Jurnal Pendidikan dan Teknologi Indonesia (JPTI)
Vol. 5, No. 6, Juni 2025, Hal. 1663-1671
DOI: https://doi.org/10.52436/1.jpti.846
p-ISSN: 2775-4227
e-ISSN: 2775-4219

Penerapan Kombinasi *Random Forest* dan *Optuna Hyperparameter Tuning* Untuk Meningkatkan Akurasi Prediksi Harga Rumah

Reza Muhammad Fadzryan[1], Eka Angga Laksana[2]

[1,2]Teknik Informatika, Universitas Widyatama, Bandung, Indonesia
Email: [1]reza.fadzryan@widyatama.ac.id, [2]eka.angga@widyatama.ac.id

Abstrak

Prediksi harga rumah merupakan aspek penting dalam pengambilan keputusan investasi di sektor properti. Penelitian ini bertujuan untuk membandingkan beberapa algoritma *Machine Learning* dan *Deep Learning* dalam memprediksi harga rumah, serta mengoptimalkan Random Forest Regressor menggunakan Optuna dengan implementasi *Tree-structured Parzen Estimators (TPE)*. Dataset yang digunakan adalah *House Price 2023 Dataset* dari Kaggle, yang mencakup 168.000 entri data properti di Pakistan. Metodologi penelitian ini meliputi tahap preprocessing data, rekayasa fitur, serta penerapan beberapa algoritma prediksi, yaitu *Artificial Neural Networks (ANN)* dengan model *Feedforward Neural Network*, *KNeighborsRegressor*, *Linear Regression,* dan *Random Forest Regressor*. Model-model ini dievaluasi menggunakan metrik MAE, MSE, RMSE, R-squared, dan Akurasi. *Random Forest Regressor* memberikan hasil terbaik dengan *R-squared* 0.91 dan Akurasi 91.33%. Untuk meningkatkan performa model, diterapkan optimasi *hyperparameter* menggunakan *Optuna* dengan pendekatan TPE yang berbasis *Bayesian Optimization*. Hasil model yang dioptimalkan mencapai peningkatan performa dengan R-squared 0.92 dan akurasi 91.75%. Hasil ini menunjukkan bahwa optimasi *hyperparameter* menggunakan *Optuna* berbasis *Bayesian Optimization* dapat meningkatkan akurasi prediksi harga rumah yang dapat diaplikasikan dalam analisis investasi properti.

Kata kunci: *House Price Prediction, Hyperparameter Tuning, Optuna, Random Forest.*

Implementation of Random Forest and Optuna Hyperparameter Tuning Combination to Improve House Price Prediction Accuracy

Abstract

House price prediction is a crucial aspect of decision-making in the real estate investment sector. This study aims to compare several Machine Learning and Deep Learning algorithms for predicting house prices and to optimize the Random Forest Regressor using Optuna with the implementation of Tree-structured Parzen Estimators (TPE). The dataset used is the House Price 2023 Dataset from Kaggle, which contains 168,000 property entries from Pakistan.The research methodology includes data preprocessing, feature engineering, and the implementation of several prediction algorithms, namely Artificial Neural Networks (ANN) with a Feedforward Neural Network model, KNeighborsRegressor, Linear Regression, and Random Forest Regressor. These models were evaluated using MAE, MSE, RMSE, R-squared, and Accuracy metrics. The Random Forest Regressor achieved the best performance, with an R-squared score of 0.91 and an accuracy of 91.33%. To enhance the model's performance, hyperparameter optimization was conducted using Optuna with a TPE approach based on Bayesian Optimization. The optimized model demonstrated improved performance, achieving an R-squared score of 0.92 and an accuracy of 91.75%. These results indicate that hyperparameter optimization using Optuna based on Bayesian Optimization can improve the accuracy of house price prediction, which can be applied in property investment analysis.

Keywords: *House Price Prediction, Hyperparameter Tuning, Optuna, Random Forest.*

1. PENDAHULUAN

Kenaikan harga perumahan yang signifikan di beberapa negara telah menyebabkan berkurangnya daya beli masyarakat. Hal ini secara langsung berdampak pada perekonomian negara dan kualitas hidup warganya [1]. Nilai rumah secara keseluruhan dipengaruhi oleh berbagai faktor, termasuk kondisi fisik, konsep, dan lokasi. Faktor fisik meliputi ukuran properti, jumlah dan luas ruangan, ketersediaan halaman, luas tanah dan bangunan, serta usia properti [2].

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Penelitian ini berfokus pada analisis perbandingan beberapa algoritma *Machine Learning* dan *Deep Learning* dalam memprediksi harga rumah. Algoritma yang dibandingkan meliputi *K-Nearest Neighbor Regression (KNN)*, *Linear Regression, Random Forest Regression, dan Artificial Neural Networks (ANN).* Keempat algoritma ini dipilih karena memiliki karakteristik yang berbeda dalam menangani data dengan pola hubungan *linear* maupun *non-linear*. Model terbaik akan dioptimalkan menggunakan *Hyperparameter Tuning* untuk mendapatkan hasil prediksi yang lebih akurat.

Berbagai penelitian sebelumnya telah membahas prediksi harga rumah dengan algoritma yang beragam. Penelitian yang dilakukan [3] menggunakan regresi linear berganda dalam prediksi harga rumah di Bandung dan menemukan bahwa metode ini mampu memberikan prediksi yang cukup akurat serta mengidentifikasi faktor-faktor utama yang memengaruhi harga rumah. Sementara itu, penelitian yang dilakukan [4] menggunakan *K-Neighbors Regressor* menunjukkan bahwa *K-Neighbors Regressor* lebih efektif dibandingkan regresi linear berganda dalam memprediksi harga rumah. KNN terbukti menjadi metode diskriminasi nonparametrik yang telah berhasil diterapkan dalam berbagai bidang, termasuk klasifikasi teks dan pemrosesan citra [5]. Penelitian yang dilakukan [6] dengan membandingkan beberapa algoritma *Machine Learning* dan menemukan bahwa *Random Forest Regression* memberikan akurasi tertinggi dibandingkan *Regresi Linear* dan *Gradient Boosted Trees*. Keunggulan utama *Random Forest* adalah kemampuannya dalam mengurangi overfitting dan meningkatkan ketangguhan model [1]. Selain itu, penelitian [7] menunjukkan bahwa *Artificial Neural Networks (ANN)* sangat efektif dalam memprediksi harga rumah di Jabodetabek, dengan hasil yang mendekati nilai aktual [8]. Dalam penelitian yang dilakukan [1], model *XGBoost* yang dioptimalkan dengan *Bayesian Optimization* menunjukkan performa yang lebih baik dibandingkan model tanpa optimasi. Namun, metode ini memiliki kelemahan dalam menangani ruang pencarian *hyperparameter* yang luas pada data pelatihan tertentu. Oleh karena itu, penelitian ini bertujuan untuk meningkatkan performa *Bayesian Optimization* dengan menggunakan *Optuna*. *Optuna* merupakan pustaka optimasi *Bayesian* berbasis *Python* yang dirancang untuk menyederhanakan pencarian *hyperparameter* model *Machine Learning*. *Optuna* menggunakan *Tree-structured Parzen Estimator (TPE)* [9], metode ini dapat dengan mudah digunakan dalam satu proses, atau dapat dipelajari secara paralel di banyak mesin [10]. Beberapa penelitian telah membuktikan efektivitas *Optuna* dalam meningkatkan performa model. Penelitian [11] yang dilakukan menunjukkan bahwa optimasi *hyperparameter* dengan *Optuna* mampu meningkatkan akurasi model *ADASYN* dan *Random Forest* secara signifikan. Penelitian [12] yang juga menggunakan *Optuna* untuk mengoptimalkan model *Machine Learning* seperti *XGBoost, Random Forest, Decision Trees, CatBoost*, dan *Extra Trees* dalam memprediksi risiko kematian akibat stroke, dan hasilnya menunjukkan bahwa optimasi dengan Optuna memberikan performa terbaik. Peneliatan [13] yang menyoroti keunggulan *Optuna* dalam hal fleksibilitas, efisiensi, serta kemampuannya dalam menghemat waktu dan sumber daya, sekaligus mencegah *overfitting*.

Dengan demikian, tujuan utama penelitian ini adalah untuk mengevaluasi performa berbagai algoritma dalam prediksi harga rumah serta meningkatkan akurasi model terbaik menggunakan *Hyperparameter* Tuning berbasis *Optuna*. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem prediksi harga rumah yang lebih akurat dan efisien, serta memiliki aplikasi potensial dalam analisis investasi properti.

2. METODE PENELITIAN

2.1. Desain Penelitian

Penelitian ini menggunakan pendekatan eksperimen untuk mengevaluasi efektivitas model *Random Forest Regressor* yang dioptimalkan menggunakan *Optuna Hyperparameter Tuning* dalam meningkatkan akurasi prediksi harga rumah dibandingkan dengan model *Random Forest Standar*. Model ini diterapkan pada dataset harga rumah dari "*House Prices 2023 Dataset"* [14] untuk memprediksi harga rumah berdasarkan fitur-fitur seperti lokasi, jumlah kamar tidur, jumlah kamar mandi, dan lain-lain.

Desain eksperimen yang digunakan adalah desain simulasi komparatif, yang di mana hasil dari model *Random Forest Standar* dibandingkan dengan hasil dari model *Random Forest Regressor* yang telah dioptimalkan menggunakan *Optuna Hyperparameter Tu*ning. Variabel dependen dalam eksperimen ini adalah akurat prediksi harga rumah yang diukur menggunakan metrik evaluasi seperti *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error (RMSE)* dan R^2 *Score.* Analisis lebih lanjut dilakukan berdasarkan distribusi harga rumah untuk mengeksplorasi variasi akurasi model pada kelompok harga yang berbeda.

2.1.1. Model Random Forest

Random Forest merupakan metode *ensemble* yang membangun beberapa pohon keputusan dan menggabungkan prediksinya untuk menghasilkan hasil yang lebih akurat dan stabil [15]. Prediksi akhir dari model *Random Forest* merupakan rata-rata dari prediksi seluruh pohon keputusan yang ada dalam *ensemble*, yang dapat dirumuskan:

p-ISSN: 2775-4227
e-ISSN: 2775-4219

$$\hat{y} = \frac{1}{n} + \sum_{i=1}^{n}(\hat{y}_i) \quad (1)$$

Dimana pada persamanaan (1), $\hat{y}_i$ adalah prediksi dari pohon keputusan ke- i dan n adalah jumlah pohon dalam *Random Forest*.

2.1.2. Hyperparameter dalam Random Forest

Beberapa *hyperparameter* kunci yang dioptimalkan dalam penelitian ini meliputi $n_{estimators}$ yang merupakan jumlah pohon dalam hutan, max_{depth} merupakan kedalaman maximum pohon, $min_{samples\_split}$ merupakan jumlah sampel minimal untuk memisahkan node, $min_{samples\_leaf}$ merupakan jumlah sampel minimal pada daun, dan $max_{features}$ yang merupakan fitur maksimum yang digunakan dalam pemisah node.

$$max_{features}\begin{cases}\sqrt{d}, \\ log_2(d), \\ d, \end{cases} \quad (2)$$

Dimana pada persamaan (2), d adalah jumlah total fitur dalam dataset.

2.1.3. Evaluasi Model

Untuk mengukur performa model, penelitian ini menggunakan tiga metrik evaluasi utama:

1) *Mean Absolute Error (MAE)*

$$MAE = \sum_{i=1}^{n}|y_1 - \hat{y}_i| \quad (3)$$

2) *Mean Squared Error (MSE)*

$$MSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (4)$$

3) *Root Mean Squared Error (RMSE)*

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (5)$$

4) *R-Squared (R^2)*

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_1 - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (6)$$

Dimana y_i adalah nilai sebenarnya, $\hat{y}_i$ adalah nilai prediksi dan $\bar{y}$ adalah rata-rata nilai sebenernya.

2.1.4. Optimasi Hyperparameter dengan Optuna

Optuna menggunakan algoritma *Tree-structured Parzen Estimator (TPE)* untuk mencari kombinasi *hyperparameter* yang optimal dengan mengeksploitasi informasi dari percobaan sebelumnya yang memodelkan distribusi probabilitas [17]. Fungsi objektif yang dioptimalkan adalah Objective = -MSE.

TPE menggunakan dua distribusi probabilitas, yaitu $p(x\,|y \leq y^*)$ = Distribusi *hyperparameter* yang menghasilkan error kecildan $p(x\,|y > y^*)$ = Distrubusi *hyperparameter* yang menghasilkan error besar.
Rasio antara kedua distribusi ini dirumuskan sebagai:

$$a(x) = \frac{p(x\,|y \leq y^*)}{p(x\,|y > y^*)} \quad (7)$$

2.1.5. Cross-Validation

Cross-validation digunakan untuk mengukur performa model secara lebih akurat. Jika menggunakan *k-fold cross-validation*, maka rata-rata MSE dihitung sebagai:

$$Cross - Validation\ Score = \frac{1}{k}\sum_{i=1}^{k} MSE_i \quad (8)$$

p-ISSN: 2775-4227
e-ISSN: 2775-4219

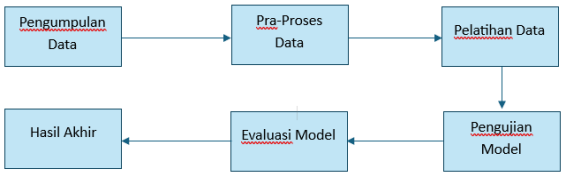
Dimana k adalah jumlah lipatan (*folds*) dan MSE_i adalah nilai MSE pada *fold* ke-i

Jumlah percobaan dalam Optuna ditentukan berdasarkan evaluasi performa model yang dihasilkan. Eksperimen ini bertujuan untuk menunjukkan bahwa penggunaan *Optuna Hyperparameter Tuning* dapat menghasilkan model prediksi harga rumah yang lebih akurat, stabil, dan dapat diandalkan dibandingkan dengan model *Random Forest Standar*.

2.2. Dataset Deskripsi dan Pra-Pemrosesan

Pra-pemrosesan data merupakan langkah yang sangat penting dalam setiap proyek pembelajaran mesin, karena memastikan data yang digunakan dalam model memiliki kualitas dan akurasi yang baik [16]. Pada tahap ini, kita mengidentifikasi dan memperbaiki masalah-masalah dalam dataset, seperti data yang hilang atau outlier, serta mengonversi variabel agar data lebih mudah dipahami dan bisa digunakan dengan tepat dalam model. Proses ini juga mencakup normalisasi data dan *encoding* variabel kategorikal untuk memastikan kompatibilitas dengan algoritma yang digunakan. Dengan melakukan ini, kita bisa memastikan bahwa model yang dibangun dapat bekerja dengan lebih efektif dan memberikan hasil yang lebih akurat [1]. Dalam penelitian ini, dataset yang digunakan adalah "*House Prices 2023 Dataset"* [14], yang diambil dari Kaggle dan berisi lebih dari 168.000 entri data properti di Pakistan. Dataset ini mencakup delapan atribut yang menggambarkan berbagai karakteristik properti, aspek geografis, serta proses terkait dengan transaksi properti tersebut.

2.3. Prosedur Metode Penelitian



Gambar 1. Metode Penelitian

Pada gambar 1 merupakan prosedur metode penelitian yang dirancang secara sistematis untuk mengevaluasi efektivitas algoritma dalam memprediksi harga rumah, baik sebelum maupun setelah optimasi hyperparameter menggunakan *optuna hyperparameter tuning*. Setiap tahapan dalam prosedur ini bertujuan untuk memastikan keakuratan prediksi model dengan menerapkan teknik *machine learning* yang sesuai. Tahapan penelitian mencakup pengumpulan data, pra-pemrosesan, pelatihan model, pengujian model, evaluasi performa, serta analisis hasil. Dengan mengikuti prosedur ini, penelitian dapat memberikan wawasan mendalam mengenai dampak optimasi hyperparameter terhadap akurasi model dalam memprediksi harga rumah.

2.3.1. Pengumpulan Data

Penelitian ini mengambil dataset dari *Kaggle.com* dengan judul dataset "*House Price 2023 Dataset*" [14] yang berisikan fitur atau atribut yang diperlukan seperti lokasi, luas tanah, jumlah kamar, dan lainnya. Dataset ini berisi 168.000 entri yang merinci berbagai properti real estate di Pakistan Kemudian proses selanjutnya mempelajari dan mengumpulkan literatur yang berkaitan erat dengan *Random Forest Regressor,* yang sumbernya bisa didapat dari jurnal atau buku.

2.3.2. Pra-pemrosesan Data

Pra-pemrosesan data menjelaskan langkah-langkah penting yang diambil untuk mempersiapkan data mentah untuk pemodelan. Langkah pertama dalam pra-pemrosesan data adalah pembersihan data (*Data Clearing*). Pembersihan data ini meliputi penanganan nilai yang hilang (*missing value*), penghapusan data yang tidak relevan atau redundan data, serta deteksi dan perbaikan kesalahan atau inkonsistensi dalam dataset. Proses ini penting untuk memastikan bahwa data yang digunakan memiliki kualitas yang baik dan tidak mempengaruhi hasil model secara negatif.

Selanjutnya dilakukan transformasi data, seperti normalisasi atau standarisasi untuk memastikan bahwa fitur-fitur dalam dataset memiliki skala yang seragam. Hal ini sangat penting, terutama ketika menggunakan algoritma yang sensitif terhadap skala data, seperti *K-Nearest Neighbor Regression* dan *Artificial Neural Networks*. Transformasi data juga mencakup pengkodean variabel kategorikal menjadi variabel numerik agar dapat diproses oleh model *Machine Learning* dan *Deep Learning*.

Lalu dilakukan tahap rekayasa fitur (*Feature Engineering*) dilakukan untuk mencipatakan fitur baru yang lebih relevan atau mengurangi dimensi data yang tidak penting. Dengan teknik rekayasa fitur, penulis dapat

p-ISSN: 2775-4227
e-ISSN: 2775-4219

meningkatkan data prediksi dari dataset, memungkinkan model untuk lebih baik dalam menangkap pola yang ada dalam data. Misalnya, penciptaan fitur interaksi variabel atau ekstrasksi informasi dari data waktu lebih berguna untuk prediksi harga rumah. Secara keseluruhan, tahap pra-pemrosesan data ini bertujuan untuk membentuk dataset yang siap digunakan untuk pelatihan model yang akan menghasilkan estimasi harga *real estate* yang lebih akurat

2.3.3. Pelatihan Data

Pelatihan data adalah langkah penting dalam pengembangan model *machine learning* yang dimana model dilatih untuk memahami pola dan hubungan antara fitur dan target menggunakan dataset pelatihan. Dalam penelitian ini, penulis menggunakan berbagai algoritma untuk membandingkan hasil dan memilih model terbaik untuk memprediksi harga rumah. Algoritma yang digunakan meliputi *Artificial Neural Networks (ANN)* menggunakan model *Feedforward Neural Network*, *Linear Regression (LR), K-Nearest Neighboar Regression (KNNR), dan Random Forest Regression (RFR).*

Setiap model dilatih menggunakan dataset pelatihan dengan parameter yang telah ditentukan. Proses pelatihan ini bertujuan untuk mempelajari pola yang ada dalam data, baik itu hubungan linear (Linear Regresi), non-linear (ANN & KNN), atau yang berbasis *ensemble* (*Random Forest*). Setelah setiap model dilatih, hasil prediksi yang diperoleh dari masing-masing algoritma dibandingkan untuk menentukan model mana yang memberikan hasil terbaik. Pembandingan dilakukan berdasarkan merik evaluasi, seperti *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (MSE) dan R^2 *Score*. Metrik ini digunakan untuk menilai kemampuan model dalam memprediksi harga rumah dengan akurat.

3. HASIL DAN PEMBAHASAN

Pada tahap ini, untuk menilai kinerja model setelah dilatih, digunakan metrik seperti *Mean Absolute Error (MAE), Mean Square Error (MSE)*, *R-squared*, dan akurasi. Metrik-metrik ini membantu untuk menentukan kecocokan prediksi model terhadap kejadian-kejadian yang terjadi dalam periode waktu tertentu.

Tabel 1 menunjukkan rincian kinerja dari enam model yang diterapkan dalam penelitian ini. Metrik yang digunakan untuk mengevaluasi kinerja adalah *Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (R²)*, dan akurasi. Model dengan nilai MAE dan MSE yang lebih rendah, R² yang lebih tinggi, serta akurasi yang lebih tinggi, dianggap lebih baik dalam memprediksi harga properti. Nilai MSE dan RMSE yang lebih kecil menunjukkan perbedaan yang lebih kecil antara harga yang diprediksi dan harga aktual. R² yang lebih tinggi menunjukkan kemampuan model dalam menjelaskan lebih banyak varians data, sementara akurasi yang lebih tinggi menunjukkan tingkat presisi yang lebih baik dalam prediksi yang benar.

Tabel 1. Perbandingan Kinerja Model

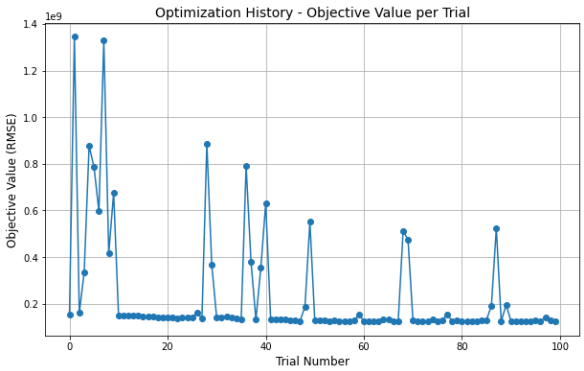
ALGORITMA	MAE	MSE	RMSE	R^2	ACC
ANN	8440.34	207863963.30	14417.49	0.85	85.18 %
KNNR	7243.62	195378566.38	13977.79	0.86	85.80%
LR	18572.52	608848896.85	4674.86	0.59	58.96%
RFR	6030.77	128655046.52	11342.62	0.91	91.33%
RFBO	6048.72	126757345.44	11258.66	0.91	91.46%
RFO	5889.00	122445317.96	11065.50	0.92	91.75%

Berdasarkan tabel di atas, ***Random Forest Regressor (RFR)*** menunjukkan hasil terbaik dengan **R-*Squared* (R²) sebesar 0.91**, yang menandakan bahwa model ini dapat menjelaskan 91% variansi harga rumah dalam dataset. Model ini juga memiliki a**kurasinya** yang sangat tinggi, yaitu **91.33%.** Model ini diikuti oleh ***KNeighbors Regressor (KNNR)***, yang memiliki ***R-Squared*** sebesar 85.80%. Sementara itu, ***Linear Regression*** *(LR)* memiliki kinerja yang lebih buruk dengan **R²** yang hanya mencapai 58.96%, menunjukkan bahwa model ini kurang mampu menjelaskan variansi harga rumah pada dataset. Setelah dilakukan proses optimasi, terlihat dari tabel 1. bahwa setelah dilakukan proses optimasi, metode *Optuna* memberikan hasil terbaik dengan MAE sebesar 5889.00, MSE sebesar 122445317.96, RMSE sebesar 11065.50, serta R² Score sebesar 0.92. Akurasi juga meningkat menjadi 91.75%, lebih tinggi dibandingkan metode lainnya. Hasil optimisasi menunjukkan bahwa metode *Optuna* lebih efektif dalam meningkatkan performa model dibandingkan *Bayesian Optimization* yang memiliki RMSE sebesar 11258.66. Hal ini disebabkan kemampuan *Optuna* untuk mengeksplorasi ruang parameter lebih fleksibel dan efisien melalui pendekatan trial berbasis iterasi. Pada implementasi *Random Forest Regressor*, penyesuaian parameter seperti peningkatan jumlah pohon (***n_estimators***) dan pengaturan kedalaman maksimum (***max_depth***)

p-ISSN: 2775-4227
e-ISSN: 2775-4219

berkontribusi signifikan dalam mengurangi error prediksi. Selain itu, penyesuaian jumlah sampel minimum (***min_samples_split)*** dan ***min_samples_leaf***) membantu model dalam menghindari *overfitting*.

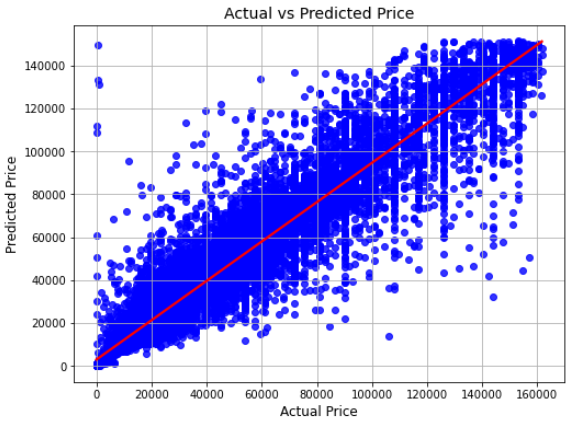
Pada gambar 2 menampilkan hasil optimisasi *hyperparameter* menggunakan Optuna dalam bentuk grafik *Optimization History - Objective Value per Trial*. Sumbu x merepresentasikan jumlah trial (0-100), sementara sumbu y menunjukkan nilai *objective function* dalam *bentuk Root Mean Square Error (RMSE)* dengan skala 1 x 10^9.



Gambar 2. *Optimization History - Objective Value per Trial*

Hasil visualisasi menunjukkan penurunan nilai RMSE yang signifikan. Pada awal optimisasi (trial 0-10), nilai RMSE menunjukkan fluktuasi tinggi mencapai 1.4 x 10^9, mengindikasikan fase eksplorasi parameter. Pada trial ke-20 hingga 100, grafik menunjukkan konvergensi yang stabil dengan nilai RMSE berkisar 0.2 x 10^9. hingga 0.4 x 10^9, meski terdapat beberapa lonjakan kecil pada trial ke-30, 40, 60, dan 80. Pada akhir optimisasi, nilai RMSE stabil di level 0.2 x 10^9, menunjukkan Optuna telah menemukan konfigurasi *hyperparameter* optimal untuk model *Random Forest*.

Pada gambar 3 ini digunakan untuk mengevaluasi sejauh mana model dapat memprediksi nilai aktual secara akurat berdasarkan dataset yang diberikan, scatter plot pada grafik ini menunjukkan perbandingan antara harga aktual dan harga prediksi. Garis diagonal merah merepresentasikan prediksi yang sempurna, dimana nilai prediksi sama dengan nilai aktual.



Gambar 3. *Predicted Prices vs Actual Prices*

Persebaran titik-titik biru di sekitar garis diagonal mengindikasikan korelasi positif yang kuat antara kedua variabel. Namun, terlihat adanya deviasi terutama pada rentang harga yang lebih tinggi, dimana model cenderung melakukan *underprediction*. Hal ini menunjukkan bahwa model memiliki performa prediksi yang baik secara keseluruhan, meskipun masih terdapat ruang untuk peningkatan akurasi pada prediksi harga tinggi.

Gambar 4 menunjukkan grafik *Residuals vs Predicted Prices*, di mana sumbu x adalah nilai prediksi harga rumah (*Predicted Prices*) dan sumbu y adalah nilai residual (*Residuals*). Residual dihitung sebagai selisih antara nilai aktual dan nilai prediksi:

$$Residual = y_{actual} - y_{predicted} \quad (9)$$

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Grafik ini digunakan untuk mengevaluasi pola kesalahan prediksi model dan memastikan bahwa asumsi regresi terpenuhi, khususnya mengenai sebaran residual.

Gambar 4. *Residuals vs Predicted Prices*

Grafik ini menunjukkan bagaimana perbedaan antara harga rumah yang diprediksi oleh model dengan harga sebenarnya (*residual*) tersebar di seluruh rentang harga prediksi. Sebagian besar titik *residual* tersebar secara acak di sekitar garis horizontal nol (ditandai dengan garis merah putus-putus). Pola ini menunjukkan bahwa model tidak memiliki bias tertentu dalam melakukan prediksi, yang berarti model cukup stabil dan kesalahannya tersebar merata. Namun, ada pola menarik di bagian harga tinggi. *Residual* cenderung lebih besar saat model memprediksi harga rumah yang lebih mahal. Hal ini menunjukkan bahwa model mengalami kesulitan dalam memprediksi rumah dengan harga tinggi secara akurat. Kemungkinan karena rumah dengan harga tinggi biasanya memiliki karakteristik yang lebih kompleks atau bervariasi, sehingga model membutuhkan informasi tambahan untuk membuat prediksi yang lebih tepat. Di sisi lain, untuk rumah dengan harga rendah hingga menengah, model terlihat lebih andal. Hal ini ditunjukkan oleh penyebaran residual yang lebih sempit, yang berarti prediksi model pada segmen ini cukup akurat.

Secara keseluruhan, grafik ini memberikan gambaran bahwa model sudah bekerja dengan baik, terutama di segmen harga rendah hingga menengah. Setelah melakukan proses training dan tuning hyperparameter, hasil model terbaik yang diperoleh memiliki akurasi sebesar 91.75%, berdasarkan nilai R^2 Score pada data uji. Pada tabel 2, berikut adalah sepuluh sampel hasil prediksi harga rumah dibandingkan dengan harga aktualnya:

Tabel 2. Hasil Prediksi Harga Rumah

Harga Aktual	Harga Prediksi
11,500,000	11,268,357.51
6,500,000	5,380,422.51
25,000	23,165.26
12,000,000	11,921,913.45
22,000,000	22,006,604.16
8,700,000	3,686,417.61
11,500,000	8,400,045.54
14,000,000	13,736,794.52
25,000	22,873.76
80,000	78,327.77

Dari tabel di atas, dapat dilihat bahwa dalam sebagian besar kasus, prediksi harga rumah cukup mendekati harga aktual. Namun, terdapat beberapa kasus di mana model masih mengalami selisih yang cukup besar, seperti pada rumah dengan harga aktual **8,700,000 PKR**, di mana prediksi model hanya **3,686,417.61 PKR**. Hal ini menunjukkan bahwa masih ada peluang untuk meningkatkan akurasi model, misalnya dengan menambahkan lebih banyak fitur atau mencoba metode optimasi yang lebih lanjut.

3.1 Diskusi

Dari hasil analisa yang telah dilakukan dan data pada tabel perbandingan regresi, keunggulan *Random Forest* dibandingkan *Artificial Neural Networks, Linear Regression,* dan *KNeighbors Regressor* dalam prediksi harga

p-ISSN: 2775-4227
e-ISSN: 2775-4219

rumah dapat dilihat pada tabel 1 *Random Forest* lebih unggul dalam menangani *hubungan non-linear* antar fitur dalam dataset properti. Ini terlihat dari nilai R^2 Score pada tabel bahwa *Random Forest Regressor* menunjukan hasil terbaik dari algoritma lainnya, dan *Random Forest* juga memiliki MAE dan MSE yang lebih rendah. Meskipun ANN memiliki R^2 Score yang cukup tinggi (0.85), performanya masih di bawah *Random Forest* (0.91). Selain itu, ANN memerlukan tuning arsitektur jaringan yang kompleks, termasuk pemilihan jumlah layer, neuron, dan fungsi aktivasi yang tepat. Jika tidak dilakukan tuning dengan baik, ANN dapat mengalami overfitting atau underfitting. Sebaliknya, *Random Forest* lebih mudah dioptimalkan dan tidak memerlukan penyesuaian arsitektur jaringan seperti ANN, membuatnya lebih efisien dari segi waktu dan sumber daya komputasi. Hasil penelitian ini juga dibandingkan dengan penelitian sebelumnya yang menunjukkan bahwa *hyperparameter tuning* dapat meningkatkan performa model *machine learning*. Studi sebelumnya[1] menunjukkan bahwa penggunaan *Bayesian Optimization* dapat meningkatkan akurasi prediksi, namun pada penelitian ini menunjukkan bahwa *Optuna* lebih efisien dalam eksplorasi ruang parameter**.** *Optuna* memungkinkan pemilihan kombinasi *hyperparameter* terbaik melalui pendekatan adaptif, yang menghasilkan penurunan RMSE dibandingkan model tanpa tuning. Dengan demikian, *Optuna* terbukti efektif dalam mengoptimalkan algoritma untuk menghasilkan model yang lebih akurat dan efisien.

4. KESIMPULAN

Berdasarkan hasil yang diperoleh melalui optimisasi model *Random Forest Regressor* menggunakan metode *Bayesian Optimization* dan *Optuna*, dapat disimpulkan bahwa *Optuna* memberikan performa yang lebih baik dalam meningkatkan akurasi prediksi harga rumah. Dengan mengoptimalkan *hyperparameter* seperti *n_estimators*, *max_depth*, *min_samples_split, min_samples_leaf*, dan *max_features*, *Optuna* berhasil mengurangi error prediksi dan meningkatkan akurasi model. Hasil evaluasi model setelah optimisasi dengan *Optuna* menunjukkan MAE sebesar 5889.00, MSE sebesar 122445317.96, RMSE sebesar 11065.50, serta R^2 Score sebesar 0.92, dengan akurasi mencapai 91.75%. Akurasinya yang mencapai 91.75% menunjukkan bahwa model mampu memprediksi harga rumah dengan cukup baik.

Secara keseluruhan, penelitian ini membuktikan bahwa penggunaan *Optuna* dalam optimisasi *hyperparameter Random Forest Regressor* sangat efektif, dengan hasil yang lebih baik dibandingkan metode lainnya, sehingga menjadikannya pilihan yang optimal untuk prediksi harga rumah pada dataset yang digunakan. Meskipun demikian, adanya keterbatasan dalam penelitian ini seperti dataset yang digunakan hanya mencakup satu negara (Pakistan), sehingga generalisasi ke wilayah lain memerlukan penelitian lebih lanjut dan juga model belum mempertimbangkan faktor eksternal seperti tren ekonomi dan kondisi pasar properti. Pada penelitian ini juga terdapat beberapa prediksi yang masih memiliki selisih signifikan, seperti pada rumah dengan harga aktual 8,700,000 PKR yang diprediksi hanya sekitar 3,686,417.61 PKR. Hal ini mengindikasikan adanya ruang untuk perbaikan lebih lanjut, baik dengan menambah fitur atau mencoba metode optimasi yang lebih lanjut.

DAFTAR PUSTAKA

[1] H. Jiang, "House Price Prediction with Optimistic Machine Learning Methods Using Bayesian Optimization," *Proceedings of the 1st International Conference on Data Science and Engineering*, vol. 1, pp. 488-496, 2024, doi: 10.5220/0012825400004547.

[2] F. M. Basysyar and G. Dwilestari, "House price prediction using exploratory data analysis and machine learning with feature selection," *Acadlore Trans. Mach. Learn.*, vol. 1, no. 1, pp. 11-21, 2022, doi: 10.56578/ataiml010103.

[3] R. N. T. Siregar, V. Sitorus, and W. P. Ananta, "Analisis Prediksi Harga Rumah di Bandung Menggunakan Regresi Linear Berganda," *Journal of Creative Student Research (JCSR)*, vol. 1, no. 6, pp. 395-404, 2023, doi: 10.55606/jcsrpolitama.v1i6.3038.

[4] L. F. Ihzaniah, A. Setiawan, and R. W. N. Wijaya, "Perbandingan Kinerja Metode Regresi K-Nearest Neighbor dan Metode Regresi Linear Berganda Pada Data Boston Housing," *Jambura Journal Probability and Statics,* vol. 4, no. 1, pp. 17-29, 2023, doi: 10.34312/jjps.v4i1.18948.

[5] R. Naz, B. Jamil, and H. Ijaz, "Real Estate Price Prediction," *International Journal of Innovations in Science & Technology*, vol. 6, no. 2, pp. 1031-1044, 2024. [Online]. Available: https://journal.50sea.com/index.php/IJIST/article/view/951

[6] 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. [Online]. Available: http://journal.isas.or.id/index.php/JACOST

p-ISSN: 2775-4227
e-ISSN: 2775-4219

[7] M. A. Hafizh, Subairi, R. D. Libriawan, N. D. Maulana, and A. M. Rizki, "Prediksi Harga Rumah Di Jabodetabek Menggunakan Metode Artificial Neural Network," *Jurnal Riset Inovasi Bidang Informatika dan Pendidikan Informatika*, vol. 5, no. 2, pp. 48-55, 2024, doi: 10.31284/j.kernel.2024.v5i2.6806.

[8] Y. Wu and J. Feng, "Development and Application of Artificial Neural Network," *Springer Nature Link*, vol. 102, pp. 1645-1656, 2017, doi: 10.1007/s11277-017-5224-x.

[9] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework," *The 25th ACM SIGKDD International Conference*, 2019, doi: 10.1145/3292500.3330701.

[10] K. Arai, I. Fujikawa, Y. Nakagawa, T. Momozaki, and S. Ogawa, "Modified Prophet + Optuna Prediction Method for Sales Estimations," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 13, no. 8, pp. 58-63, 2022, doi: 10.14569/IJACSA.2022.0130809.

[11] J. A. Amien, Y. Rizki, and M. A. R. Nasution, "Implementasi ADASYN untuk Imbalance Data pada Dataset UNSW-NB15," *Jurnal Computer Science and Information Technology*, vol. 3, no. 3, pp. 242-248, 2022, doi: 10.37859/coscitech.v3i3.4339.

[12] A. Tikaningsih, P. Lestari, A. Nurhopipah, I. Tahyudin, E. Winarto, and N. Hassa, "Optuna Based Hyperparameter Tuning for Improving the Performance Prediction Mortality and Hospital Length of Stay for Stroke Patients," *Telematika*, vol. 17, no. 1, pp. 1-16, 2024, doi: 10.35671/telematika.v17i1.2816.

[13] R. Kausar, F. Iqbal, A. Raziq, N. Sheikh, and A. Rehman, "Enhanced Foreign Exchange Volatility Forecasting using CEEMDAN with Optuna-Optimized Ensemble Deep Learning Model," *Sains Malaysiana*, vol. 53, no. 9, pp. 3229-3239, 2024. [Online]. Available: https://journalarticle.ukm.my/24507/

[14] U. Ali, "House Prices 2023 Dataset," *Kaggle*, 2023. [Online]. Available: https://www.kaggle.com/datasets/howisusmanali/house-prices-2023-dataset

[15] I. Maulita and A. M. Wahid, "Prediksi Magnitudo Gempa Menggunakan Random Forest, Support Vector Regression, XGBoost, LightGBM, dan Multi-Layer Perceptron Berdasarkan Data Kedalaman dan Geolokasi," *Jurnal Pendidikan dan Teknologi Indonesia* (JPTI), vol. 4, no. 5, pp. 221-232, 2024, doi: 10.52436/1.jpti.470.

[16] B. Nugroho and A. Denih, "Perbandingan Kinerja Metode Pra-Pemrosesan Dalam Pengklasifikasian Otomatis Dokumen Paten," *Jurnal Ilmiah Ilmu Komputer dan Matematika*, vol. 17, no. 2, pp. 381-387, 2020, doi: 10.33751/komputasi.v17i2.2148.

[17] Y. Ozaki, Y. Tanigaki, S. Watanabe, M. Nomura, and M. Onishi, "Multiobjective Tree-Structured Parzen Estimator," *Journal of Artificial Intelligence Research*, vol. 73, pp. 1209-1250, 2022, doi: 10.1613/jair.1.13188.