

BAB I

PENDAHULUAN

1.1 Latar Belakang

Permasalahan perumahan merupakan isu penting yang dihadapi masyarakat modern, terutama di daerah dengan pertumbuhan ekonomi yang tinggi seperti kota-kota besar di Indonesia. Salah satu aspek krusial dalam persoalan ini adalah harga properti yang sangat dinamis dan dipengaruhi oleh berbagai faktor, sehingga fluktuasi harga rumah menjadi fenomena yang tidak dapat dihindari (Fauzi et al., 2025). Kenaikan harga rumah dari tahun ke tahun juga dipicu oleh meningkatnya permintaan akan tempat tinggal, yang secara langsung memengaruhi daya beli masyarakat (Aqsha, 2025). Kondisi ini menjadi perhatian tidak hanya bagi masyarakat umum, tetapi juga bagi investor, pengembang properti, dan pemerintah daerah dalam merancang kebijakan hunian jangka panjang.

Jakarta dan wilayah penyangganya, termasuk Depok, serta kota metropolitan seperti Bandung, merupakan pusat dinamika pasar properti yang tinggi di Indonesia, didorong oleh urbanisasi yang pesat, pertumbuhan penduduk, dan tingginya permintaan hunian (Aswin et al., 2022; Putri et al., 2024). Pertumbuhan kawasan perkotaan yang terus berlangsung telah menyebabkan kelangkaan lahan dan mendorong peningkatan nilai tanah serta harga properti, terutama di kawasan yang berdekatan dengan pusat kegiatan ekonomi (Gamal, 2023). Variasi harga lahan juga dipengaruhi oleh distribusi spasial pendapatan dan preferensi lokasi, di mana kelompok berpenghasilan tinggi cenderung terkonsentrasi di pusat kota dengan aksesibilitas lebih baik (Janssen et al., 2022), sementara faktor-faktor seperti tingkat pendapatan, jumlah penduduk, dan kebijakan kredit perumahan turut memengaruhi harga properti secara signifikan di berbagai kota besar Indonesia (Igamo et al., 2023). Tingginya aktivitas pasar ini tidak selalu diiringi kestabilan harga, karena fluktuasi dipengaruhi oleh kondisi ekonomi makro, karakteristik lokasi, serta perilaku pembeli dan investor yang beragam (Aswin et al., 2022; Gamal, 2023).

Untuk memahami dan memprediksi fenomena tersebut, diperlukan sumber data pasar yang komprehensif dan terkini. Dalam penelitian ini, data yang digunakan bersumber dari tiga *dataset* publik yang tersedia di platform *Kaggle*, yang secara keseluruhan mencakup 32.412 data properti dari wilayah Jakarta Selatan, Depok, dan Bandung. Setiap *dataset* memuat atribut-atribut utama seperti harga properti, jumlah kamar tidur, kamar mandi, garasi, luas tanah, luas bangunan, dan lokasi. Ketiga *dataset* tersebut kemudian digabungkan menjadi satu *dataset* terpadu dengan penambahan fitur *sumber_kota* sebagai penanda asal kota dari setiap data, sehingga model dapat mempelajari pola harga lintas kota secara bersamaan (Hibatulloh et al., 2025; Rayadin et al., 2024).

Namun, karakteristik data harga properti perkotaan menghadirkan tantangan tersendiri dalam proses pemodelan, salah satunya adalah keberadaan *outlier* atau nilai ekstrem yang tidak dapat diabaikan begitu saja. Pada data harga properti perkotaan, nilai ekstrem lazim muncul akibat adanya properti-properti premium yang memiliki karakteristik berbeda secara fundamental dari properti residensial pada umumnya. Keberadaan *outlier* semacam ini berpotensi mendistorsi parameter model regresi sehingga menurunkan akurasi prediksi secara keseluruhan, terutama pada model yang sensitif terhadap nilai ekstrem. Oleh karena itu, penanganan *outlier* menjadi bagian penting dalam tahap pra-pemrosesan data sebelum proses pelatihan model dilakukan. Selain itu, fluktuasi harga rumah juga mencerminkan hubungan non-linear yang kompleks antara harga dan berbagai faktor penentu seperti lokasi, ukuran, dan fasilitas, sehingga model konvensional seperti regresi linier kerap kesulitan menangkap pola tersebut secara menyeluruh. *Ensemble learning* menjadi solusi efektif karena menggabungkan beberapa algoritma untuk meningkatkan akurasi dan stabilitas prediksi (Hibatulloh et al., 2025; Liang et al., 2021; Rayadin et al., 2024; Zhang et al., 2022).

Dalam kerangka *ensemble learning*, terdapat tiga pendekatan utama yang umum digunakan, yaitu *bagging*, *boosting*, dan *stacking*. *Bagging*, seperti yang diterapkan pada *Random Forest*, bekerja dengan melatih beberapa model secara paralel pada subset data yang berbeda untuk mengurangi variansi, namun hanya menggabungkan model sejenis secara sederhana. *Boosting*, seperti *XGBoost*,

melatih model secara bertahap untuk mengurangi bias, namun pendekatannya yang sekuensial membuat setiap iterasi berfokus pada kesalahan sebelumnya sehingga kurang fleksibel dalam memanfaatkan keunggulan model yang berbeda jenis. *Stacking* mengatasi keterbatasan kedua pendekatan tersebut dengan menggabungkan prediksi dari beberapa model yang heterogen melalui sebuah *meta learner* yang secara adaptif mempelajari kombinasi optimal dari setiap *base learner* (Zhang et al., 2022). Penelitian ini menerapkan *Stacking Regressor* yang memadukan *Linear Regression*, *Random Forest*, dan *XGBoost* sebagai *base learner*. *Linear Regression* dipilih karena interpretasinya yang sederhana dan efektif sebagai model baseline (Liang et al., 2021), *Random Forest* karena kemampuannya menangani data kompleks dan hubungan non-linear (Rayadin et al., 2024; Zhang et al., 2022), serta *XGBoost* karena efisiensi dan performanya yang tinggi pada data tabular (Rayadin et al., 2024). Ketiga model ini memiliki karakteristik *bias-variance trade-off* yang berbeda, di mana setiap model memiliki tingkat keseimbangan yang unik antara kemampuan generalisasi dan sensitivitas terhadap data latih, sehingga kombinasinya melalui *Stacking* diharapkan mampu saling melengkapi dan menghasilkan prediksi yang lebih akurat dibandingkan model tunggal maupun pendekatan bagging maupun boosting. Hal ini didukung oleh temuan (Hibatulloh et al., 2025) yang mengindikasikan bahwa *Stacking Ensemble* menunjukkan performa terbaik dengan nilai R^2 sebesar 0,9076 dalam konteks prediksi harga rumah di Indonesia.

Sejumlah penelitian telah menunjukkan efektivitas metode regresi dalam memprediksi harga rumah dengan berbagai pendekatan algoritma. Model *regresi linier* banyak dimanfaatkan karena kemudahan interpretasinya, meskipun kinerjanya sering kali menurun ketika berhadapan dengan hubungan antarvariabel yang kompleks. Model berbasis pohon keputusan seperti *Random Forest* dan *XGBoost* menunjukkan performa lebih baik dalam menangani data non-linear dan menghadapi keberadaan *outlier*, serta mampu menghasilkan prediksi yang lebih stabil. Dalam kerangka *Stacking*, keberadaan *meta learner* yang belajar dari prediksi gabungan *base learner* memungkinkan model untuk tidak terlalu bergantung pada satu algoritma yang mungkin lebih rentan terhadap nilai ekstrem tertentu, sehingga secara keseluruhan prediksi menjadi lebih stabil dan tidak

terdistorsi oleh *outlier* secara berlebihan. Lebih lanjut, pendekatan *Ensemble* seperti *Stacking Regressor*, yang mengombinasikan kekuatan beberapa algoritma, mengindikasikan kemampuannya dalam meningkatkan akurasi dan generalisasi model secara signifikan. Studi pada berbagai wilayah menunjukkan bahwa kombinasi model ini dapat mengompensasi kelemahan masing-masing algoritma dasar, sehingga hasil prediksi menjadi lebih presisi dan tahan terhadap variasi data (Erdebilli & Devrim-İçtenbaş, 2022; Hibatulloh et al., 2025; Liang et al., 2021; Rayadin et al., 2024; Zhang et al., 2022).

Berdasarkan uraian sebelumnya, permasalahan harga properti yang fluktuatif di kota-kota besar Indonesia seperti Jakarta Selatan, Depok, dan Bandung memerlukan pendekatan prediksi yang akurat dan adaptif. Kondisi pasar yang dipengaruhi oleh kepadatan penduduk, urbanisasi, serta tingginya permintaan hunian menjadikan estimasi harga rumah sebagai tantangan yang kompleks, terlebih dengan keberadaan *outlier* nilai ekstrem yang perlu ditangani secara sistematis sebelum pemodelan dilakukan. Model regresi, khususnya yang menggabungkan beberapa algoritma dalam kerangka *ensemble learning*, menawarkan solusi potensial untuk menangkap pola non-linear dan mengatasi variasi data yang tinggi. Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi kinerja metode *Ensemble Stacking Regressor* dalam memprediksi harga rumah tinggal di Jakarta Selatan, Depok, dan Bandung. Hasil yang diperoleh diharapkan dapat memberikan kontribusi bagi pengembangan model prediksi harga properti yang lebih akurat dan dapat menjadi referensi bagi pengambil keputusan, baik dari kalangan masyarakat, pengembang, maupun pemerintah.

1.2 Rumusan Masalah

Penelitian ini mengidentifikasi beberapa permasalahan dalam memprediksi harga rumah tinggal, khususnya di wilayah perkotaan Indonesia, yaitu Jakarta Selatan, Depok, dan Bandung:

1. Harga rumah di Jakarta Selatan, Depok, dan Bandung memiliki variasi yang tinggi akibat pengaruh faktor lokasi, luas bangunan, jumlah kamar, dan

fasilitas lainnya, sehingga dibutuhkan model prediksi yang mampu menangkap kompleksitas hubungan antar variabel tersebut secara akurat.

2. Model prediksi tunggal seperti *Regresi Linear* atau *Random Forest* memiliki keterbatasan dalam menangani variasi data yang kompleks, terutama pada pola non-linear atau ketika terdapat *outlier*.
3. Penelitian yang menerapkan *Stacking Regressor* untuk prediksi harga rumah di wilayah perkotaan Indonesia umumnya berfokus pada satu kota secara terpisah, sehingga belum banyak yang mengeksplorasi pendekatan model gabungan lintas kota yang memungkinkan pembelajaran pola harga secara bersamaan dari Jakarta Selatan, Depok, dan Bandung.
4. Belum diketahui apakah pendekatan *Stacking Regressor* secara konsisten mampu memberikan performa prediksi yang lebih baik dibandingkan model tunggal seperti *Linear Regression*, *Random Forest*, dan *XGBoost* pada data harga properti perkotaan di Indonesia.
5. Belum tersedia pendekatan yang dapat menginterpretasikan hasil prediksi model machine learning secara praktis sebagai rekomendasi bagi masyarakat, investor, atau pengembang dalam konteks pengambilan keputusan properti di Jakarta Selatan, Depok, dan Bandung.

Berdasarkan permasalahan tersebut, maka rumusan masalah yang diajukan dalam penelitian ini adalah:

1. Bagaimana perbandingan performa model *Stacking Regressor* terhadap model tunggal (*Linear Regression*, *Random Forest*, *XGBoost*) dalam memprediksi harga rumah tinggal di Jakarta Selatan, Depok, dan Bandung berdasarkan fitur-fitur seperti lokasi, luas bangunan, luas tanah, dan jumlah kamar tidur serta kamar mandi?
2. Bagaimana hasil prediksi harga rumah dapat dimanfaatkan untuk menghasilkan rekomendasi yang relevan dalam konteks penentuan harga jual atau pembelian properti di Jakarta Selatan, Depok, dan Bandung?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dikemukakan, maka tujuan dari penelitian ini adalah:

1. Mengukur perbandingan performa model *Stacking Regressor* terhadap model tunggal (*Linear Regression*, *Random Forest*, *XGBoost*) dalam memprediksi harga rumah tinggal di Jakarta Selatan, Depok, dan Bandung berdasarkan variabel-variabel seperti lokasi, luas bangunan, luas tanah, dan jumlah kamar tidur serta kamar mandi.
2. Menghasilkan rekomendasi yang relevan berdasarkan hasil prediksi harga rumah, yang dapat digunakan sebagai acuan dalam menentukan harga jual atau memilih properti di Jakarta Selatan, Depok, dan Bandung.

1.4 Ruang Lingkup Penelitian

Penelitian ini akan memiliki ruang lingkup dan batasan sebagai berikut agar fokus dan hasil yang diperoleh lebih terukur:

1. Penelitian berfokus pada prediksi harga rumah tinggal berdasarkan data properti di wilayah Jakarta Selatan, Depok, dan Bandung. *Dataset* yang digunakan bersumber dari tiga *dataset* publik yang tersedia di platform *Kaggle* dengan total 32.412 data, mencakup atribut seperti harga properti, lokasi, jumlah kamar tidur, kamar mandi, garasi, luas tanah, luas bangunan, dan `sumber_kota` sebagai penanda asal kota.
2. Penelitian tidak membahas faktor eksternal yang tidak tercantum dalam *dataset*, seperti pengaruh kebijakan pemerintah, peraturan zonasi, kondisi ekonomi makro, atau faktor lain di luar atribut yang tersedia dalam *dataset*.
3. Data yang digunakan hanya mencakup rumah tinggal (residential), tidak termasuk apartemen, vila komersial, atau properti komersial lainnya.
4. Pembuatan *interface* sebagai sarana uji coba *model* prediksi harga rumah, namun tidak dibahas dalam konteks pengembangan *UI/UX*.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. **Bagi Peneliti dan Akademisi**, penelitian ini dapat memberikan kontribusi ilmiah dalam pengembangan metode prediksi harga rumah berbasis *ensemble learning*, khususnya dengan pendekatan *Stacking Regressor* yang menggabungkan tiga algoritma, serta menjadi referensi untuk penelitian lanjutan di bidang prediksi harga properti berbasis *machine learning*.
2. **Bagi Masyarakat Umum**, model prediksi ini dapat membantu memperkirakan harga rumah tinggal berdasarkan spesifikasi dan lokasi, sehingga mendukung pengambilan keputusan yang lebih tepat dan informatif dalam proses membeli atau menjual rumah.
3. **Bagi Investor dan Agen Properti**, hasil prediksi yang dihasilkan dari model ini dapat menjadi alat bantu untuk menganalisis tren pasar secara efisien, menilai potensi nilai properti secara objektif, serta mengidentifikasi peluang investasi yang menguntungkan.
4. **Bagi Pengembang Aplikasi Properti Digital**, penelitian ini dapat menjadi dasar untuk mengembangkan fitur estimasi harga otomatis (*price estimator*) yang bermanfaat bagi pengguna dalam memperoleh gambaran harga pasar secara lebih akurat dan real-time.