

BAB I

PENDAHULUAN

1.1 Latar Belakang

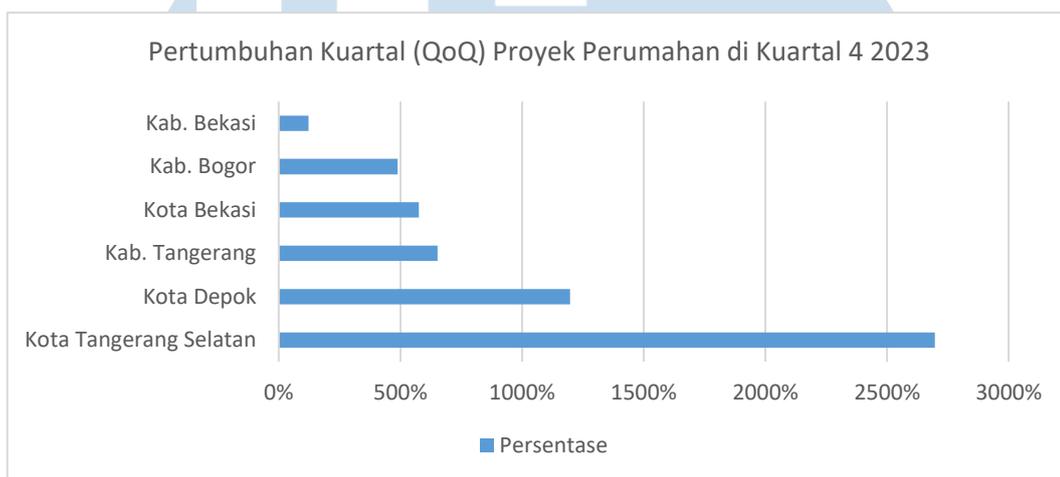
Di era digital, kemajuan dalam teknologi telah mengubah secara drastis cara manusia melengkapi kebutuhan primernya. Kebutuhan primer merupakan kebutuhan pokok termutlak yang dibutuhkan manusia guna melanjutkan keberlangsungan hidup. Kebutuhan primer yaitu pangan, sandang, dan papan masih menjadi pokok yang mesti selalu menempati urutan teratas dalam hal permintaan kebutuhan manusia. Contoh kebutuhan primer meliputi pangan yaitu sumber makanan bagi manusia, sandang yaitu pakaian yang digunakan dalam kehidupan manusia, dan yang terakhir papan yaitu tempat tinggal yang layak[1].

Berdasarkan paragraf sebelumnya, kebutuhan primer papan merupakan tempat tinggal yang layak seperti contohnya rumah. Rumah merupakan tempat tinggal yang memberikan perlindungan secara fisik dari cuaca ekstrem, keamanan dari ancaman luar, dan kenyamanan untuk beristirahat. Namun, seiring perkembangan zaman, rumah juga berfungsi sebagai tempat berkumpul keluarga dan sebagai simbol status sosial[2].

Masyarakat Indonesia lebih memilih tempat tinggal di rumah tapak. Pernyataan tersebut didukung oleh survey yang dilakukan Telkomsel pada 2023. Hasil dari survei itu menyatakan bahwa Jakarta dengan yang paling padat penduduknya, serta paling banyak dan paling bervariasi terhadap pengembangan jenis hunian apartemen, hanya 7% saja yang memilih apartemen sebagai hunian[3]. Beberapa alasan dari 93% responden yang memilih rumah tapak sebagai pilihan hunian diantaranya; memiliki luas hunian yang lebih besar, fleksibilitas dalam penggunaan ruang, lingkungan yang lebih tenang, potensi investasi yang besar dan juga akses kendaraan pribadi yang mudah[3].

Salah satu kota dengan peminat rumah tapak terbanyak adalah Kota Tangerang Selatan. Hal ini menjadi alasan pemilihan dari objek penelitian di Kota Tangerang Selatan dan diperkuat oleh penelitian yang dilakukan Pinhome Research tentang pertumbuhan proyek perumahan yang berkembang pesat di Kota Tangerang Selatan. Berdasarkan penelitian tersebut, Kota Tangerang Selatan menempati posisi

pertama dalam pertumbuhan proyek perumahan sebesar 2697% di Kuartal 4 2023. Salah satu alasan utama perkembangan dari proyek perumahan di Kota Tangerang Selatan adalah adanya pembangunan infrastruktur Jalan Tol Serpong-Bogor via Parung, hal ini memberikan dampak yang kuat pada penambahan proyek perumahan di Kota Tangerang Selatan dan Kabupaten Bogor[4].



Gambar 1. 1 Pertumbuhan Kuartal (QoQ) Proyek Perumahan di Kuartal 4 2023

Dengan dipilihnya Kota Tangerang Selatan sebagai objek penelitian ini, wilayah seperti Ciputat, Pamulang, BSD, Serpong, Alam Sutera, Bintaro, dan Pondok Aren[5] menjadi fokus utama di dalam penelitian ini. Pemilihan tujuh wilayah Ciputat, Pamulang, BSD, Serpong, Alam Sutera, Bintaro, dan Pondok Aren sebagai fokus utama penelitian ini didasarkan pada urgensi dan relevansi mereka sebagai episentrum pertumbuhan properti di Tangerang Selatan. Wilayah-wilayah ini bukan hanya sekadar area perumahan, melainkan kawasan kota mandiri dan satelit yang terus berkembang pesat, didorong oleh permintaan tinggi dari populasi yang bermigrasi dari Jakarta untuk mencari hunian yang lebih terjangkau namun tetap memiliki aksesibilitas prima. Perkembangan infrastruktur strategis, seperti jaringan jalan tol yang menghubungkan Serpong dengan area vital lainnya dan jalur transportasi publik seperti KRL Commuter Line, secara signifikan meningkatkan konektivitas dan menjadikan lokasi-lokasi ini pilihan utama bagi keluarga muda dan para komuter[6].

Signifikansi ketujuh area ini diperkuat oleh data pertumbuhan proyek perumahan yang masif, di mana Tangerang Selatan menempati posisi teratas

dengan pertumbuhan mencapai 2697% pada kuartal keempat tahun 2023. Lonjakan ini, ditambah dengan kenaikan median harga rumah yang signifikan dari Rp 1,2 miliar pada tahun 2021 menjadi Rp 2,0 miliar pada tahun 2025, menunjukkan dinamika pasar yang sangat aktif dan kompleks. Oleh karena itu, memfokuskan penelitian pada ketujuh wilayah ini menjadi krusial untuk dapat menangkap esensi dari tren pasar properti di Tangerang Selatan secara komprehensif, sehingga model prediksi harga yang dihasilkan dapat menjadi relevan dan akurat bagi masyarakat luas. [7].

Dengan seiring berkembangnya teknologi, bidang properti menjadi salah satu industri yang memanfaatkan kemajuan tersebut. Di masa kini, orang dapat mencari rumah yang dijual dengan mudah secara daring (*online*). dengan menggunakan platform internet berbasis website seperti Rumah.com, OLX, Rumah123, Lamudi dan yang lainnya. Platform internet dapat memudahkan penjual dan pembeli dalam mengakses dan mencari harga, spesifikasi, dan lokasi dari rumah yang diinginkan, keuntungan platform internet dalam bidang properti menjadi salah satu cara untuk meningkatkan industri properti. Akan tetapi dengan banyaknya varian harga dari jenis rumah yang dijual, membuat para pembeli merasa bingung dengan pilihan tersebut, apakah uang yang dikeluarkan sepadan dengan spesifikasi rumah yang diinginkan. Oleh karena itu, diperlukan sebuah algoritma *machine learning* yang dapat melakukan prediksi harga rumah sesuai dengan spesifikasi yang diinginkan oleh calon pembeli maupun juga penjual[8].

Machine Learning adalah sebuah pendekatan berbasis algoritma dan model statistik yang memungkinkan komputer untuk menyelesaikan tugas tertentu tanpa memerlukan instruksi eksplisit dari *programmer*. Salah satu aplikasi utama dari *machine learning* adalah prediksi, yang bisa diterapkan dalam masalah klasifikasi ataupun regresi[9]. Penelitian ini mengkhususkan diri pada masalah regresi, dengan fokus pada prediksi harga rumah di Kota Tangerang Selatan wilayah Ciputat, Pamulang, BSD, Serpong, Alam Sutera, Bintaro, dan Pondok Aren. Dalam penelitian ini, algoritma *machine learning* yang digunakan untuk melakukan prediksi harga rumah adalah *Multiple Linear Regression*, *Gradient Boosting*, dan *Random Forest*. Ketiga algoritma ini dipilih karena memiliki karakteristik yang

berbeda dalam menangani data dan mampu memberikan hasil yang beragam dalam masalah regresi[10].

Multiple Linear Regression merupakan algoritma yang digunakan untuk memodelkan hubungan antara satu variabel dependen dan lebih dari satu independen dengan menyesuaikan persamaan linier pada data yang diamati[11]. Sementara itu, berbeda dengan algoritma *Gradient Boosting* yang menghasilkan model prediksi dalam bentuk ensemble dari model prediksi yang lemah. *Gradient Boosting* bertujuan untuk meningkatkan model dengan mengubah prediksi yang lemah menjadi lebih kuat, dengan memfokuskan pada kesalahan antara nilai prediksi dan nilai aktual pada setiap iterasi[12]. Algoritma *Random Forest* merupakan sekumpulan dari banyak pohon keputusan (*Decision Tree*) yang digabungkan untuk mengambil keputusan berdasarkan hasil mayoritas dari pohon yang dilatih, hal ini dapat mengurangi variansi yang ada pada model, sehingga memberikan prediksi yang lebih stabil[12].

Pada penelitian terdahulu berjudul “Analisis Perbandingan Metode Regresi Linier, Random Forest Regression dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah” yang ditulis oleh Evita Fitri. Penelitian ini membandingkan tiga metode regresi *machine learning* untuk memprediksi harga rumah, yaitu *Multiple Linear Regression*, *Random Forest*, dan *Gradient Boosted Trees Regression*. Dataset yang digunakan terdiri dari 2011 data dengan 6 fitur seperti harga rumah, luas tanah, luas bangunan, jumlah kamar mandi, kamar tidur, dan garasi[13]. Hasil pengujian menunjukkan bahwa *Random Forest Regression* menghasilkan nilai RMSE terkecil, yaitu 0,440, dengan akurasi prediksi mencapai 81,5%. Metode lain, *Linear Regression* dan *Gradient Boosted Trees Regression*, memperoleh RMSE masing-masing sebesar 0,515 dan 0,508, dengan akurasi yang lebih rendah dibandingkan *Random Forest*.

Penelitian lain dengan judul “*House Price Prediction using Machine Learning Algorithms*” yang ditulis oleh Angulakshmi M, Deepa M, Mala Serene I, Thilagavathi M, dan Aarthi P. Penelitian ini menggunakan empat algoritma *machine learning* untuk memprediksi harga rumah, yaitu *Linear Regression*,

Polynomial Regression, *Decision Tree*, dan *Random Forest*. Dataset yang digunakan berasal dari *Open Data* Pakistan, dengan 168,447 data dan 20 fitur yang meliputi berbagai atribut properti seperti lokasi, ukuran, dan jumlah kamar[14]. Hasil evaluasi menunjukkan bahwa algoritma *Random Forest* menghasilkan akurasi tertinggi sebesar 89%, jauh lebih baik dibandingkan dengan algoritma lainnya, seperti *Decision Tree* yang memiliki kesalahan rata-rata yang lebih tinggi. *Random Forest* juga menunjukkan nilai error terkecil dalam hal mean absolute error, menjadikannya pilihan yang paling akurat untuk prediksi harga rumah.

Penelitian lain dengan judul “*Predicting property prices with machine learning algorithms*” yang ditulis oleh Winky K.O. Ho, Bo-Sin Tang, dan Siu Wai Wong. Penelitian ini menggunakan tiga algoritma *machine learning*, yaitu *Support Vector Machine*, *Random Forest*, dan *Gradient Boosting Machine*, untuk memprediksi harga properti. Data yang digunakan mencakup 39.554 transaksi properti yang terjadi di Hong Kong selama lebih dari 18 tahun[12]. Hasil penelitian menunjukkan bahwa *Random Forest* dan *Gradient Boosting Machine* memberikan akurasi yang lebih baik dibandingkan *Support Vector Machine*. *Random Forest* dan *Gradient Boosting* mencapai nilai R^2 tertinggi masing-masing 0,90333 dan 0,90365, dengan MSE dan RMSE yang lebih rendah, menjadikannya metode yang lebih unggul untuk prediksi harga properti.

Penelitian lain dengan judul “*Research on Prediction and Analysis of Real Estate Market Based on the Multiple Linear Regression Model*” yang ditulis oleh Guangjie Liu. Penelitian ini menggunakan algoritma *Multiple Linear Regression* untuk memprediksi harga properti di pasar properti. Data yang digunakan mencakup faktor-faktor seperti tingkat pendapatan bulanan penduduk, pendapatan *disposable* per kapita, pengeluaran per kapita untuk perumahan, dan luas area properti yang terbangun[15]. Model prediksi harga rumah ini menunjukkan akurasi yang tinggi, dengan kesalahan prediksi maksimum tidak lebih dari 8%, yang menandakan efektivitas model dalam memprediksi perubahan harga properti.

Penelitian lain dengan judul “Perbandingan Kinerja Algoritma Extreme Gradient Boosting dan Random Forest Untuk Prediksi Harga Rumah di

Jabodetabek” yang ditulis oleh Dhiwa Aqsha. Data yang digunakan dalam penelitian ini diambil dari *dataset* "Daftar Harga Rumah Jabodetabek" yang tersedia di *Kaggle*, yang mencakup 3.554 sampel dan 27 variabel, seperti harga properti, luas bangunan, dan jumlah kamar tidur. Hasil evaluasi menunjukkan bahwa model *Random Forest* memberikan hasil yang lebih baik dengan nilai *Mean Absolute Error* (MAE) sebesar 95.200.513,25, *Mean Squared Error* (MSE) 1,47e+19, dan R^2 0,77, dibandingkan dengan *XGBoost* yang memiliki MAE 121.836.703,27, MSE 3,03e+19, dan R^2 0,52[16].

Data yang akan digunakan dalam penelitian ini adalah data iklan rumah yang tayang pada situs Lamudi.co.id. Adapun fitur yang digunakan antara lain, data *link* iklan, judul iklan, lokasi, luas bangunan, luas tanah, jumlah lantai, kamar mandi, kamar tidur, garasi, taman, lapangan terbuka, keamanan, dan harga rumah. Algoritma *Linear Regression*, *Gradient Boosting*, dan *Random Forest* akan digunakan dalam penelitian ini karena ketiga algoritma ini banyak dipakai pada penelitian terdahulu dan mampu memberikan performa yang cukup baik dalam kasus prediksi harga rumah. Selain itu, akan dilakukan evaluasi algoritma antara algoritma *Linear Regression*, *Gradient Boosting*, dan *Random Forest* untuk mengetahui algoritma dengan performa terbaik dalam memprediksi harga rumah. Algoritma akan di *deploy* ke dalam website untuk melakukan prediksi harga rumah sesuai dengan spesifikasi yang pengguna masukkan ke dalam *website*.

1.2 Rumusan Masalah

Dengan latar belakang yang sudah dijelaskan diatas, maka rumusan masalah dari penelitian ini adalah sebagai berikut:

1. Bagaimana hasil performa yang dihasilkan dari model algoritma *Multiple Linear Regression*, *Gradient Boosting*, dan *Random Forest* dengan data rumah di kota Tangerang Selatan?
2. Faktor apa yang paling mempengaruhi prediksi harga di kota Tangerang Selatan?

3. Bagaimana implementasi model algoritma *Multiple Linear Regression*, *Gradient Boosting*, dan *Random Forest* dalam penelitian ini menggunakan *prototype*?

1.3 Batasan Masalah

Batasan masalah yang terdapat di penelitian ini adalah sebagai berikut:

1. Implementasi prediksi harga rumah yang berbasis *website* dengan *framework Flask*.
2. Data untuk prediksi harga rumah diambil dari data rumah Kota Tangerang Selatan dengan wilayah seperti Ciputat, Pamulang, BSD, Serpong, Alam Sutera, Bintaro, dan Pondok Aren menjadi fokus utama.
3. Data untuk prediksi harga rumah diambil dari data jual beli rumah dari *platform Lamudi* pada bulan April-Mei 2025.
4. Faktor yang menjadi indikator dalam penelitian ini adalah data lokasi, luas bangunan, luas tanah, jumlah lantai, kamar mandi, kamar tidur, garasi, taman, lapangan terbuka, keamanan, dan harga rumah.
5. *Prototype* akan menampilkan hasil berupa harga rumah yang merupakan hasil estimasi atau prediksi dan parameter yang dimasukkan oleh *user*.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Berikut adalah tujuan dari penelitian ini:

1. Memberikan hasil dari performa prediksi harga rumah di kota Tangerang selatan dengan algoritma *Multiple Linear Regression*, *Gradient Boosting* dan *Random Forest*.
2. Mencari faktor yang berpengaruh terhadap prediksi harga rumah di kota Tangerang Selatan
3. Memberikan prediksi harga rumah sesuai dengan *input* dari *user* menggunakan *prototype* yang dibentuk didalam penelitian ini.

1.4.2 Manfaat Penelitian

Berikut adalah manfaat dari penelitian ini:

1. Memberikan fitur prediksi harga rumah sesuai dengan *input* dari *user* menggunakan *prototype* yang dibentuk didalam penelitian ini dan dapat dikembangkan lagi kedepannya.
2. Menjadi inspirasi terhadap penelitian tentang properti kedepannya untuk membantu meningkatkan properti Indonesia.

1.5 Sistematika Penulisan

BAB 1 PENDAHULUAN

Dalam pendahuluan terdapat penjelasan mengenai latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat dari penelitian ini.

BAB II LANDASAN TEORI

Pada landasan teori terdapat penjelasan mengenai dasar-dasar teori yang digunakan dalam penelitian ini seperti teori properti, lamudi dan juga algoritma yang digunakan

BAB III METODOLOGI PENELITIAN

Pada metodologi penelitian ini menjelaskan beberapa tahap penelitian secara berurut. Diawali dari tinjauan pusaka, algoritma yang digunakan, software yang digunakan, hingga penelitian terdahulu.

BAB IV ANALISIS DAN HASIL PENELITIAN

Bab IV berisi implementasi mengenai analisis dan hasil dari penelitian, yang sesuai dengan proses yang telah dijelaskan pada Bab III dan tujuan penelitian yang telah didefinisikan pada Bab II.

BAB V SIMPULAN DAN SARAN

Bab V berisi kesimpulan dari hasil penelitian dan juga memberikan saran yang dapat digunakan untuk terhadap penelitian kedepannya.