

BAB I PENDAHULUAN

1.1 Latar Belakang

Harga properti, khususnya rumah, memainkan peran yang krusial dalam sektor ekonomi dan investasi, karena dipengaruhi oleh berbagai faktor seperti lokasi, ukuran bangunan, luas tanah, dan kondisi pasar. Harga rumah sering kali menjadi indikator ekonomi lokal, dan di sisi lain juga menjadi aset yang dinilai sebagai investasi jangka panjang bagi Masyarakat [1]. Bagi calon pembeli, memiliki informasi harga yang akurat dan dapat diandalkan sangat penting dalam mengambil keputusan yang informatif dan strategis. Di sisi lain, bagi pelaku industri real estate, seperti pengembang dan agen properti, perkiraan harga yang akurat memungkinkan mereka untuk menyesuaikan strategi pemasaran dan penawaran sesuai dengan nilai pasar yang sebenarnya. Dalam konteks kota Yogyakarta, yang terus berkembang pesat dan mengalami peningkatan permintaan terhadap properti residensial, menentukan harga rumah dengan akurat menjadi tantangan tersendiri. Berdasarkan data dari berbagai laporan properti nasional, di kota-kota besar secara umum, harga properti meningkat sekitar 5 hingga 15 persen per tahun [2], dipengaruhi oleh pertumbuhan ekonomi, urbanisasi, serta kebutuhan akan hunian dan investasi[1]. Kondisi ini menjadikan pasar properti semakin dinamis dan kompleks, sehingga membutuhkan pendekatan yang lebih canggih dalam melakukan prediksi harga.

Dalam perkembangan teknologi terkini, machine learning menjadi salah satu solusi yang diandalkan untuk meningkatkan akurasi dalam prediksi harga properti. *Machine learning* memungkinkan analisis data yang lebih mendalam dan otomatis, dengan memanfaatkan kemampuan komputasi untuk mengolah data dalam jumlah besar secara cepat. Model-model *machine learning* mampu mendeteksi pola dan tren yang mungkin tidak terlihat melalui metode konvensional. Salah satu teknik dalam machine learning yang telah terbukti efektif adalah *Support Vector Regression* (SVR), sebuah algoritma yang didesain untuk bekerja dengan

data non-linear maupun linear. SVR menggunakan fungsi kernel, seperti Radial Basis Function (RBF), untuk memetakan data ke dalam dimensi yang lebih tinggi sehingga pola kompleks dalam data dapat diidentifikasi dengan lebih baik. Selain SVR, metode ensemble learning seperti Random Forest dan XGBoost juga telah terbukti efektif dalam meningkatkan akurasi prediksi dengan mengombinasikan hasil dari beberapa model. Penggunaan kombinasi metode ini diharapkan dapat menghadirkan prediksi harga properti yang lebih andal dan dapat mengakomodasi kompleksitas data yang bervariasi, terutama di wilayah seperti Yogyakarta yang memiliki perbedaan nilai properti yang signifikan antar wilayah.[3], [4], [5], [6], [7], [8]

Penelitian ini berfokus pada penerapan model *Support Vector Regression*, *Random Forest*, dan *XGBoost* dalam memprediksi harga rumah di Yogyakarta berdasarkan berbagai variabel penting, seperti luas bangunan, luas tanah, jumlah kamar tidur, jumlah kamar mandi, dan lokasi. Data yang digunakan diperoleh dari platform Kaggle, yang menyediakan dataset properti residensial di Yogyakarta dengan berbagai spesifikasi yang relevan untuk analisis prediksi harga. Tujuan dari penelitian ini adalah untuk mengevaluasi efektivitas masing-masing model dalam konteks prediksi harga properti, serta mengetahui tingkat akurasi yang dapat dicapai dengan pendekatan ini. Untuk mendapatkan gambaran yang komprehensif mengenai kinerja model, penelitian ini menggunakan tiga metrik evaluasi utama, yaitu *R-squared*, *Mean Squared Error (MSE)*, dan *Mean Absolute Error (MAE)*. *R-squared* akan membantu dalam menilai seberapa baik model dapat menjelaskan variabilitas harga rumah berdasarkan fitur-fitur yang ada, sementara MSE dan MAE akan memberikan gambaran mengenai rata-rata kesalahan prediksi, baik dalam satuan kuadrat maupun absolut, guna memberikan informasi yang lebih detail mengenai akurasi model[9], [10], [11].

Harapannya, penelitian ini dapat memberikan kontribusi bagi pengembangan aplikasi machine learning dalam prediksi harga properti, sekaligus menjadi acuan untuk kasus prediktif serupa yang membutuhkan analisis berbasis fitur atau spesifikasi tertentu. Jika metode ensemble learning maupun *Support*

Vector Regression ini terbukti efektif, pendekatan serupa dapat diterapkan untuk memprediksi harga pada berbagai produk atau jasa yang memiliki karakteristik variabel kompleks dan saling berkaitan, seperti barang elektronik, kendaraan, dan lain-lain. Dengan demikian, hasil penelitian ini diharapkan dapat mendukung pemangku kepentingan untuk mengambil keputusan berbasis data yang lebih akurat dan efisien, tidak hanya dalam konteks properti namun juga dalam konteks prediksi harga lainnya di masa mendatang.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang sudah dijelaskan di atas, maka rumusan masalah yang didapat pada penelitian ini sebagai berikut:

1. Bagaimana penerapan model *Support Vector Regression* (SVR), dan model ensemble seperti *random forest* dan *XGBoost*, dalam prediksi harga rumah di kota Yogyakarta?
2. Bagaimana hasil akurasi masing-masing model berdasarkan evaluasi *Mean Absolute Error*, *Mean Squared Error*, dan *R-Squared* terhadap pengaruh harga rumah

1.3 Batasan Masalah

Penelitian ini memiliki batasan masalah sebagai berikut:

1. Dataset yang digunakan dalam penelitian ini diperoleh dari *Kaggle* melalui tautan <https://www.kaggle.com/datasets/pramudyadika/yogyakarta-housing-price-ndonesia/data>. Fitur yang digunakan berupa *price*, *listing-location*, *bed*, *bath*, *carport*, *surface_area*, *building_area*.
2. Evaluasi performa model prediksi dibatasi pada tiga metrik utama, yaitu *R-squared*, *Mean Absolute Error* (MAE), dan *Mean Squared Error* (MSE). Hasil evaluasi akan digunakan untuk mengukur akurasi dan efektivitas model, sementara metrik evaluasi lain tidak digunakan dalam penelitian ini.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Tujuan penelitian ini adalah menilai efektivitas dan kinerja masing-masing model dalam menghasilkan prediksi harga rumah yang akurat dengan menggunakan metrik evaluasi seperti R-squared, Mean Absolute Error (MAE), dan Mean Squared Error (MSE), serta menemukan model dengan performa terbaik dalam konteks data yang tersedia.

1.5 Manfaat Penelitian

Berdasarkan tujuan penelitian diatas, adapun manfaat dari penelitian yaitu:

1. Penelitian ini akan memberikan wawasan mendalam tentang efektifitas masing-masing model dalam memprediksi harga rumah
2. Hasil penelitian ini dapat membantu pelaku industri real estate, agen properti, dan calon pembeli rumah dalam membuat keputusan. Dengan wawasan yang diperoleh dari penggunaan model ensemble, mereka dapat lebih memahami dinamika harga rumah dan faktor-faktor yang mempengaruhinya, sehingga dapat mengambil langkah strategis dalam investasi properti.

1.6 Sistematika Penulisan

Penelitian ini disusun secara sistematis agar memudahkan pembaca dalam memahami isi dan alur penelitian. Sistematika penulisan dalam skripsi ini dijelaskan sebagai berikut:

BAB I PENDAHULUAN, bab ini berisi latar belakang masalah yang mendasari penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan sebagai panduan keseluruhan isi skripsi.

BAB II TINJAUAN PUSTAKA, bab ini memuat berbagai landasan teori yang relevan dengan penelitian, seperti teori machine learning, algoritma *Support Vector Regression* (SVR), *Random Forest*, dan *XGBoost*. Sumber pustaka berasal dari jurnal ilmiah, buku, dan referensi lain yang terpercaya.

BAB III METODE PENELITIAN, bab ini menjelaskan metodologi yang digunakan dalam penelitian, mencakup deskripsi dataset, tahapan pengolahan data, eksplorasi data (Exploratory Data Analysis), implementasi model machine learning, evaluasi model, dan alur penelitian secara keseluruhan.

BAB IV HASIL DAN PEMBAHASAN, bab ini menyajikan hasil penelitian, termasuk performa model yang diukur menggunakan evaluasi seperti R-squared, Mean Squared Error (MSE), dan Mean Absolute Error (MAE). Bab ini juga membahas analisis perbandingan performa model serta pengaruh fitur dalam prediksi harga rumah.

BAB V PENUTUP, bab terakhir ini berisi kesimpulan dari hasil penelitian serta saran untuk pengembangan penelitian lebih lanjut, baik dari sisi teknis maupun implementasi pada kasus nyata.

