Google yang memungkinkan pembuatan aplikasi *multi-platform* hanya dengan menggunakan satu basis kode (Kamil et al., 2023). *Flutter* menggunakan bahasa pemrograman *Dart* dan dikenal memiliki keunggulan performa tinggi karena menerapkan mekanisme kompilasi *Ahead-of-Time (AOT)* menjadi kode mesin *(native code)* serta menggunakan mesin rendering grafis sendiri *(Skia Engine)* yang menjamin tampilan antarmuka tetap responsif dan halus (Wicaksono et al., 2025).

Mean Absolute Error (MAE)

Mean Absolute Error (MAE) adalah metrik evaluasi yang digunakan untuk mengukur rata-rata kesalahan absolut antara nilai hasil prediksi model dengan nilai aktual yang sebenarnya. Dalam perhitungan *MAE*, semua selisih kesalahan diperlakukan sama tanpa memperhitungkan arah kesalahan (positif atau negatif) karena menggunakan nilai mutlak (Hodson, 2022). *MAE* dinilai lebih mudah diinterpretasikan karena satuan nilainya sama dengan satuan data asli (misalnya Rupiah). Rumus *MAE* didefinisikan sebagai:

$$MAE = \frac{1}{n}\sum_{i=1}^{\{n\}}|y_i - \hat{y}_i|$$

Keterangan:

n = Jumlah total data observasi.
y_i = Nilai aktual (harga sewa sebenarnya).
y^_i = Nilai prediksi model.
|…| = Simbol nilai mutlak (absolute value).

Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) adalah akar kuadrat dari rata-rata kuadrat selisih antara nilai prediksi dan nilai aktual. Berbeda dengan *MAE, RMSE* memberikan bobot yang lebih besar pada kesalahan yang bernilai besar karena proses pengkuadratan sebelum dirata-ratakan (Hodson, 2022). Hal ini membuat *RMSE* sangat sensitif terhadap *outliers* atau data pencilan yang ekstrem. Oleh karena itu, *RMSE* sangat efektif digunakan jika tujuan model adalah untuk menghindari kesalahan prediksi yang fatal atau sangat jauh dari nilai sebenarnya. Rumus *RMSE* adalah:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Coefficient of Determination (R² Score)

Koefisien Determinasi (R^2) merupakan metrik statistik yang digunakan untuk mengukur sejauh mana variabilitas atau proporsi varians dari variabel dependen (target harga) dapat dijelaskan secara akurat oleh fitur-fitur independen yang ada di dalam model prediksi. Nilai R^2 berkisar antara 0 hingga 1. Nilai R^2 yang mendekati angka 1 mengindikasikan bahwa model memiliki kemampuan generalisasi pola data yang sangat baik dan mampu memetakan harga aktual dengan tingkat kecocokan yang tinggi. Rumus perhitungan R^2 dinyatakan sebagai berikut

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Pengujian Penerimaan Pengguna (*User Acceptance Test*)

User Acceptance Test (UAT) atau pengujian penerimaan pengguna adalah tahapan akhir dalam siklus hidup pengembangan perangkat lunak (*Software Development Life Cycle*) sebelum sistem secara resmi diserahkan atau diluncurkan ke lingkungan produksi. Menurut (Pressman, 2010), *UAT* berbeda dengan pengujian sistem yang dilakukan oleh pengembang. *UAT* dilakukan langsung oleh pengguna akhir (*end-user*) atau klien untuk memvalidasi apakah sistem yang dibangun telah sesuai dengan kebutuhan bisnis dan dapat digunakan dalam skenario dunia nyata.

Penelitian Terdahulu

Beberapa penelitian terdahulu telah mengkaji secara mendalam penerapan berbagai algoritma *Machine Learning* dalam memprediksi harga di sektor properti. Penelitian yang dilakukan oleh (Fauzi et al., 2025) melakukan analisis perbandingan kinerja algoritma *Linear Regression*, *Random Forest*, dan *XGBoost* untuk memprediksi harga rumah berdasarkan variabel fisik seperti luas tanah, jumlah kamar serta faktor lokasi geografis. Hasil pengujian pada penelitian tersebut menunjukkan bahwa model *Linear Regression*

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

memberikan performa prediksi yang paling presisi dengan nilai *Root Mean Square Error (RMSE)* terendah. Peneliti menyimpulkan bahwa dalam kasus dataset tersebut, hubungan antara variabel fisik bangunan dan lokasi dengan harga rumah lebih cocok dimodelkan menggunakan pendekatan linier yang lebih sederhana dibandingkan algoritma ansambel.

Di sisi lain, penelitian yang lebih spesifik mengenai algoritma ansambel dilakukan oleh (Sari & Prabowo, 2025) yang membandingkan performa *Random Forest*, *Gradient Boosting*, dan *XGBoost* untuk memprediksi harga rumah di wilayah Yogyakarta menggunakan data tabular. Berbeda dengan temuan Fauzi, dkk., studi ini menemukan bahwa algoritma *XGBoost* menghasilkan performa yang paling unggul dengan nilai koefisien determinasi (R^2) tertinggi mencapai 0,7746 dan tingkat error yang paling rendah. Keunggulan *XGBoost* pada studi ini didasari oleh efektivitasnya dalam menangani interaksi fitur yang kompleks pada data properti serta kemampuannya mengendalikan *overfitting* melalui teknik regularisasi bawaan.

Meskipun kedua penelitian di atas telah memberikan wawasan yang komprehensif mengenai performa algoritma *XGBoost* dan regresi lainnya di sektor properti, terdapat beberapa celah penelitian yang belum terselesaikan. Kedua studi terdahulu berfokus murni pada prediksi harga aset properti permanen berjangka panjang, yakni rumah tinggal. Belum ada kajian yang menyentuh dinamika harga sewa harian pada akomodasi pariwisata spesifik (seperti hotel, villa, dan resort) di wilayah Bali.

Pemodelan *XGBoost* pada penelitian sebelumnya umumnya mengandalkan parameter bawaan atau penyesuaian dasar. Belum mengintegrasikan teknik otomatisasi pencarian parameter seperti *Optuna* untuk lebih menekan metrik error pada dataset properti pariwisata.

Hasil dari kedua penelitian terdahulu masih sebatas pembuktian performa komputasional di lingkungan evaluasi metrik analitik. Oleh karena itu, penelitian ini hadir untuk menjembatani celah tersebut dengan tidak hanya mengembangkan model *XGBoost* yang dioptimasi secara otomatis menggunakan *Optuna*, tetapi juga mendemonstrasikan kebaruannya melalui implementasi model tersebut ke dalam arsitektur sistem perangkat lunak berupa aplikasi *mobile* secara nyata dan *real-time*.

METODE PENELITIAN

Jenis Penelitian

Penelitian ini termasuk dalam jenis penelitian terapan dan penelitian pengembangan (*research and development/R&D*). Penelitian ini menerapkan metode *Machine Learning* dengan pendekatan regresi untuk menyelesaikan permasalahan ketidakpastian harga akomodasi yang bervariasi pada pasar properti di Bali. Hasil akhirnya berupa sebuah produk perangkat lunak yaitu aplikasi estimasi harga akomodasi berbasis *Android*.

Penelitian ini menerapkan pendekatan kuantitatif dengan dukungan teknik komputasional. Pendekatan ini dipilih karena seluruh tahapan analisis didasarkan pada perhitungan matematis yang terukur, mulai dari transformasi data spesifikasi akomodasi (seperti lokasi, fasilitas, dan tipe) menjadi numerik melalui proses *encoding* dan normalisasi. Selain itu, prediksi yang dihasilkan berupa nilai kontinu (*continuous value*) dalam satuan Rupiah, di mana tingkat akurasi model terhadap harga pasar aktual dievaluasi serta divalidasi secara objektif menggunakan parameter statistik berupa *Mean Absolute Error (MAE)* dan *Root Mean Square Error (RMSE)*.

Batasan Penelitian

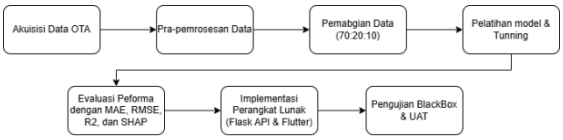
Agar penelitian ini lebih terarah dan spesifik, ditetapkan beberapa batasan metodologis, khususnya terkait akuisisi dan pemrosesan data

1. Temporal Data
   Penelitian ini belum mengintegrasikan variabel deret waktu yang dipengaruhi oleh seasonality, event pariwisata, atau hari libur nasional.
2. Proses web scraping difokuskan hanya pada satu platform *Online Travel Agent (OTA)*, yaitu tiket.com.
3. Cakupan Wilayah Geografis
   Pemodelan ini secara eksklusif dikembangkan menggunakan data spasial Provinsi Bali. Fitur-fitur lokasi yang diekstraksi (seperti koordinat GPS dan representasi nilai kecamatan) bersifat lokal.

Alur Penelitian

Untuk memberikan gambaran menyeluruh mengenai jalannya penelitian ini, disusun sebuah diagram alur metodologi penelitian yang sistematis. Tahapan penelitian ini dimulai dari tahap akuisisi data melalui *web scraping*, pra-pemrosesan data, pembagian dataset, pelatihan dan optimasi model *XGBoost* menggunakan *Optuna*, hingga implementasi ke dalam aplikasi mobile dan pengujian sistem kepada pengguna akhir. Alur pengerjaan lengkap dapat dilihat pada Gambar 2.

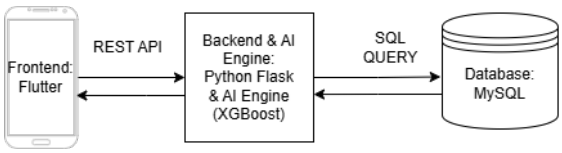
This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.



Gambar 1. Alur Penelitian

Spesifikasi dan Arsitektur Sistem

Untuk merealisasikan model prediktif ke dalam produk yang aplikatif, perangkat lunak dikembangkan menggunakan pendekatan arsitektur *client-server*. Secara fungsional, sistem bertugas untuk mengakuisisi parameter spesifikasi akomodasi dari pengguna (lokasi, tipe, dan fasilitas), mentransmisikan data tersebut melalui *REST API*, dan menjalankan inferensi machine learning menggunakan model *XGBoost* yang telah dilatih pada sisi backend. Hasil kalkulasi kemudian direpresentasikan kembali ke antarmuka pengguna dalam bentuk estimasi nominal harga sewa secara real-time.



Gambar 2. Arsitektur Sistem

Pelatihan Model

a. Pembagian Data
Peneliti melakukan pembagian dataset dengan skema 70:20:10, Pemilihan skema ini didasarkan pada pedoman standar machine learning untuk memastikan model memiliki data yang cukup untuk pembelajaran sekaligus mencegah kebocoran informasi pada saat evaluasi akhir (Goodfellow et al., 2016). Sebanyak 70% dialokasikan sebagai data latih, 20% sebagai data validasi yang berperan krusial bagi kerangka kerja Optuna untuk memonitor fungsi objektif dan mendeteksi overfitting selama proses pencarian parameter, serta 10% dikunci murni sebagai data uji.

b. Pemilihan XGBoost dan Optuna
Pelatihan utama menggunakan algoritma Extreme Gradient Boosting (XGBoost) Regressor. Pemilihan algoritma ini dijustifikasi oleh kemampuannya yang efisien dalam memproses dataset tabular berdimensi tinggi yang memiliki fitur kategorikal campuran, serta adanya mekanisme regularisasi bawaan yang secara teori lebih unggul dalam mengendalikan varians dibanding regresi klasik (Chen & Guestrin, 2016). Untuk mendapatkan performa optimal, dilakukan Hyperparameter Tuning. Kerangka kerja Optuna dipilih karena mengimplementasikan pendekatan Tree-structured Parzen Estimator (TPE) yang secara komputasi terbukti lebih efisien dan lebih cepat mencapai titik konvergensi hyperparameter global yang optimal dibandingkan metode konvensional seperti Grid Search atau Random Search (Akiba et al., 2019).

Pengujian dan Evaluasi Model

Tahapan pengujian dan evaluasi model dilakukan untuk mengukur seberapa presisi model dalam memetakan fungsi prediksi terhadap harga sewa aktual. Metrik evaluasi yang digunakan mencakup *Mean Absolute Error (MAE)*, *Root Mean Square Error (RMSE)*, dan *Coefficient of Determination (R^2)*. Pemilihan *MAE* dan *RMSE* didasarkan pada kemampuan kedua metrik tersebut dalam mengukur deviasi absolut menggunakan satuan data asli (Rupiah), di mana *RMSE* secara khusus memberikan penalti matematis yang lebih berat terhadap nilai prediksi yang meleset terlalu jauh (Chicco et al., 2021).

Batas toleransi error dan kriteria kelayakan model dalam penelitian ini ditentukan berdasarkan dua acuan utama, Model *XGBoost* yang diusulkan dinyatakan layak dan memiliki akurasi tinggi apabila mampu mencapai nilai R2 > 0.70.

Selain itu toleransi kesalahan dinilai berdasarkan performa komparatif. Model akhir diwajibkan menghasilkan nilai kesalahan MAE dan RMSE yang signifikan lebih rendah dibandingkan dengan model baseline (Naive Mean dan Linear Regression) maupun algoritma berbasis ensemble yaitu Radom Forest.

Validasi Sistem dan Pengujian Penerimaan Pengguna

Selain evaluasi performa algoritma secara matematis, validasi juga dilakukan terhadap kelayakan fungsional perangkat lunak yang dibangun. Pengujian kelayakan tahap pertama dilakukan menggunakan metode *Black Box Testing* untuk memastikan transmisi data antara aplikasi klien dan infrastruktur *backend* (*API*) tereksekusi tanpa kendala teknis. Tahap selanjutnya adalah *User Acceptance Test (UAT)* guna memvalidasi logika bisnis dan kebergunaan (*usability*) antarmuka aplikasi di lingkungan pengguna akhir.

Pengujian UAT melibatkan 30 responden yang ditarik menggunakan teknik purposive sampling, merepresentasikan dua target pengguna utama: pengelola properti dan pencari akomodasi. Evaluasi diukur menggunakan instrumen kuesioner berskala Likert 5 poin. Untuk menjamin keabsahan data empiris, instrumen kuesioner terlebih dahulu dievaluasi kelayakan statistiknya secara ketat; tingkat validitas

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

butir pernyataan diuji menggunakan Pearson Product Moment dengan syarat kelulusan r_hitung > 0,361 (pada N=30 dan taraf signifikansi 5%), sementara konsistensi internal instrumen divalidasi melalui perhitungan Cronbach's Alpha dengan ambang batas keandalan $\alpha > 0{,}60$.

HASIL DAN PEMBAHASAN

Pengumpulan data dan Pra-pemrosesan Data

Tahap pengumpulan data dilakukan secara otomatis menggunakan teknik *web scraping* berbasis *python* dengan pustaka *selenium*. Melalui proses ini, dataset mentah berhasil dikumpulkan 4.876 baris data mentah yang merepresentasikan tipe akomodasi hotel, villa, dan resort di wilayah Bali.

Sebelum digunakan dalam proses pelatihan model XGBoost, dataset melalui serangkaian tahapan pra-pemrosesan agar memiliki format yang terstandarisasi dan dapat diproses oleh algoritma secara optimal. Adapun tahapan pra-pemrosesan yang diimplementasikan meliputi:

a. Data Cleaning & Handling Outliner
   Pembersihan awal mencakup penghapusan karakter non-numerik seperti simbol "Rp" dan tanda baca pada variabel target, serta penanganan nilai kosong (missing values). Selanjutnya, untuk menjaga distribusi harga agar tidak bias, dilakukan penyaringan data pencilan (outliers) ekstrem menggunakan pendekatan statistic Interquartile Range (IQR). Proses ini menyisakan 4.618 data bersih.
b. Parsing Text & Ekstraksi Token
   Transformasi data tidak terstruktur pada kolom fasilitas dilakukan menggunakan teknik string parsing. String fasilitas majemuk dipecah menjadi token-token individu untuk memisahkan fitur ketersediaan fasilitas publik dan fasilitas privat.
c. Transformasi Data
   Data kategorikal dikonversi menjadi format numerik (float/integer) agar kompatibel dengan komputasi XGBoost. Fitur biner dan tipe pemandangan ditransformasi menggunakan One-Hot Encoding. Fitur bertingkat seperti tipe properti diproses dengan Label Encoding, sementara fitur lokasi (kecamatan) yang memiliki kardinalitas tinggi diproses menggunakan algoritma Target Encoding untuk menghindari peningkatan dimensi matriks yang eksponensial. Variabel target harga ditransformasi menggunakan operasi logaritmik natural (log1p) untuk menstabilkan varians dan mendekati distribusi normal.

Hasil akhir dari seluruh rangkaian pra-pemrosesan ini adalah sebuah matriks data numerik berdimensi tinggi berjumlah 4.618 baris dan 32 kolom. Dataset bersih ini kemudian dialokasikan untuk tahap pemodelan selanjutnya menggunakan skema partisi data 70:20:10, yakni untuk kebutuhan data latih, data validasi, dan data uji.

Tabel 1. Struktur data hasil scraping

Kelompok Fitur	Nama Atribut/fitur	Tipe Data	Deskripsi / Hasil *Encoding*
Target	price_numeric	Numeric (float)	Nilai transformasi log1p dari harga akomodasi asli
Informasi Dasar	star_rating, rating, log_review, room_size_m2	Numeric	Data kontinu hasil pembersihan dan normalisasi
Geografis	latitude, longitude, kecamatan	Float / Encoded	Koordinat GPS dan fitur lokasi hasil *Target Encoder*
Karakteristik	is_luxury_brand, property_type_enc, room type	Binary / Integer	Status *brand* mewah ,tipe properti (*Label Encoder*), tipe kamar (*Label Encoder*)
Fasilitas Umum	has_pool, has_wifi, has_ac, has_gym, has_spa, has_restaurant, has_parking,	Binary (0/1)	Hasil *One-Hot Encoding* untuk detail kamar
Fasilitas Kamar	has_breakfast, has_kitchen, has_bathtub, has_balcony, has_private_pool		Hasil *One-Hot Encoding* untuk detail kamar
Fitur Pemandangan	has_pool_view, has_ocean_view, has_garden_view,	Binary (0/1)	Hasil *One-Hot Encoding* untuk tipe pemandangan

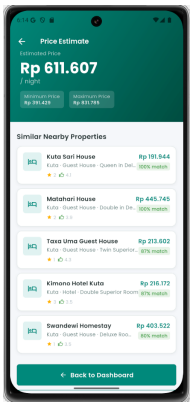
Antarmuka Sistem

Implementasi antarmuka pengguna untuk aplikasi estimasi harga akomodasi berbasis mobile dibangun mengggunakan *Flutter*. Aplikasi ini berfungsi untuk input parameter hunian mencakup detail akomodasi seperti nama akomodasi, tipe akomodasi, lokasi, fasilitas dari pengguna, lalu menampilkannya kembali hasil prediksi harga yang diterima dari server melalui *API*.

.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)



Gambar 3. Halaman Hasil Prediksi

Hasil Pelatihan Model dan *Hyperparameter Tuning*

Penelitian ini menerapkan metode *hyperparameter tuning* menggunakan framework *Optuna*. Proses pencarian kombinasi parameter terbaik dilakukan menggunakan *Tree-structured Parzen Estimator (TPE)* sampler sebanyak 100 percobaan.

Berdasarkan hasil eksperimen, algoritma berhasil menemukan titik konvergensi optimal yang menghasilkan nilai error validasi terendah. Kombinasi *hyperparameter* terbaik yang diterapkan pada model final dalah sebagai berikut:

Tabel 2. Tabel Hyperparamenter Terbaik dari Optuna

Parameter	*Best Value*
eta	0.06375
max_depth	10
subsample	0.5674
colsample_bytree	0.6193
min_child_weight	19
lambda	0.273200
alpha	0.08495
gamma	0.2630
n_estimators	1001

Evaluasi Kinerja Model XGBoost

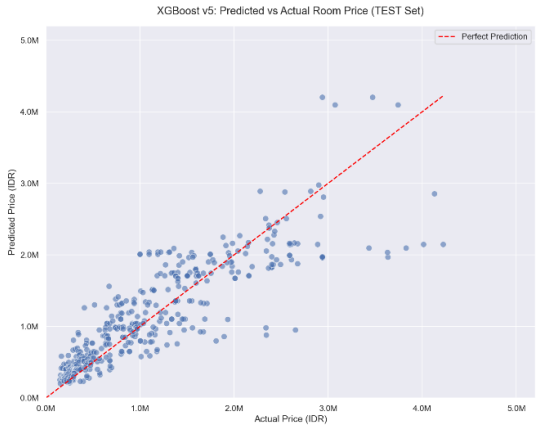
Pengukuran dilakukan dengan membandingkan model *XGBoost* yang telah di-*tuning* melawan empat model *baseline*. Hasil pengujian menunjukkan bahwa implementasi *hyperparameter tuning* pada *XGBoost* memberikan peningkatan akurasi yang sangat signifikan. Model *XGBoost* final berhasil mencatatkan nilai koefisien determinasi (R^2) sebesar 0,797. Hal ini menunjukkan bahwa 79,7% variabilitas harga sewa akomodasi di Bali dapat dijelaskan secara akurat oleh fitur-fitur yang diekstraksi.

Sebagai perbandingan, model Random Forest hanya mencapai R^2 sebesar 0,607 dan Linear Regression sebesar 0,56. Sehingga XGBoost terbukti lebih superior dalam hal akurasi prediksi pada dataset ini karena mekanisme regularisasi bawaannya yang sangat ketat dalam mencegah overfitting pada fitur-fitur kategorikal. Peningkatan metrik R^2 serta penurunan tingkat kesalahan (MAE, RMSE) ini secara empiris mengonfirmasi klaim bahwa XGBoost adalah algoritma yang andal dan presisi untuk memodelkan kompleksitas harga akomodasi harian di Bali dibandingkan algoritma lainnya.

Tabel 3. Hasil Pengujian Model

Model	RMSE	MAE	MAPE	R^2
Naive (train mean)	0.8294	0.7115	75.91%	-0.03
Linear Regression	0.5416	0.4467	49.72%	0.56
Random Forest	0.5123	0.4005	46.69%	0.607
XGBoost initial (val set)	0.5042	0.3921	41.99%	0.62
XGBoost final tuned (TEST set)	0.3752	0.3007	34.88%	0.797

Untuk memvisualisasikan tingkat akurasi prediksi, Gambar 8 menyajikan *Scatter Plot* perbandingan antara harga sewa aktual dan harga sewa hasil prediksi oleh model *XGBoost*.



Gambar 4. *Scatter Plot* harga predikisi vs harga aktual

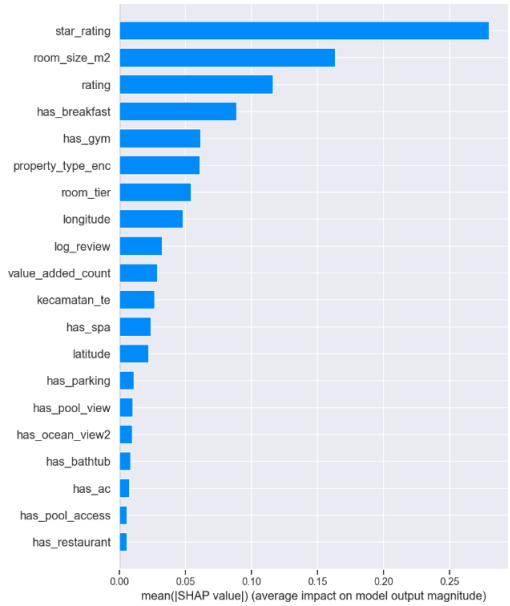
Berdasarkan plot tersebut, dapat diamati bahwa sebaran titik-titik observasi (*scatter points*) sebagian besar mengelompok dan mengikuti pola garis diagonal merah (Y = X), yang merepresentasikan garis prediksi sempurna. Konvergensi titik-titik data di sepanjang garis ini secara visual mengonfirmasi tingginya nilai *Coefficient of Determination* (R^2) sebesar 0,797, yang berarti model mampu memetakan harga aktual dengan tingkat deviasi (*error*) yang relatif rendah. Meskipun terdapat beberapa titik pencilan pada rentang harga atas, secara keseluruhan model *Extreme Gradient*

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

Boosting yang telah di-*tuning* berhasil menunjukkan performa generalisasi yang sangat baik dan stabil dalam memprediksi harga akomodasi harian di Bali.

Untuk memahami bagaimana model XGBoost menetapkan estimasi harga akomodasi, dilakukan analisis interpretasi model menggunakan metode *Shapley Additive Explanations* (*SHAP*). Analisis ini bertujuan untuk mengukur kontribusi rata-rata nilai absolut dari setiap fitur terhadap output prediksi harga.



Gambar 5. Grafik *Feature Importance*

Berdasarkan grafik *Feature Importance* (Gambar 10), model menunjukkan bahwa penentuan harga akomodasi harian di Bali sangat didorong oleh kombinasi antara dimensi fisik, lokasi strategis, dan kelas properti.

Komparasi dengan Studi Terdahulu dan Implikasi

Tingginya metrik akurasi yang dicapai oleh model ini sekaligus mengonfirmasi dan memperluas temuan dari penelitian-penelitian terdahulu. Pada studi pemodelan algoritma ensemble yang dilakukan oleh (Sari & Prabowo, 2025)untuk memprediksi harga rumah tapak, model XGBoost baseline mereka menghasilkan nilai koefisien determinasi (R^2) tertinggi sebesar 0,7746. Melalui integrasi hyperparameter tuning menggunakan Optuna, penelitian ini berhasil melampaui performa tersebut dengan mencatatkan R^2 sebesar 0,797 pada konteks dataset harga akomodasi yang memiliki tingkat volatilitas lebih tinggi.

Lebih lanjut, temuan ini memberikan sudut pandang baru terhadap kesimpulan riset (Fauzi et al., 2025) yang sebelumnya menyatakan bahwa Linear Regression lebih unggul daripada XGBoost dalam memprediksi harga aset properti statis. Berdasarkan hasil komparasi eksperimen pada penelitian ini Linear Regression justru mengalami penurunan performa yang drastis (R^2=0,560) ketika dihadapkan pada karakteristik dinamis dari harga sewa harian pariwisata. Hal ini membuktikan implikasi teoretis yang kuat bahwa interaksi multivariabel yang kompleks antara fasilitas privat, tipe pemandangan, dan lokasi geografis mikro pada akomodasi pariwisata bersifat non-linier, sehingga jauh lebih optimal diselesaikan menggunakan arsitektur Gradient Boosting yang diregularisasi ketimbang pendekatan regresi linear klasik.

Hasil Pengujian *Black Box*

Sebelum sistem diujikan kepada pengguna akhir, dilakukan evaluasi fungsionalitas perangkat lunak menggunakan metode *Black Box Testing*. Pengujian ini bertujuan untuk memvalidasi kesesuaian antara data masukan (*input*) yang dikirim melalui antarmuka aplikasi *mobile* dengan hasil keluaran (*output*) yang diproses oleh *backend API (Python Flask)* dan mesin model *XGBoost*, tanpa perlu meninjau struktur kode internal.

Fokus pengujian mencakup fitur-fitur krusial sistem, mulai dari validasi formulir parameter akomodasi, eksekusi proses *geocoding* di latar belakang, hingga kecepatan respon sistem dalam mengembalikan nilai prediksi harga. Ringkasan skenario dan hasil pengujian *Black Box* dapat dilihat pada Tabel 4.

Tabel 4. Hasil pengujian *Blax Box Testing*

No	Skenario Pengujian	Hasil Yang Diharapkan	Hasil Aktual	Status
1	Pengguna memasukkan lokasi, tipe akomodasi, dan mencentang fasilitas, lalu menekan tombol "Hitung".	Sistem menerima input dan memvalidasi kelengkapan data (mencegah *form* kosong).	Sistem berhasil memvalidasi input dan mengirim format JSON ke *backend*.	Valid
2	Sistem memproses data ke dalam model XGBoost yang telah dilatih.	Sistem mengembalikan nilai estimasi harga (Rp) dan merespon dalam waktu kurang dari 3 detik.	Sistem berhasil menampilkan prediksi harga di layar pengguna dalam waktu <1 detik.	Valid

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

3	Tombol hitung ditekan saat perangkat tidak memiliki koneksi internet.	Sistem tidak *crash* dan menampilkan notifikasi peringatan koneksi kepada pengguna.	Muncul *pop-up* "Gagal terhubung ke server, periksa koneksi internet Anda".	Valid

Hasil Uji Validitas dan Reliabilitas Instrumen

Sebelum melakukan perhitungan tingkat penerimaan pengguna akhir, dilakukan pengujian validitas dan reliabilitas terhadap data kuesioner yang dikumpulkan dari 30 responden. Pengujian ini menggunakan bantuan perangkat lunak statistik.

a. Hasil Uji Validitas

Berdasarkan uji Pearson Product Moment, tingkat validitas kedelapan butir pernyataan kuesioner dievaluasi dengan membandingkan r_{hitung} hadap nilai r_{tabel} sebesar 0,361 (pada N=30 dan taraf signifikansi 5%). Hasil uji validitas disajikan pada Tabel X berikut:

Tabel 5. Uji Validitas Kuisioner UAT

Kode UAT	Indikator Aspek	Nilai r_{hitung}	Nilai r_{tabel}	Status
UI-01	Antarmuka (UI)	0.618	0.361	Valid
UI-02	Antarmuka (UI)	0.577	0.361	Valid
US-01	Kemudahan (*Usability*)	0.448	0.361	Valid
US-02	Kemudahan (*Usability*)	0.453	0.361	Valid
US-03	Kemudahan (*Usability*)	0.607	0.361	Valid
FN-01	Fungsional (*Utility*)	0.454	0.361	Valid
FN-02	Fungsional (*Utility*)	0.523	0.361	Valid
FN-03	Fungsional (*Utility*)	0.431	0.361	Valid

b. Hasil Uji Reliabilitas

Setelah seluruh butir pernyataan dalam kuesioner dipastikan valid, langkah berikutnya adalah melakukan pengujian keandalan instrumen menggunakan metode *Cronbach's Alpha*. Pengujian ini bertujuan untuk mengukur tingkat konsistensi internal dan keandalan dari indikator-indikator yang menyusun instrumen kuesioner *User Acceptance Test (UAT)*. Secara statistik, sebuah instrumen dinyatakan reliabel dan memiliki konsistensi yang baik apabila nilai koefisien *Cronbach's Alpha* (α) yang dihasilkan lebih besar dari 0,60.

Hasil kalkulasi statistik terhadap jawaban dari 30 responden menunjukkan bahwa nilai koefisien *Cronbach's Alpha* (α) untuk kuesioner estimasi harga akomodasi ini adalah sebesar 0,609. Mengingat nilai koefisien tersebut telah melampaui ambang batas kriteria minimum yang ditetapkan (> 0,60), maka dapat ditarik simpulan bahwa instrumen kuesioner yang digunakan terbukti reliabel. Hal ini menunjukkan bahwa butir-butir pernyataan dalam kuesioner memiliki tingkat stabilitas dan konsistensi yang baik, sehingga aman dan layak digunakan secara ilmiah untuk mengukur tingkat penerimaan sistem perangkat lunak tanpa risiko bias komputasi.

Hasil *User Acceptance Test (UAT)*

Pengujian penerimaan pengguna (*User Acceptance Test*) dilakukan untuk mengevaluasi fungsionalitas dan tingkat kebergunaan (*usability*) antarmuka aplikasi dari sudut pandang pengguna akhir. Pengujian ini melibatkan 30 responden yang dipilih menggunakan teknik *purposive sampling*, yang terdiri dari kelompok sasaran utama, yakni perwakilan pengelola akomodasi dan masyarakat umum selaku pencari akomodasi.

Setiap responden diinstruksikan untuk menjalankan serangkaian skenario tugas (*task scenario*), mulai dari proses *input* parameter lokasi dan fasilitas akomodasi, mengeksekusi proses perhitungan model, hingga menginterpretasikan hasil prediksi harga yang ditampilkan oleh sistem. Setelah skenario diselesaikan, responden mengevaluasi sistem melalui kuesioner berbasis Skala *Likert* 5 poin yang terbagi ke dalam tiga aspek indikator utama: Antarmuka (*UI*), Kemudahan Penggunaan (*Usability*), dan Fungsionalitas (*Utility*).

Tabel 6. Hasil pengujian **User Acceptance Test**

Kode	Indikator Pernyataan	Total Skor	Persentase	Kategori
Antarmuka (*UI*)				
UI-01	Tampilan aplikasi menarik dan rapi	138	**92,0%**	Sangat Baik
UI-02	Informasi harga dan fasilitas mudah dibaca.	132	88,0%	Sangat Baik
Rata-rata Aspek Antarmuka (UI)		270	90,0%	Sangat Baik
Kemudahan (US)				
US-01	Aplikasi mudah dioperasikan tanpa panduan rumit	136	90,6%	Sangat Baik
US-02	Proses input data lokasi dan fasilitas berjalan cepat	133	88,6%	Sangat Baik

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

US-03	Navigasi antar halaman tidak membingungkan	131	87,3%	Sangat Baik
Rata-rata Aspek Kemudahan (Usability)		400	88,8%	Sangat Baik
Fungsional (*Utility*)				
FN-01	Fitur estimasi harga memberikan hasil masuk akal	132	88,0%	Sangat Baik
FN-02	Info rentang harga membantu penentuan sewa	135	90,0%	Sangat Baik
FN-03	Secara keseluruhan, aplikasi bermanfaat	134	89,3%	Sangat Baik
Rata-rata Aspek Fungsional (*Utility*)		401	89,1%	Sangat Baik
Rata-Rata Keseluruhan		1071	89,2%	Sangat Baik

Berdasarkan hasil rekapitulasi pengolahan data kuesioner, diperoleh temuan penerimaan sistem sebagai berikut

a. Aspek Antarmuka
   Tata letak visual dan kejelasan informasi aplikasi mendapatkan persentase kelayakan sebesar 90,0%. Hal ini menunjukkan bahwa informasi estimasi harga dan atribut kelengkapan fasilitas dapat dibaca serta dipahami dengan baik oleh pengguna.
b. Aspek Kemudahan Penggunaan (*Usability*)
   Kecepatan proses pengisian formulir data dan kejelasan navigasi menu memperoleh skor persentase sebesar 88,8%, yang mengindikasikan bahwa alur aplikasi dinilai sangat intuitif untuk dioperasikan tanpa memerlukan panduan teknis khusus.
c. Aspek Fungsionalitas (*Utility*)
   Manfaat hasil estimasi hargac berupa nilai tebakan dan rentang kewajaran harga terhadap penentuan keputusan penggguna mencatatkan tingkat persetujuan sebesar 89,1%.

Secara keseluruhan, rata-rata tingkat penerimaan pengguna terhadap aplikasi estimasi harga akomodasi ini mencapai 89,2%. Merujuk pada parameter evaluasi sistem, perolehan nilai rata-rata keseluruhan yang berada di atas ambang batas 75% mengategorikan sistem ini dalam predikat "Sangat Baik". Dengan demikian, dapat ditarik simpulan empiris bahwa sistem cerdas berbasis *XGBoost* yang dikembangkan telah memenuhi logika bisnis secara praktis dan sangat layak digunakan sebagai instrumen pendukung dalam penentuan maupun verifikasi harga sewa akomodasi di pasar pariwisata Provinsi Bali.

Implikasi Penelitian

Hasil dari pengujian model dan sistem dalam penelitian ini memberikan dua implikasi utama yang signifikan. Secara implikasi teoretis Penelitian ini berkontribusi dalam memperkaya literatur rekayasa data dan *machine learning* pada sektor pariwisata. Secara keilmuan, temuan ini membuktikan bahwa pemodelan harga sewa harian yang fluktuatif yang dipengaruhi oleh interaksi spasial dan kelengkapan fasilitas privat memiliki sifat non-linier yang kompleks. Oleh karena itu, pendekatan *ensemble learning* berbasis pohon keputusan (*tree-based*) yang dioptimasi seperti XGBoost dengan Optuna terbukti secara empiris lebih superior dan relevan untuk diterapkan pada dinamika industri *hospitality* dibandingkan pemodelan regresi klasik.

Secara implikasi praktis luaran penelitian berupa aplikasi *mobile* berbasis *XGBoost* ini dapat langsung dimanfaatkan sebagai instrumen acuan harga yang objektif di pasar pariwisata Bali. Bagi para pengelola atau pemilik akomodasi, sistem ini membantu mencegah kerugian akibat underpricing atau hilangnya daya saing akibat overpricing melalui rekomendasi tarif dasar yang berbasis data.

KESIMPULAN

Penelitian ini berhasil mengimplementasikan algoritma *Extreme Gradient Boosting (XGBoost)* untuk mengestimasi harga sewa akomodasi (hotel, villa, dan resort) di Bali. Melalui penerapan teknik *hyperparameter Tuning* menggunakan kerangka kerja *Optuna* pada 4.618 baris data bersih yang diperoleh dari proses web scraping, model *XGBoost* yang dioptimasi terbukti mampu menghasilkan tingkat akurasi yang tinggi. Kinerja model final mencatatkan nilai *Coefficient of Determination* (R^2) sebesar 0,797, *Root Mean Square Error (RMSE)* sebesar 0,3752, dan *Mean Absolute Error (MAE)* sebesar 0,3007. Hasil evaluasi ini membuktikan bahwa algoritma *XGBoost* yang telah di-*tuning* jauh lebih stabil dan presisi dalam menangani kompleksitas serta pola non-*linier* pada penentuan harga akomodasi harian dibandingkan metode regresi klasik.

Selain tingkat akurasi model, penelitian ini juga berhasil mengimplementasikan model tersebut ke dalam perangkat lunak berupa aplikasi estimasi harga berbasis mobile. Berdasarkan hasil pengujian *User Acceptance Test (UAT)* kepada pengguna akhir, sistem aplikasi ini mencatatkan tingkat penerimaan rata-rata keseluruhan sebesar 89,2%, yang mengkategorikannya ke dalam predikat "Sangat Baik". Secara teoretis, pencapaian ini membuktikan bahwa algoritma ensemble berbasis tree sangat relevan dan superior dalam menangani kompleksitas regresi pada data properti pariwisata yang dinamis. Dengan demikian, dapat ditarik simpulan empiris bahwa sistem cerdas

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

berbasis XGBoost yang dikembangkan sangat akurat, intuitif, dan layak diaplikasikan sebagai instrumen acuan harga yang objektif bagi pengelola properti maupun wisatawan di pasar pariwisata Bali.

Meskipun menunjukkan performa dan tingkat penerimaan yang tinggi, penelitian ini masih terbatas pada ruang lingkup data akomodasi di wilayah Bali serta belum mempertimbangkan faktor musiman (*seasonality*), *event* pariwisata, dan dinamika fluktuasi permintaan wisatawan. Oleh karena itu, penelitian selanjutnya diharapkan dapat mengintegrasikan variasi data temporal tersebut serta mengeksplorasi penggunaan pendekatan deep learning guna semakin meningkatkan akurasi, stabilitas, dan ketahanan prediksi model dalam jangka panjang.

DAFTAR PUSTAKA

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631. https://doi.org/10.1145/3292500.3330701

Badan Pusat Statistik Provinsi Bali. (2025). *Provinsi Bali dalam angka 2025*. Badan Pusat Statistik Provinsi Bali. https://bali.bps.go.id/

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, *13*, 281–305.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-August-2016*, 785–794. https://doi.org/10.1145/2939672.2939785

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*. https://doi.org/10.7717/PEERJ-CS.623

Claesen, M., & De Moor, B. (2015). Hyperparameter Search in Machine Learning. *ArXiv Preprint ArXiv:1502.02127*. http://arxiv.org/abs/1502.02127

Fauzi, M. R., Handika, M., Awinanto, A., Wahidin, A. J., Rahmatullah, B., & Kurniawati, I. (2025). Analisis Perbandingan Kinerja Algoritma Linear Regression, Random Forest, dan XGBoost dalam Prediksi Harga Rumah. *RIGGS: Journal of Artificial Intelligence and Digital Business*, *4*(4), 1541–1548. https://doi.org/10.31004/riggs.v4i4.3620

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Han, J., Kamber, M., & Pei, J. (2011). *Data mining:Concepts and techniques* (3rd ed.). Morgan Kaufmann.

Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development*, *15*(14), 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022

Kamil, M., Lestari Lokapitasari B, P., Lutfi, D., & Ilmawan, B. (2023). Implementasi framework flutter pada rancang bangun aplikasi konsultasi dokter hewan. *Buletin Sistem Informasi Dan Teknologi Islam*, *4*(4), 296–305.

Megawati, N. K. R., Wisnawa, I. M. B., & Trianingrum, N. N. N. (2024). Pengaruh lokasi harga kamar dan kualitas pelayanan terhadap keputusan tamu menginap di Capella Ubud Bali. *Bali Journal of Hospitality, Tourism and Culture Research*, *1*(2).

Pressman, R. S. (2010). *Software Engineering: A Practitioner's Approach* (7th ed.). www.mhhe.com/pressman

Rizquina, A. Z., & Ratnasari, C. I. (2023). Implementasi web scraping untuk pengambilan data pada website e-commerce. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, *5*(4), 377–383. https://doi.org/10.47233/jteksis.v5i4.913

Sari, W. B., & Prabowo, D. (2025). Analisis Perbandingan Prediksi Harga Rumah Dengan Random Forest, Gradient Boosting, dan XGBoost. *Intellect: Indonesian Journal of Learning and Technological Innovation*, *4*(1), 42–51. https://doi.org/10.57255/intellect.v4i1.1385

Wicaksono, F., Prasety, R. R., & Fadilah, M. N. A. (2025). Analisis Perbandingan Framework Pengembangan Aplikasi Mobile: Studi Kasus Flutter dan React Native. *INDOTERA-Indonesian Dissemination of Engineering and Applied Technology Research Applications, 1*(1), 17-25.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.