

BAB 1 PENDAHULUAN

1.1 Latar Belakang Masalah

Salah satu kebutuhan dasar manusia untuk bertahan hidup adalah memiliki akses terhadap tempat tinggal, khususnya rumah. Selain berfungsi sebagai tempat untuk tinggal dan beraktivitas, rumah juga dapat membantu meningkatkan kualitas kehidupan, mengurangi dampak perubahan iklim, mencegah penyakit, dan bahkan menyelamatkan nyawa [1]. Namun demikian, memperoleh rumah dengan harga yang terjangkau menjadi semakin sulit seiring berjalannya waktu [2]. Kesulitan ini dipengaruhi oleh berbagai faktor, seperti pesatnya urbanisasi yang mendorong peningkatan populasi, pendapatan yang tidak merata, dan fluktuasi harga tanah serta ketersediaan sumber daya, yang pada akhirnya menyebabkan kenaikan harga rumah [3]. Selain faktor-faktor eksternal yang telah disebutkan, unsur-unsur objektif yang berkaitan dengan rumah itu sendiri—seperti karakteristik bangunan, fasilitas, lokasi, dan lingkungan sekitarnya—juga memainkan peran penting dalam menentukan harga rumah [4]. Dengan banyaknya faktor internal dan eksternal yang memengaruhi harga rumah, penentuan harga yang wajar menjadi semakin kompleks. Oleh karena itu, diperlukan pendekatan prediksi yang mampu menangani kompleksitas hubungan antar faktor yang memengaruhi harga rumah secara lebih objektif berdasarkan karakteristik yang dimiliki.

Di era digital ini, ketersediaan data telah meningkat secara signifikan seiring dengan meluasnya penggunaan teknologi. Namun, peningkatan jumlah data tersebut juga menuntut adanya metode yang mampu mengekstraksi informasi secara efektif dan cerdas dari data yang tersedia [5]. Salah satu metode yang dapat digunakan adalah *Machine Learning*. Salah satu pendekatan yang banyak digunakan untuk mengolah data dalam jumlah besar adalah *Machine Learning*. Perkembangan *Machine Learning* dalam beberapa tahun terakhir memungkinkan komputer untuk mempelajari pola dari data, menghasilkan prediksi, serta membantu proses pengambilan keputusan berdasarkan informasi yang tersedia [6]. Dalam permasalahan regresi seperti prediksi harga rumah, kemampuan tersebut menjadi penting karena harga rumah dipengaruhi oleh banyak faktor yang saling berkaitan dan tidak selalu memiliki hubungan linear [7]. Selain itu, berbagai pendekatan *Machine Learning*, termasuk metode *ensemble*, memiliki karakteristik

dan performa yang berbeda dalam menangani pola data yang kompleks [8]. Oleh karena itu, *Machine Learning* menjadi pendekatan yang relevan untuk digunakan dalam prediksi harga rumah berdasarkan faktor-faktor internal maupun eksternal yang dimiliki suatu rumah.

Penelitian mengenai prediksi harga rumah atau properti telah banyak dilakukan dengan menggunakan berbagai pendekatan *Machine Learning*. Akan tetapi, banyak penelitian sebelumnya dilakukan di berbagai negara, dan karakteristik serta faktor-faktor internal maupun eksternal yang memengaruhi harga rumah cenderung berbeda pada setiap wilayah atau negara [9, 10]. Indonesia, sebuah negara yang saat ini merupakan ekonomi terbesar di Asia Tenggara dan terbesar kesepuluh dalam hal paritas daya beli [11], mengalami kesulitan dalam menentukan harga rumah baik sebagai pihak pembeli maupun pihak pengembang dan penjual karena banyaknya faktor yang memengaruhi harga rumah [12]. Tantangan ini menjadi semakin kuat untuk area dengan tingkat kepadatan penduduk atau populasi yang tinggi seperti Jakarta [13], karena tingkat kepadatan penduduk memengaruhi harga tanah [14]. Selain itu, sejumlah penelitian sebelumnya telah menggunakan berbagai pendekatan *Machine Learning*, termasuk metode *ensemble* seperti *Random Forest* dan *Extreme Gradient Boosting (XGBoost)*, untuk melakukan prediksi harga rumah maupun properti [15, 16, 17]. Hasil penelitian terdahulu menunjukkan bahwa metode-metode tersebut efektif dalam menangkap karakteristik data yang kompleks, sehingga dapat menghasilkan prediksi dengan tingkat akurasi yang relatif tinggi. Namun demikian, sebagian besar penelitian masih berfokus pada evaluasi model secara individual atau *standalone*, sementara penelitian yang mengevaluasi kombinasi atau pendekatan *hybrid* antar model *ensemble* masih relatif terbatas, khususnya pada konteks prediksi harga rumah di Jakarta. Padahal, kombinasi beberapa model memiliki potensi untuk menghasilkan karakteristik prediksi yang berbeda dibandingkan penggunaan model secara individual [18, 19]. Oleh karena itu, penelitian ini bertujuan untuk menganalisis dan membandingkan performa model *Random Forest*, *XGBoost*, serta pendekatan *hybrid ensemble* dalam melakukan prediksi harga rumah di Jakarta.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah sebagai berikut:

- Bagaimana implementasi model *Random Forest* dan *Extreme Gradient Boosting (XGBoost)* dalam melakukan prediksi harga rumah di Jakarta?

- Bagaimana implementasi pendekatan *hybrid ensemble* yang mengombinasikan model *Random Forest* dan *XGBoost* dalam melakukan prediksi harga rumah di Jakarta?
- Bagaimana hasil perbandingan performa antara model *Random Forest*, *XGBoost*, dan *hybrid ensemble* yang telah diimplementasikan dalam prediksi harga rumah di Jakarta?

1.3 Batasan Permasalahan

Batasan permasalahan pada penelitian ini adalah sebagai berikut:

- Dataset yang digunakan merupakan dataset harga rumah di Jakarta yang terdiri dari 10.000 data dengan 8 fitur.
- Penelitian ini hanya menggunakan data yang tersedia pada dataset dan tidak mempertimbangkan faktor eksternal lain di luar dataset.
- Model yang digunakan dalam penelitian ini adalah *Random Forest*, *Extreme Gradient Boosting (XGBoost)*, dan pendekatan *hybrid ensemble*.
- Evaluasi performa model dilakukan menggunakan metrik evaluasi regresi.
- Penelitian ini berfokus pada prediksi harga rumah di Jakarta dan tidak membahas aspek ekonomi makro maupun kebijakan sektor properti.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut:

- Mengimplementasikan model *Random Forest* dan *Extreme Gradient Boosting (XGBoost)* dalam melakukan prediksi harga rumah di Jakarta.
- Mengimplementasikan pendekatan *hybrid ensemble* yang mengombinasikan model *Random Forest* dan *XGBoost*.
- Membandingkan performa model *Random Forest*, *XGBoost*, dan *hybrid ensemble* dalam melakukan prediksi harga rumah di Jakarta.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

- Penelitian ini diharapkan dapat memberikan gambaran mengenai perbedaan karakteristik dan performa model *Random Forest*, *XGBoost*, serta pendekatan *hybrid ensemble* dalam prediksi harga rumah.
- Penelitian ini diharapkan dapat menjadi referensi bagi penelitian selanjutnya yang membahas penerapan maupun pengembangan metode *ensemble* dan *hybrid ensemble* pada permasalahan prediksi harga rumah.
- Penelitian ini diharapkan dapat memberikan informasi tambahan bagi pihak yang tertarik pada prediksi harga rumah berbasis *Machine Learning*, khususnya dalam pemilihan pendekatan pemodelan yang sesuai.

1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN
Bagian ini berfungsi sebagai pendahuluan penelitian yang menjelaskan alasan dilaksanakannya penelitian, permasalahan yang dikaji, ruang lingkup penelitian, tujuan yang hendak dicapai, manfaat penelitian, dan struktur laporan yang digunakan.
- Bab 2 LANDASAN TEORI
Bagian ini memuat berbagai teori, konsep, dan kajian yang relevan sebagai landasan dalam penelitian, meliputi *Machine Learning*, *Supervised Learning*, *Random Forest*, *XGBoost*, *Hybrid Ensemble* berbasis *Weighted Average*, metrik evaluasi, teknik *encoding* data kategorikal, serta *GridSearchCV*.
- Bab 3 METODOLOGI PENELITIAN
Bagian ini menjabarkan metode yang diterapkan selama penelitian, termasuk tahapan penelitian, karakteristik dataset yang digunakan, proses EDA, pra-pemrosesan data, pembagian dataset, optimasi parameter model, pembangunan model, dan prosedur evaluasi yang digunakan.
- Bab 4 HASIL DAN DISKUSI
Bagian ini menyampaikan hasil implementasi dan analisis penelitian

yang diperoleh dari seluruh tahapan yang telah dilakukan. Pembahasan meliputi hasil eksplorasi data, proses persiapan data, optimasi parameter, implementasi model, pembentukan model hybrid, serta evaluasi dan perbandingan performa model.

- Bab 5 SIMPULAN DAN SARAN

Bagian ini menyajikan ringkasan hasil penelitian dalam bentuk kesimpulan serta beberapa saran yang dapat digunakan sebagai bahan pertimbangan untuk penelitian berikutnya.

