

BAB I PENDAHULUAN

A. Latar Belakang

Rumah adalah kebutuhan dasar manusia yang memiliki peranan sangat penting. Selain berfungsi sebagai tempat tinggal, rumah juga kerap dijadikan sebagai instrumen investasi. Kebutuhan terhadap rumah kerap dimanfaatkan sebagai peluang investasi oleh para pelaku bisnis properti, yang mampu memberikan keuntungan signifikan [1]. Sehingga munculah harga yang tidak terprediksi pada harga rumah yang ada. Penentuan harga yang tidak akurat dapat menimbulkan ketidakseimbangan pasar, yang berpotensi merugikan baik pihak penjual maupun pembeli. Di Indonesia, proses penilaian harga rumah masih banyak dilakukan secara manual atau berdasarkan *intuisi* dan pengalaman pribadi. Pendekatan ini bersifat subjektif dan kurang konsisten, sehingga dibutuhkan solusi berbasis teknologi yang mampu melakukan prediksi harga secara lebih objektif dan akurat guna mendukung pengambilan keputusan baik penjual maupun pembeli [2].

Prediksi merupakan alat yang penting dan kini sering digunakan sebagai salah satu acuan dalam proses pengambilan keputusan pembelian. Kemampuan menghasilkan prediksi dengan tingkat akurasi yang tinggi dapat membantu dalam membuat keputusan yang lebih akurat dan tepat sasaran[3]. Model prediksi yang baik seharusnya mempertimbangkan berbagai variabel penting seperti lokasi, luas bangunan, luas tanah, jumlah kamar dan faktor-faktor lain yang relevan sebagai bahan pertimbangan dalam pengambilan keputusan pembelian [4]. Dalam konteks ini, *data mining* dapat dimanfaatkan untuk membangun model prediktif berdasarkan data historis dan menghasilkan estimasi harga rumah yang lebih akurat.

Metode yang digunakan dalam menganalisis data adalah *data mining*. *Data mining* adalah proses pengumpulan dan pengolahan data untuk di ekstrak menjadi informasi yang lebih spesifik. *Data mining* merupakan proses yang mencakup pengumpulan dan pemanfaatan data historis guna mengidentifikasi pola, keteraturan, atau hubungan tersembunyi dalam kumpulan data yang besar [5]. Data yang dihasilkan dari proses *data mining* dapat digunakan sebagai dasar untuk meningkatkan kualitas pengambilan keputusan di masa yang akan datang. *Data mining* umumnya dipahami sebagai proses menggali dan menemukan pola yang terdapat dalam sekumpulan data. Berdasarkan fungsi penggunaannya, *data mining* dapat dibagi menjadi beberapa kategori, yaitu deskripsi, estimasi atau prediksi, klasifikasi, klastering serta analisis asosiasi.

Berbagai algoritma *machine learning* sering dimanfaatkan dalam pemodelan prediktif, di antaranya *Decision Tree*, *Neural Network*, *K-Nearest Neighbor*, *Gradient Boosted Trees*, *Support Vector Regression*, serta *Random Forest*. *Decision tree* adalah salah satu *algoritma data mining* yang sering digunakan sebagai penyelesaian untuk klasifikasi maupun prediksi suatu masalah. *Algoritma*

ini memiliki struktur menyerupai pohon keputusan dengan akar, cabang, dan daun. *Algoritma* ini dikenal sederhana, mudah diinterpretasi, dan cukup andal dalam menangani data kompleks [6]. *Decision Tree* adalah metode yang mudah dipahami dan diinterpretasikan karena bentuk visualnya menyerupai diagram alur (*flowchart*), sehingga dapat dengan mudah diikuti dan dimengerti oleh cara berpikir manusia [7]. Sementara itu, *Neural Network* adalah *algoritma* yang lebih kompleks dan terinspirasi dari cara kerja jaringan saraf manusia. *Algoritma* ini memiliki keunggulan dalam mengenali pola nonlinear dan sering kali memberikan hasil prediksi dengan tingkat akurasi yang tinggi. *K-Nearest Neighbor* bekerja dengan pendekatan instancebased learning melalui perhitungan jarak kedekatan antara data baru dengan data latih. Metode ini relatif sederhana namun tetap efektif, khususnya untuk data dengan distribusi yang tidak terlalu rumit. *Gradient Boosted Trees* termasuk algoritma ensemble yang membangun model secara bertahap melalui kombinasi pohon keputusan lemah. Algoritma ini dikenal mampu meningkatkan akurasi prediksi secara signifikan, meskipun membutuhkan waktu komputasi yang lebih tinggi. Kemudian *Support Vector Regression* yang merupakan pengembangan dari *Support Vector Machine* untuk kasus regresi, efektif dalam menangani data berdimensi tinggi. Algoritma ini dapat menghasilkan prediksi presisi, meski sensitif terhadap pemilihan parameter. Sementara itu, *Random Forest* merupakan metode ensemble yang mengombinasikan banyak pohon keputusan sehingga menghasilkan prediksi yang stabil dan akurat. Untuk membandingkan kinerja beberapa *algoritma* tersebut, penelitian ini menggunakan aplikasi *Altair Studio* sebagai alat analisis.

Altair Studio merupakan perangkat lunak yang digunakan untuk mempelajari *data mining*, mulai dari analisis dan pengolahan data hingga pengembangan model prediktif, serta tersedia sebagai platform *open source* [8]. *Altair Studio* memiliki antarmuka yang ramah pengguna sehingga mempermudah proses pengolahan data. *Altair studio* digunakan dalam tahap pra-prosesing data dan evaluasi model. Ada beberapa tahapan yang dilakukan dalam pra-prosesing data antara lain *loading data*, *data cleaning*, *feature selection*, *splitting data*, *visualisasi model*, *testing data* dan *evaluasi model*. Tujuan dari *praprosesing* data adalah untuk mengolah data dari mentah menjadi data yang valid dan siap untuk digunakan, data mentah umumnya masih mengandung gangguan seperti data yang hilang, format yang tidak seragam, atau atribut yang tidak diperlukan, sehingga perlu dilakukan pembersihan dan penyesuaian terlebih dahulu.

Proses analisis untuk membandingkan performa antara algoritma tersebut akan diukur berdasarkan hasil evaluasi model prediktif yang dibuat menggunakan software *altair studio*. Terdapat 3 evaluasi yang dilakukan yaitu *performa*, *waktu komputasi* dan *kompleksitas model*, berdasarkan hasil evaluasi maka didapatkan algoritma mana yang memiliki performa lebih baik dalam memprediksi harga rumah. Evaluasi pada performa akan diukur berdasarkan dari hasil nilai *RMSE* dan *MAE* yang didapatkan oleh model algoritma. *RMSE* digunakan untuk mengukur

rata-rata akar kuadrat dari selisih antara nilai prediksi dan nilai aktual, sedangkan *MAE* digunakan untuk mengukur rata-rata dari selisih absolut antara nilai prediksi dan nilai aktual. Evaluasi pada waktu komputasi bertujuan untuk mengukur seberapa cepat model dapat belajar dari data latih dan seberapa efisiensi model dalam melakukan prediksi terhadap data baru. Evaluasi kompleksitas model bertujuan untuk menilai sejauh mana sebuah algoritma memerlukan sumber daya, struktur, dan parameter dalam proses pembentukan serta penerapan model. Melalui evaluasi ini, dapat diketahui keseimbangan antara tingkat akurasi dan efisiensi komputasi, sehingga pengguna dapat menghindari risiko *overfitting* maupun penggunaan sumber daya yang berlebihan.

Berdasarkan hal tersebut, penelitian ini bertujuan untuk menganalisis dan membandingkan performa beberapa algoritma *machine learning* dalam memprediksi harga rumah menggunakan *Altair Studio*. Diharapkan hasil penelitian ini dapat memberikan gambaran yang jelas mengenai kelebihan dan kekurangan masing-masing *algoritma*, sehingga dapat menjadi referensi dalam pengembangan sistem prediksi harga rumah yang lebih akurat dan efisien.

B. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana melakukan konfigurasi Altair Studio untuk penggunaan beberapa algoritma Machine Learning untuk memprediksi harga rumah?
2. Bagaimana melakukan evaluasi kinerja algoritma dalam melakukan prediksi harga rumah?

C. Batasan Masalah

Batasan masalah yang diterapkan dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini menggunakan software Altair Studio versi 2025.1.0.
2. Variabel yang digunakan dalam datasheet terdiri dari harga rumah, luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi dan garasi.
3. Datasheet yang digunakan adalah data harga rumah di daerah Jakarta Selatan dengan jumlah data 1010 baris data yang diperoleh dari kaggle dalam format excel.

D. Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Melakukan konfigurasi Altair Studio untuk penggunaan beberapa algoritma Machine Learning untuk memprediksi harga rumah.
2. Melakukan evaluasi kinerja algoritma dalam melakukan prediksi harga rumah.

E. Manfaat Penelitian

Adapun manfaat penelitian yang diharapkan setelah penelitian ini selesai dilakukan adalah:

1. Manfaat Teoritis
 - a. Menambah wawasan mengenai perbandingan kinerja algoritma Machine Learning dalam memprediksi harga rumah.
 - b. Menyediakan wawasan baru terkait penggunaan Altair Studio sebagai perangkat lunak untuk analisis data dan pembangunan model prediktif.
 - c. Menyajikan informasi baru terkait dengan evaluasi efektivitas algoritma dalam memprediksi harga rumah dengan menggunakan berbagai metrik kinerja.
2. Manfaat Praktis
 - a. Pengembang Sistem: Memberikan referensi dalam memilih algoritma yang lebih efisien dan akurat untuk digunakan dalam sistem prediksi harga rumah.
 - b. Pelaku Bisnis Properti: Membantu dalam pengambilan keputusan yang lebih baik terkait harga rumah dengan memanfaatkan model prediksi yang lebih tepat.
 - c. Penulis: Memberikan pengalaman langsung dalam menggunakan Altair Studio, serta meningkatkan keterampilan dalam menganalisis dan memodelkan data menggunakan algoritma Machine Learning.