JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1264-1271
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1338

Optimasi Hyperparameter Gradient Boosting Menggunakan RandomizedSearchCV untuk Prediksi Harga Rumah di Wilayah Jabodetabek

Baka Dayla Mahaga Br Tarigan[1*], Muhammad Iqbal[2], Mia Rosmiati[3]

[1,2,3]Fakultas Teknik dan Informatika, Program Studi Informatika Kampus Pontianak, Universitas Bina Sarana Informatika, Kota Pontianak, Indonesia
Email: [1]15220488@bsi.ac.id, [2]iqbal.mdq@bsi.ac.id, [3]mia.mrm@bsi.ac.id
(*Email Corresponding Author: 15220488@bsi.ac.id)
Received: 25 Juni 2026 | Revision: 30 Juni 2026 | Accepted: 1 Juli 2026

Abstrak

Backlog perumahan di Jabodetabek yang tembus 2,93 juta unit bikin kebutuhan sistem prediksi harga rumah yang akurat jadi makin penting, supaya masyarakat dan pengembang bisa ambil keputusan jual-beli properti dengan lebih terukur. Penelitian ini mencoba membandingkan performa enam algoritma *machine learning*, yaitu *Ridge Regression*, *Random Forest*, *Gradient Boosting*, *XGBoost*, *Artificial Neural Network* (ANN) *Backpropagation*, dan *Deep Neural Network* (DNN), untuk memprediksi harga rumah di Jabodetabek menggunakan *dataset open source* dari Kaggle berisi 3.553 data. Tahapan penelitian meliputi eksplorasi data, penanganan *missing value*, penghapusan *outlier* dengan metode Interquartile Range (IQR), rekayasa fitur, *encoding*, standardisasi, pelatihan model dengan validasi silang 10-fold, serta penyetelan *hyperparameter* menggunakan *Randomized Search*. Hasil pengujian pada 571 data uji (20%) menunjukkan model *Gradient Boosting* yang sudah disetel (*tuned*) memberikan performa paling bagus dengan R^2 93,06%, MAE Rp265.951.001, RMSE Rp480.524.642, dan MAPE 13,42%, mengungguli *XGBoost* (R^2 92,33%), *Random Forest* (R^2 91,23%), ANN *Backpropagation* (R^2 86,70%), DNN (R^2 86,37%), dan *Ridge Regression* (R^2 85,65%). Hasil ini juga lebih tinggi dibandingkan penelitian-penelitian acuan sebelumnya yang memakai algoritma serupa pada *dataset* yang sama. Penelitian ini memberi kontribusi berupa perbandingan yang lebih lengkap antara algoritma berbasis pohon keputusan, regresi linear teregularisasi, dan jaringan saraf tiruan untuk kasus prediksi harga properti di kawasan urban Indonesia

Kata Kunci: Prediksi Harga Rumah, Jabodetabek, *Gradient Boosting, Random Forest, Machine Learning*

Abstract

The housing backlog in Jabodetabek, which has reached 2.93 million units, makes the need for an accurate house price prediction system increasingly important so that the public and developers can make more measurable property transaction decisions. This study compares the performance of six machine learning algorithms, namely Ridge Regression, Random Forest, Gradient Boosting, XGBoost, Artificial Neural Network (ANN) Backpropagation, and Deep Neural Network (DNN), in predicting house prices in Jabodetabek using an open-source Kaggle dataset of 3,553 records. The research stages include data exploration, missing value handling, outlier removal using the Interquartile Range (IQR) method, feature engineering, encoding, standardization, model training with 10-fold cross validation, and hyperparameter tuning using Randomized Search. Testing on 571 test records (20%) shows that the tuned Gradient Boosting model achieved the best performance with an R^2 of 93.06%, MAE of Rp265,951,001, RMSE of Rp480,524,642, and MAPE of 13.42%, outperforming XGBoost (R^2 92.33%), Random Forest (R^2 91.23%), ANN Backpropagation (R^2 86.70%), DNN (R^2 86.37%), and Ridge Regression (R^2 85.65%). These results also exceed previous reference studies using similar algorithms on the same dataset. This research contributes a more complete comparison between tree-based algorithms, regularized linear regression, and artificial neural networks for the house price prediction case in Indonesia's urban area.

***Keywords**: House Price Prediction, Jabodetabek, Gradient Boosting, Random Forest, Machine Learning*

1. PENDAHULUAN

Urbanisasi yang terus naik di Indonesia, khususnya di kawasan Jabodetabek (Jakarta, Bogor, Depok, Tangerang, dan Bekasi), bikin permintaan tempat tinggal makin tinggi [1]. Kawasan ini dikenal sebagai wilayah megapolitan dengan kepadatan penduduk tinggi dan pertumbuhan ekonomi yang cepat, sehingga harga rumah di sini cenderung naik tiap tahun [2]. Data dari Kementerian Pekerjaan Umum dan Perumahan Rakyat menunjukkan *backlog* perumahan di Jabodetabek

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1264-1271
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1338

sudah mencapai sekitar 2,93 juta unit atau 30% dari total *backlog* perumahan nasional, yang berdampak pada makin ketatnya persaingan masyarakat dalam mendapatkan rumah yang layak dan terjangkau [3].

Harga jual rumah dipengaruhi banyak faktor, mulai dari lokasi, luas tanah dan bangunan, jumlah kamar, fasilitas penunjang, sampai kondisi kelistrikan [3], [4]. Banyaknya faktor ini bikin calon pembeli maupun investor cukup kesulitan menilai kewajaran harga sebuah properti, sehingga dibutuhkan sistem yang bisa membantu mengestimasi harga rumah berdasarkan spesifikasinya secara lebih objektif. Pendekatan *machine learning* banyak dipakai untuk menjawab kebutuhan ini, karena mampu mempelajari pola hubungan non-linear antara fitur-fitur properti dan harga jualnya tanpa harus pakai asumsi statistik yang kaku.

Beberapa penelitian terdahulu sudah membandingkan berbagai algoritma *machine learning* untuk memprediksi harga rumah di kawasan Jabodetabek. [3] mengimplementasikan algoritma *Random Forest Regression* dengan *Grid Search Cross Validation* dan memperoleh R^2 sebesar 0,8751 atau 87,51% pada data ternormalisasi. [4] membandingkan algoritma *Extreme Gradient Boosting* dan *Random Forest* pada *dataset* yang sama, dan hasilnya *Random Forest* (R^2 = 0,77) ternyata lebih unggul dibanding *Extreme Gradient Boosting* (R^2 = 0,52) tanpa proses *hyperparameter tuning*. Sementara itu, [5] menerapkan algoritma *Gradient Boosting* berbasis pohon keputusan dan memperoleh RMSE sebesar Rp277.369.397 dengan MAPE 17,37% atau akurasi 82,63%, dan menunjukkan kinerja *Gradient Boosting* cukup kompetitif dibandingkan *Multi-Layer Perceptron*, *Linear Regression*, *Decision Tree*, dan *K-Nearest Neighbor*, meskipun sedikit di bawah *XGBoost* (83,2%). Penelitian lain di luar kawasan Jabodetabek, seperti perbandingan *Random Forest Regression* dan *Linear Regression* di kawasan elit [6], serta perbandingan *Random Forest*, *Linear Regression*, dan *Gradient Boosted Trees* di Tebet dan Jakarta Selatan [7], juga konsisten menunjukkan algoritma berbasis *ensemble* pohon keputusan lebih unggul dibanding regresi linear konvensional.

Selain studi soal harga rumah, ada juga studi sejenis di domain pendidikan yang membandingkan dua algoritma *deep learning*, yaitu *Deep Neural Network* (DNN) dan *Multi-Layer Perceptron* (MLP), untuk memprediksi kelulusan tepat waktu mahasiswa. Hasilnya menunjukkan bahwa DNN lebih unggul pada metrik Recall (0,9766), sedangkan MLP lebih unggul pada Accuracy (0,8812), Precision (0,9037), MCC, dan Cohen Kappa, yang berarti kedua model *deep learning* punya *trade-off* masing-masing tergantung metrik yang diprioritaskan [8]. Temuan ini jadi salah satu pertimbangan kenapa penelitian ini juga menyertakan model jaringan saraf tiruan (ANN dan DNN) dalam perbandingan, bukan cuma algoritma berbasis pohon keputusan, supaya bisa dilihat juga bagaimana performanya kalau diterapkan pada kasus regresi harga rumah.

Dari tinjauan terhadap penelitian-penelitian terkait di atas, bisa diidentifikasi sebuah *gap analysis*: penelitian sebelumnya umumnya cuma membandingkan satu sampai tiga algoritma saja, belum menyertakan model regresi teregularisasi (regularized regression) seperti *Ridge Regression* maupun model jaringan saraf tiruan yang lebih dalam (*deep learning*) secara bersamaan dalam satu kerangka eksperimen, dan belum semuanya menerapkan kombinasi rekayasa fitur, validasi silang k-fold, dan penyetelan *hyperparameter* secara menyeluruh pada *dataset* Jabodetabek yang sama. Selain itu, hasil R^2 pada penelitian Aqsha [4] (*Random Forest* 77%, *XGBoost* 52%) relatif jauh lebih rendah dibanding penelitian Putraa dan Suhartanaa [3] (87,51%) maupun [5] (82,63%), yang mengindikasikan bahwa tahapan praproses data dan optimasi parameter cukup berpengaruh besar terhadap akurasi prediksi walaupun *dataset*nya sama.

Berdasarkan gap tersebut, penelitian ini bertujuan membandingkan secara lebih lengkap kinerja enam algoritma *machine learning*, yaitu *Ridge Regression*, *Random Forest*, *Gradient Boosting*, *XGBoost*, *ANN Backpropagation*, dan *Deep Neural Network* (DNN), dalam memprediksi harga rumah di Jabodetabek pada *dataset* yang sama dengan penelitian-penelitian acuan, dengan menerapkan praproses data yang lebih lengkap (penanganan *outlier* IQR, rekayasa fitur, dan standardisasi), validasi silang 10-fold, serta penyetelan *hyperparameter* menggunakan *Randomized Search Cross Validation*. Hasil penelitian ini diharapkan bisa jadi acuan bagi masyarakat dan pengembang properti dalam memilih algoritma prediksi harga rumah yang paling akurat dan efisien, sekaligus melengkapi kekosongan kajian yang membandingkan algoritma *ensemble* pohon keputusan dengan algoritma berbasis jaringan saraf tiruan secara bersamaan pada kasus yang serupa. Selain itu, penelitian ini juga diharapkan dapat menjadi referensi metodologis bagi penelitian selanjutnya yang ingin menerapkan pendekatan serupa pada *dataset* properti dari wilayah lain di Indonesia, mengingat karakteristik pasar properti di luar Jabodetabek yang bisa jadi berbeda secara signifikan dari sisi distribusi harga maupun pola fitur yang dominan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Secara garis besar, penelitian ini dilakukan lewat beberapa tahapan utama, yaitu: (1) pengumpulan data, (2) eksplorasi data (Exploratory Data Analysis), (3) praproses data yang meliputi penanganan *missing value*, penghapusan *outlier*, rekayasa fitur, *encoding*, dan standardisasi, (4) pembagian data latih dan data uji beserta validasi silang k-fold, (5) pelatihan enam model *machine learning*, (6) penyetelan *hyperparameter* pada model dengan performa terbaik, dan (7) evaluasi serta perbandingan kinerja seluruh model memakai metrik R^2, MAE, RMSE, dan MAPE. Alur keseluruhan tahapan penelitian ini disajikan secara visual pada

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1264-1271
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1338

Gambar 1. Tahapan Penelitian

2.2 Dataset

Dataset yang dipakai adalah data sekunder *open source* berjudul "Daftar Harga Rumah Jabodetabek" yang diambil dari platform Kaggle [9], hasil *web scraping* dari situs rumah123.com. Dataset terdiri dari 3.553 baris data dengan 27 kolom atribut, mencakup informasi lokasi (district, city, latitude, longitude), spesifikasi fisik rumah (jumlah kamar tidur, kamar mandi, luas tanah, luas bangunan, jumlah lantai, carport, garasi), fasilitas (kamar dan kamar mandi pembantu, perabotan), serta atribut administratif (jenis sertifikat, kapasitas listrik, kondisi properti). Variabel price_in_rp dipakai sebagai variabel target.

2.3 Praproses Data

Kolom yang dirasa kurang relevan secara prediktif (seperti url, title, address, ads_id, dan facilities) maupun kolom dengan proporsi *missing value* yang sangat tinggi seperti building_orientation dihapus dari *dataset*, sejalan dengan pendekatan pada penelitian acuan [3]. Variabel kategorikal seperti district, city, property_type, certificate, dan furnishing dikodekan jadi nilai numerik memakai Label Encoding. Outlier pada variabel price_in_rp ditangani memakai metode Interquartile Range (IQR) dengan batas bawah dan batas atas yang disesuaikan terhadap kondisi pasar properti riil, sehingga dari 3.553 data awal didapat 2.851 data bersih (terhapus 702 data atau 19,8%) yang dipakai untuk pemodelan.

Tahap rekayasa fitur menambahkan beberapa variabel baru, yaitu rasio luas bangunan terhadap luas tanah (building_to_land_ratio), total jumlah kamar (total_rooms), total kapasitas parkir (total_parking), indikator rumah berlahan luas (is_large), kepadatan kamar per lantai (rooms_per_floor), serta transformasi logaritmik pada luas tanah dan luas bangunan untuk mengurangi skewness data. Variabel target price_in_rp juga ditransformasi jadi log_price untuk menstabilkan varians. Setelah itu, seluruh fitur numerik (total 25 fitur) distandardisasi memakai StandardScaler sebelum data dibagi menjadi 80% data latih (2.280 data) dan 20% data uji (571 data), dilengkapi validasi silang 10-fold pada data latih untuk mengestimasi performa model secara lebih robust.

2.4 Algoritma yang Dibandingkan

Ridge Regression adalah metode regresi linear yang menambahkan regularisasi L2 pada fungsi kerugian untuk mengurangi varians model dan mencegah *overfitting* akibat multikolinearitas antar fitur [10].

a. *Random Forest Regression* adalah algoritma *ensemble learning* berbasis pohon keputusan yang membangun banyak pohon regresi secara acak lewat proses *bootstrap aggregating* (*bagging*) dan *random subspace*, lalu merata-ratakan prediksi seluruh pohon untuk menghasilkan estimasi akhir yang lebih stabil [3], [11].
b. *Gradient Boosting Regression* membangun model secara berurutan, di mana setiap pohon baru dilatih untuk memperbaiki kesalahan (residual) dari pohon sebelumnya berdasarkan arah gradien negatif dari fungsi kerugian [5], [12].
c. *XGBoost* (*Extreme Gradient Boosting*) merupakan pengembangan dari gradient boosting yang menerapkan ekspansi Taylor orde kedua pada fungsi kerugian, regularisasi, serta pemrosesan paralel sehingga lebih efisien dan bisa menangani *dataset* besar dengan baik [4], [13].
d. *ANN Backpropagation* adalah jaringan saraf tiruan *feedforward* yang dilatih memakai algoritma *backpropagation* untuk memperbarui bobot antar neuron berdasarkan gradien fungsi kerugian terhadap target[14] .
e. *Deep Neural Network* (DNN) merupakan perluasan dari ANN dengan jumlah *hidden layer* yang lebih banyak (lima *hidden layer* pada penelitian ini), dilengkapi *Batch Normalization* dan *Dropout* untuk menjaga stabilitas pelatihan pada arsitektur yang lebih dalam [15]. Pemilihan DNN sebagai pembanding ANN juga terinspirasi dari studi perbandingan DNN dan MLP pada kasus klasifikasi kelulusan mahasiswa, yang menunjukkan bahwa kedalaman arsitektur jaringan saraf bisa memengaruhi trade-off antar metrik evaluasi [8].

2.5 Evaluasi Model

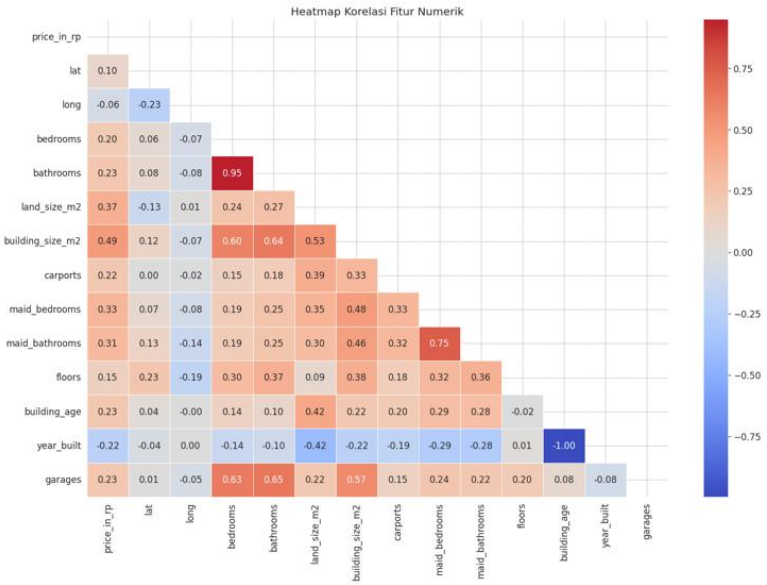
Evaluasi performa keenam model dilakukan memakai empat metrik, yaitu *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), *Mean Absolute Percentage Error* (MAPE), dan koefisien determinasi (R^2), sebagaimana dirumuskan pada penelitian acuan [3], [16]. Nilai R^2 yang lebih tinggi serta nilai MAE, RMSE, dan MAPE yang lebih rendah menunjukkan kinerja model yang lebih baik.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1264-1271
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1338

3. HASIL DAN PEMBAHASAN

3.1 Eksplorasi Data dan Korelasi Fitur

Analisis korelasi antar fitur numerik dilakukan untuk melihat variabel mana yang punya hubungan kuat terhadap harga rumah (price_in_rp). Hasilnya, variabel building_size_m2 punya korelasi positif paling tinggi terhadap harga (0,49), diikuti land_size_m2 (0,37), maid_bedrooms (0,33), dan maid_bathrooms (0,31). Selain itu, ditemukan juga korelasi yang cukup kuat antara bedrooms dan bathrooms, yang mengindikasikan rumah dengan jumlah kamar tidur banyak biasanya juga punya jumlah kamar mandi yang sebanding. Temuan ini konsisten dengan hasil eksplorasi data pada penelitian acuan [3], [4], yang juga menemukan bahwa luas bangunan dan luas tanah jadi faktor dominan dalam menentukan harga jual rumah. Variabel fitur rekayasa seperti building_to_land_ratio juga menunjukkan korelasi positif yang cukup signifikan terhadap harga, yang mengindikasikan bahwa efisiensi pemanfaatan lahan turut menjadi pertimbangan dalam penentuan nilai properti. Sebaliknya, fitur-fitur seperti floors dan carport memiliki korelasi yang relatif rendah, yang menunjukkan bahwa kedua atribut tersebut bukan faktor utama pembeda harga pada *dataset* ini. Secara keseluruhan, hasil eksplorasi data ini memperkuat asumsi bahwa model prediksi harga rumah di Jabodetabek perlu mampu menangkap hubungan non-linear antar fitur, sehingga pendekatan *machine learning* berbasis *ensemble* jauh lebih tepat dibandingkan regresi linear sederhana.

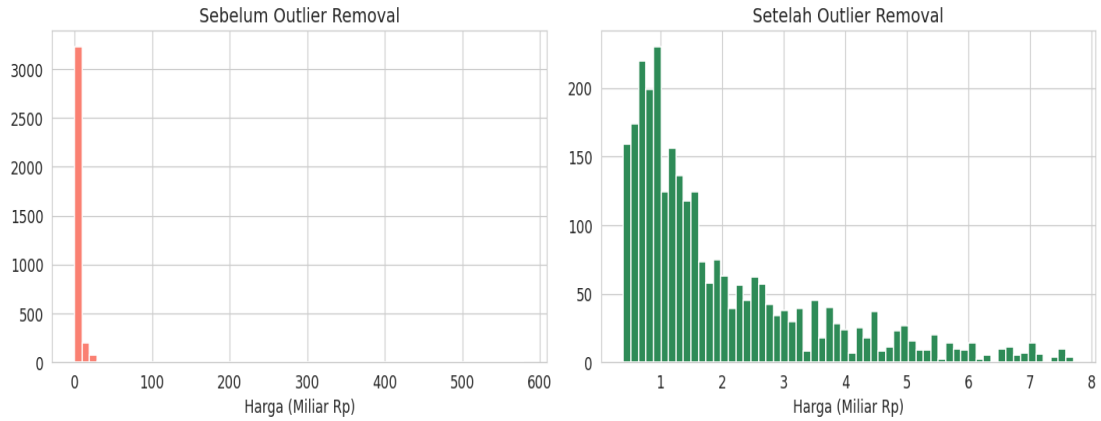


Gambar 2. Heatmap Korelasi Fitur Numerik

3.2 Penanganan Outlier

Distribusi awal variabel price_in_rp sangat condong ke kanan (*right-skewed*) karena adanya sejumlah kecil properti mewah dengan harga jauh di atas rata-rata. Penanganan *outlier* dilakukan memakai metode Interquartile Range (IQR) dengan batas bawah dan batas atas yang disesuaikan terhadap kondisi pasar, sehingga jumlah data berkurang dari 3.553 menjadi 2.851 data (terhapus 702 data atau 19,8%). Distribusi data setelah pembersihan jadi lebih representatif terhadap kondisi pasar rumah pada umumnya. Metode IQR dipilih karena bersifat robust terhadap distribusi yang tidak normal dan tidak mengasumsikan bentuk distribusi tertentu, berbeda dengan metode berbasis Z-*score* yang mengasumsikan data berdistribusi normal [16]. Selain itu, batas atas dan batas bawah IQR yang digunakan pada penelitian ini disesuaikan secara manual berdasarkan pertimbangan kontekstual pasar properti Jabodetabek, sehingga properti dengan harga yang sangat tinggi namun masih realistis dalam konteks pasar mewah tidak serta-merta dihapus hanya karena melampaui batas statistik baku. Pendekatan ini dinilai lebih tepat untuk data harga properti yang memiliki distribusi fat-tailed dibandingkan penghapusan outlier berbasis persentil tetap. Transformasi logaritmik pada variabel target price_in_rp menjadi log_price juga dilakukan setelah pembersihan outlier untuk lebih menstabilkan varians dan mendekati distribusi normal, yang pada akhirnya membantu model-model linear maupun boosting bekerja lebih optimal pada rentang nilai yang lebih seragam.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1264-1271
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1338



Gambar 3. Histogram Sebelum dan Sesudah Proses Outlier Remover

3.3 Hasil Pelatihan dan Perbadingan Model

Setelah melalui tahap praproses, sebanyak 25 fitur dipakai untuk melatih enam algoritma *machine learning*. Data dibagi menjadi 80% data latih (2.280 data) dan 20% data uji (571 data), dengan validasi silang 10-fold pada *Ridge Regression*, *Random Forest*, *Gradient Boosting*, dan *XGBoost*. Model *Gradient Boosting* dan *XGBoost* selanjutnya disetel ulang (*hyperparameter tuning*) memakai *Randomized Search Cross Validation* dengan 10-fold *cross validation*. Tabel 1 menyajikan hasil evaluasi lengkap dari delapan konfigurasi model (enam algoritma, dua di antaranya ditampilkan dalam versi sebelum dan setelah *tuning*) yang diurutkan berdasarkan nilai R^2 tertinggi pada data uji. Perlu dicatat bahwa *ANN Backpropagation* dan DNN tidak melalui proses *Randomized Search* seperti *Gradient Boosting* dan *XGBoost*, karena kompleksitas pencarian *hyperparameter* pada model neural network jauh lebih tinggi dari sisi komputasi. Arsitektur ANN yang digunakan terdiri dari dua *hidden layer* dengan masing-masing 128 dan 64 neuron, fungsi aktivasi ReLU, dan optimizer Adam. Sementara itu, arsitektur DNN menggunakan lima *hidden layer* dengan ukuran neuron yang semakin mengecil (256, 128, 64, 32, 16), dilengkapi *Batch Normalization* setelah setiap *hidden layer* serta *Dropout* dengan rate 0,3 untuk mencegah *overfitting*.

Tabel 1. Hasil Pelatihan dan Perbandingan Model

Model	R^2	MAE (Rp)	RMSE (Rp)	MAPE (%)
Gradient Boosting (Tuned)	93,06%	265.951.001	480.524.642	13,42
Gradient Boosting	92,53%	292.676.941	501.442.487	14,85
XGBoost (Tuned)	92,33%	284.723.448	509.673.717	14,17
XGBoost	91,85%	317.851.400	539.139.667	15,83
Random Forest	91,23%	324.271.397	578.984.284	15,68
ANN Backpropagation	86,70%	429.652.871	740.666.463	19,31
DNN (5 Hidden Layer)	86,37%	431.967.993	741.072.131	19,96
Ridge Regression	85,65%	454.386.504	714.851.571	23,03

Dari Tabel 1 bisa dilihat kalau *Gradient Boosting* yang sudah di-*tuning* jadi model paling unggul dengan R^2 93,06%, diikuti *XGBoost* (Tuned) di posisi kedua dengan R^2 92,33%. *Random Forest* masih kompetitif di angka 91,23%, sedangkan dua model *deep learning* (ANN dan DNN) justru kalah dari ketiga algoritma *ensemble* pohon keputusan, masing-masing cuma mencapai R^2 86,70% dan 86,37%. *Ridge Regression* jadi model dengan performa paling rendah (85,65%), yang masuk akal karena *Ridge Regression* cuma bisa menangkap hubungan linear, sementara harga rumah jelas dipengaruhi banyak interaksi fitur yang sifatnya non-linear. Dari sisi metrik MAE, *Gradient Boosting* (Tuned) menghasilkan rata-rata selisih prediksi sebesar Rp265.951.001 per unit properti, yang berarti prediksi model rata-rata meleset sekitar Rp265 juta dari harga aktual. Angka ini tergolong cukup baik mengingat rentang harga properti di Jabodetabek yang sangat luas, mulai dari kisaran Rp500 juta hingga lebih dari Rp5 miliar. Nilai MAPE sebesar 13,42% juga berada di bawah ambang 15% yang secara umum dianggap sebagai batas akurasi yang layak untuk model prediksi harga properti [16]. Perbandingan antara *Gradient Boosting* sebelum dan sesudah *tuning* menunjukkan peningkatan R^2 dari 92,53% menjadi 93,06% dan penurunan MAE dari Rp292.676.941 menjadi Rp265.951.001, yang membuktikan

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1264-1271
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1338

bahwa proses *hyperparameter tuning* memberikan kontribusi yang nyata terhadap peningkatan akurasi. Pola serupa juga terlihat pada *XGBoost*, dengan peningkatan R^2 dari 91,85% menjadi 92,33% setelah *tuning*. Menariknya, meskipun DNN memiliki arsitektur yang lebih dalam dari ANN, performanya tidak secara signifikan berbeda (R^2 86,37% vs 86,70%), yang mengindikasikan bahwa penambahan kedalaman lapisan tidak selalu memberikan keuntungan pada *dataset* tabular berukuran menengah seperti yang digunakan dalam penelitian ini [17], [8].

3.4 Perbadingan dengan Penelitian Terdahulu

Untuk menilai kontribusi penelitian ini, hasil model terbaik dibandingkan dengan tiga penelitian acuan yang menggunakan *dataset* Jabodetabek House Price yang identik. Perbandingan tersebut disajikan pada Tabel 2.

Tabel 2. Perbandingan dengan Penelitian Terdahulu

Penelitian	Algoritma Terbaik	R^2	Catatan
Putraa & Suhartanaa	*Random Forest*	87,51%	Grid Search CV
Aqsha	*Random Forest*	77%	Tanpa *tuning*
Lisnawati & Nugroho	*Gradient Boosting*	82,63%	Dihitung dari MAPE
Penelitian ini	***Gradient Boosting* (Tuned)**	**93,06%**	10-fold CV + RandomizedSearch

Tabel 2 menunjukkan bahwa model *Gradient Boosting* (Tuned) pada penelitian ini memperoleh R^2 sebesar 93,06%, melampaui *Random Forest* pada penelitian [3] (87,51%), *Random Forest* maupun *Extreme Gradient Boosting* pada penelitian [4] (77% dan 52%), serta *Gradient Boosting* pada penelitian [5] (82,63%). Peningkatan akurasi ini kemungkinan besar dipengaruhi beberapa faktor metodologis, yaitu penerapan rekayasa fitur tambahan (rasio luas bangunan-tanah, transformasi logaritmik), penggunaan transformasi log pada variabel target untuk menstabilkan varians, penanganan *outlier* yang disesuaikan dengan kondisi pasar, validasi silang 10-fold yang lebih ketat dibandingkan validasi silang 5-fold pada penelitian acuan [3], serta penyetelan *hyperparameter* memakai *Randomized Search Cross Validation*. Catatan: nilai R^2 pada penelitian [5] dihitung berdasarkan rasio MAPE (82,63% = 100% − 17,37%) karena penelitian tersebut tidak melaporkan nilai R^2 secara langsung. Jika dibandingkan lebih detail, selisih R^2 antara penelitian ini (93,06%) dan penelitian [3] (87,51%) adalah sebesar 5,55 poin persentase. Selisih ini cukup signifikan secara praktis, mengingat pada domain prediksi harga properti, peningkatan R^2 sebesar 1 poin persentase saja sudah bisa berdampak signifikan pada akurasi estimasi nilai properti yang bernilai miliaran rupiah. Perbedaan yang paling mencolok terlihat pada hasil penelitian [4], di mana *XGBoost* yang tidak melalui proses *tuning* hanya menghasilkan R^2 sebesar 52%, jauh di bawah hasil penelitian ini. Ini secara kuat mengindikasikan bahwa kualitas praproses data dan *hyperparameter tuning* merupakan faktor penentu yang jauh lebih dominan dibandingkan sekadar pemilihan algoritma. Dengan kata lain, algoritma yang sama dapat menghasilkan performa yang sangat berbeda tergantung seberapa baik data disiapkan dan parameter model dioptimalkan [12], [13].

Secara umum, hasil ini memperkuat kesimpulan pada literatur sebelumnya bahwa algoritma *ensemble* berbasis pohon keputusan, khususnya gradient boosting, secara konsisten lebih unggul dalam memprediksi harga rumah dibandingkan algoritma regresi linear maupun jaringan saraf tiruan sederhana pada *dataset* tabular berukuran menengah seperti *dataset* Jabodetabek House Price ini [6], [7], [17]. Hal ini juga sejalan dengan temuan pada studi perbandingan DNN dan MLP di domain lain [8], yang sama-sama menunjukkan bahwa model *deep learning* tidak selalu otomatis lebih unggul dibanding algoritma lain kalau jumlah data dan kompleksitas fiturnya belum terlalu besar. Dari perspektif praktis, model *Gradient Boosting* (Tuned) yang dihasilkan pada penelitian ini berpotensi diimplementasikan dalam bentuk sistem pendukung keputusan berbasis web yang dapat digunakan oleh calon pembeli, agen properti, maupun pengembang untuk memperoleh estimasi harga rumah secara cepat berdasarkan spesifikasi yang dimasukkan. Namun demikian, perlu dicatat bahwa model ini dilatih menggunakan data yang dikumpulkan pada periode tertentu, sehingga akurasi prediksinya dapat menurun seiring perubahan kondisi pasar properti dari waktu ke waktu. Pembaruan model secara berkala dengan data terbaru menjadi kebutuhan mutlak jika model ini ingin digunakan dalam aplikasi nyata yang bersifat jangka panjang [18]. Selain itu, meskipun model ini menunjukkan akurasi yang tinggi pada data uji, penting untuk diingat bahwa prediksi berbasis *machine learning* tetap memiliki tingkat ketidakpastian yang inheren, sehingga sebaiknya digunakan sebagai acuan awal yang kemudian diverifikasi dengan penilaian ahli properti yang mempertimbangkan kondisi spesifik masing-masing unit properti.

4. KESIMPULAN

Penelitian ini berhasil membandingkan kinerja enam algoritma *machine learning*, yaitu *Ridge Regression*, *Random Forest*, *Gradient Boosting*, *XGBoost*, *ANN Backpropagation*, dan DNN, dalam memprediksi harga rumah di kawasan Jabodetabek menggunakan *dataset* sebanyak 2.851 data hasil praproses. Hasil evaluasi menunjukkan bahwa model

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1264-1271
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1338

Gradient Boosting yang sudah disetel *hyperparameter*-nya lewat *Randomized Search Cross Validation* memberikan kinerja terbaik dengan R^2 sebesar 93,06%, MAE Rp265.951.001, RMSE Rp480.524.642, dan MAPE 13,42%, diikuti oleh *XGBoost* (Tuned) dengan R^2 92,33%, dan *Random Forest* dengan R^2 91,23%. Algoritma berbasis *ensemble* pohon keputusan secara konsisten mengungguli *ANN Backpropagation*, DNN, dan *Ridge Regression*, yang mengindikasikan bahwa hubungan antar variabel spesifikasi rumah dengan harganya bersifat non-linear dan lebih cocok dimodelkan dengan pendekatan boosting. Hasil penelitian ini juga melampaui akurasi pada tiga penelitian acuan yang menggunakan *dataset* identik, sehingga bisa disimpulkan bahwa kombinasi rekayasa fitur, penanganan *outlier* yang tepat, transformasi target, validasi silang, dan penyetelan *hyperparameter* memberikan kontribusi yang cukup signifikan terhadap peningkatan akurasi prediksi harga rumah. Temuan ini juga menegaskan bahwa kualitas tahapan praproses data memiliki pengaruh yang sama pentingnya—bahkan melebihi—pemilihan algoritma itu sendiri. Penelitian selanjutnya bisa mengeksplorasi algoritma *ensemble* lain seperti LightGBM atau CatBoost, menambahkan fitur eksternal seperti jarak ke fasilitas umum, aksesibilitas transportasi publik, dan tingkat kepadatan lingkungan, serta menerapkan model pada *dataset* dengan rentang waktu yang lebih baru untuk meningkatkan generalisasi dan relevansi prediksi terhadap kondisi pasar properti Jabodetabek yang terus berkembang.

REFERENCES

[1] Badan Pusat Statistik, Statistik Indonesia 2023. Jakarta: Badan Pusat Statistik, 2023.

[2] N. F. Arsaf, Bakhtiar, and Ahmadin, "Dampak Urbanisasi terhadap Ketersediaan dan Keterjangkauan Perumahan di Kota Besar," QISTINA: Jurnal Multidisiplin Indonesia, vol. 4, no. 1, pp. 190–197, 2025.

[3] I. M. G. A. B. Putraa and I. K. G. Suhartanaa, "Implementasi Algoritma Random Forest Regression dalam Sistem Prediksi Harga Rumah di Jabodetabek," Jurnal Nasional Teknologi Informasi dan Aplikasinya (JNATIA), vol. 4, no. 1, pp. 27–38, 2025, doi: 10.1234/jnatia.v4i1.27.

[4] D. Aqsha, "Perbandingan Kinerja Algoritma Extreme Gradient Boosting dan Random Forest untuk Prediksi Harga Rumah di Jabodetabek," Jurnal Ilmu Komputer dan Sistem Informasi, vol. 13, no. 1, pp. 1–7, 2025, doi: 10.24912/jiksi.v13i1.32863.

[5] I. Lisnawati and A. A. Nugroho, "Decision Tree-Based Gradient Boosting: Algorithm to Approach House Price Prediction in Jakarta, Bogor, Depok, Tangerang, and Bekasi (Jabodetabek)," Jurnal Statistika Universitas Muhammadiyah Semarang, vol. 12, no. 2, pp. 1–11, 2024, doi: 10.14710/JSUNIMUS.12.2.2024.1-11.

[6] N. A. C. Putri and D. B. Arianto, "Komparasi Penggunaan Information Gain Pada Machine Learning untuk Memprediksi Harga Rumah di Jabodetabek," Jurnal Sains dan Teknologi, vol. 5, no. 3, pp. 756–762, 2024, doi: 10.55338/saintek.v5i1.2052.

[7] E. Fitri, "Analisis Perbandingan Metode Regresi Linier, Random Forest Regression dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah," Journal of Applied Computer Science and Technology (JACOST), vol. 4, no. 1, pp. 58–64, 2023, doi: 10.52158/jacost.491.

[8] M. Iqbal, Y. N. Dewi, Lisnawanty, Maisyaroh, and Suhardjono, "Optimalisasi Prediksi Dalam Kelulusan Berbasis Deep Learning: Perbandingan Kinerja Multi-Layer Perceptron dan Deep Neural Network," Infotek: Jurnal Informatika dan Teknologi, 2025.

[9] N. Barizki, "Daftar Harga Rumah Jabodetabek," Kaggle Dataset, 2022.

[10] T. Kotsilieris, I. Anagnostopoulos, and I. E. Livieris, "Special Issue: Regularization Techniques for Machine Learning and Their Applications," Electronics, vol. 11, no. 4, p. 521, 2022, doi: 10.3390/electronics11040521.

[11] M. Schonlau and R. Y. Zou, "The Random Forest Algorithm for Statistical Learning," The Stata Journal, vol. 20, no. 1, pp. 3–29, 2020, doi: 10.1177/1536867X20909688.

[12] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A Comparative Analysis of Gradient Boosting Algorithms," Artificial Intelligence Review, vol. 54, no. 3, pp. 1937–1967, 2021, doi: 10.1007/s10462-020-09896-5.

[13] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, 2016, doi: 10.1145/2939672.2939785.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1264-1271
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1338

[14] M. Li, "Comprehensive Review of Backpropagation Neural Networks," Academic Journal of Science and Technology, vol. 9, no. 1, pp. 150–154, 2024, doi: 10.54097/51y16r47.

[15] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA: MIT Press, 2016.

[16] C. Schröer, F. Kruse, and J. M. Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," Procedia Computer Science, vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.

[17] G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning with Applications in R, 2nd ed. New York: Springer, 2021.

[18] A. A. G. S. Utama, "The Best Model and Variables Affecting Housing Values of Big Cities in Indonesia," Galaxy International Interdisciplinary Research Journal, vol. 10, no. 6, pp. 782–793, 2022.