

BAB V

SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan hasil evaluasi yang telah dilakukan, penelitian ini berhasil menunjukkan cara dalam memilih algoritma *machine learning* terbaik untuk prediksi harga rumah di Kabupaten Tangerang dengan membandingkan performa masing-masing model. Dari lima model yang diuji, model *Optuna XGBoost* terbukti menjadi model paling optimal dengan performa terbaik dan tingkat generalisasi yang tinggi. Model ini menunjukkan nilai koefisien determinasi (R^2) sebesar 0,9280 pada data *training*, 0,8937 pada data *validation*, dan 0,9094 pada data *testing*. Selain itu, nilai RMSE yang dihasilkan adalah 0,2141 (*training*), 0,2594 (*validation*), dan 0,2460 (*testing*), serta nilai MAPE sebesar 0,0075 (*training*), 0,0091 (*validation*), dan 0,0083 (*testing*). Nilai R^2 yang tinggi serta RMSE dan MAPE yang rendah dan konsisten di ketiga *dataset* ini mengindikasikan bahwa model *Optuna XGBoost* mampu memprediksi harga rumah dengan akurasi tinggi serta menunjukkan kemampuan generalisasi yang baik tanpa mengalami *overfitting*. Hal ini menunjukkan bahwa model tersebut termasuk ke dalam kategori *goodfit*.

Selain itu, penelitian ini juga menunjukkan bahwa *Base Model Multiple Linear Regression* memiliki performa yang stabil, meskipun tidak sebaik *Optuna XGBoost*. Model ini mencatat nilai R^2 antara 0,8690 hingga 0,8844, dengan RMSE berkisar antara 0,2778 hingga 0,2877, serta MAPE di rentang 0,0099 hingga 0,0102 pada seluruh data *training*, *validation*, dan *testing*. Meski demikian, performa ini belum cukup untuk mengungguli *Optuna XGBoost*, sehingga model tersebut tidak dijadikan sebagai pilihan utama. Di sisi lain, model *Base Random Forest*, *Base XGBoost*, dan *Optuna Random Forest* menunjukkan adanya gejala *overfitting*, yang terlihat dari perbedaan performa yang cukup mencolok antara data *training* dan data *testing*. Ketiga model ini mencatat performa sangat baik pada data *training*, namun mengalami penurunan signifikan pada data *validation* dan *testing*, menandakan bahwa model tidak mampu melakukan generalisasi secara optimal dan tidak layak untuk digunakan pada tahap implementasi.

Dalam implementasinya, model *Optuna XGBoost* telah berhasil diintegrasikan ke dalam sebuah *website* berbasis *Streamlit* sebagai alat bantu bagi pengguna dengan *budget* tertentu dalam mencari estimasi harga rumah di Kabupaten Tangerang. *Website* ini memungkinkan pengguna untuk memasukkan spesifikasi rumah yang diinginkan dan mendapatkan estimasi harga dengan tepat, sehingga membantu pengguna dalam proses perencanaan pembelian rumah. Keakuratan model ini yang ditunjukkan melalui nilai R^2 yang tinggi serta nilai RMSE dan MAPE yang rendah dan stabil pada ketiga data—*training*, *validation*, dan *testing*—membuktikan efektivitas pendekatan ini dalam pengembangan sistem prediksi harga rumah yang andal.

Jika dibandingkan dengan penelitian sebelumnya, hasil *tuning* dengan *Optuna* pada penelitian ini memberikan wawasan baru bahwa peningkatan performa model tidak selalu ditunjukkan oleh kenaikan nilai R^2 . Dalam salah satu penelitian yang menggunakan *XGBoost* dan *tuning* dengan *Optuna*, nilai R^2 meningkat dari 0,78 menjadi 0,84, menunjukkan dampak positif *tuning* terhadap model. Namun, pada penelitian ini, nilai R^2 justru sedikit menurun dari 0,92 (*base model XGBoost*) menjadi 0,90 setelah *tuning* dengan *Optuna*. Penurunan tersebut tidak serta-merta menunjukkan performa model yang lebih buruk. Justru, hasil evaluasi menyatakan bahwa *tuning* *Optuna* berhasil mengurangi *overfitting* dan menciptakan model yang lebih seimbang serta andal dalam menghadapi data baru. Hal ini sangat relevan mengingat data yang digunakan memiliki banyak fitur (34 kolom), sehingga *tuning* mampu membantu model menangkap pola yang lebih umum. Dengan demikian, meskipun nilai R^2 sedikit menurun, model hasil *tuning* tetap dikategorikan sebagai *goodfit* dan menjadi solusi prediksi harga rumah yang paling optimal untuk diimplementasikan ke dalam sistem *digital*.

5.2 Saran

Untuk meningkatkan akurasi model dalam prediksi harga rumah di Kabupaten Tangerang, disarankan pada penelitian selanjutnya untuk memperbanyak jumlah data agar model dapat mengenali lebih banyak pola dan variasi yang ada dalam *dataset*. Selanjutnya, penerapan *hyperparameter tuning* dengan ruang pencarian yang lebih luas serta jumlah iterasi yang lebih banyak dapat meningkatkan peluang

untuk menemukan kombinasi *hyperparameter* yang lebih baik lagi, sehingga model dapat menghasilkan prediksi yang lebih akurat. Terakhir, pendekatan *machine learning* berbasis *time series* juga dapat dipertimbangkan, khususnya apabila data yang digunakan memiliki komponen temporal yang relevan, sehingga model dapat menangkap tren dan pola perubahan harga dari waktu ke waktu secara lebih baik. Diharapkan, dengan dilakukannya ketiga langkah tadi dapat meningkatkan performa model sehingga hasil prediksi bisa menjadi lebih akurat dan kuat terhadap berbagai kondisi data.

