

# BAB I

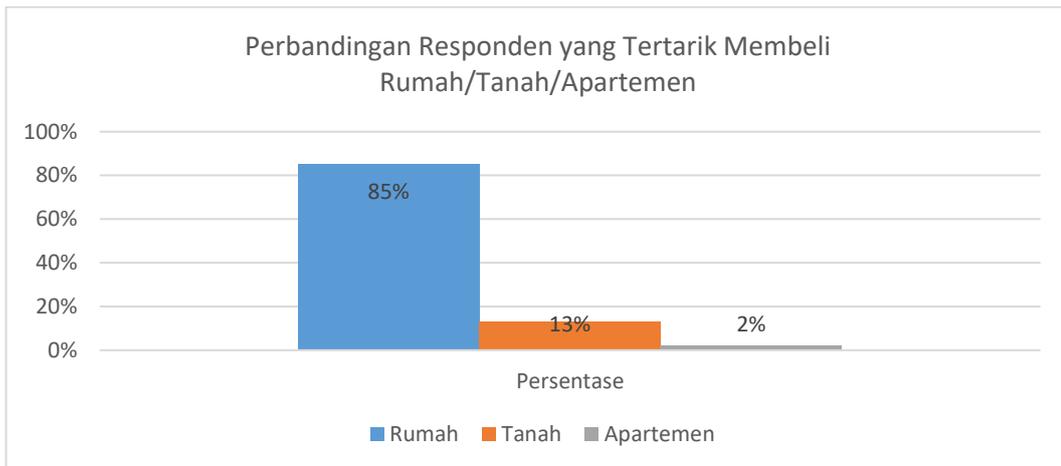
## PENDAHULUAN

### 1.1 Latar Belakang

Dalam beberapa tahun terakhir, perkembangan teknologi berkembang dengan pesat. Agar dapat bertahan hidup di zaman yang serba cepat ini, salah satu cara yang bisa dilakukan manusia adalah dengan memenuhi kebutuhan primernya. Kebutuhan primer merupakan kebutuhan yang wajib terpenuhi, apabila tidak terpenuhi, maka manusia akan mengalami kesulitan dalam hidupnya [1]. Kebutuhan primer terdiri atas kebutuhan sandang, papan, dan pangan [1]. Contoh dari kebutuhan primer, yaitu kebutuhan sandang yang berupa pakaian yang layak pakai, kebutuhan papan yang berupa tempat tinggal yang layak, dan kebutuhan pangan yang berupa makanan yang layak untuk di konsumsi.

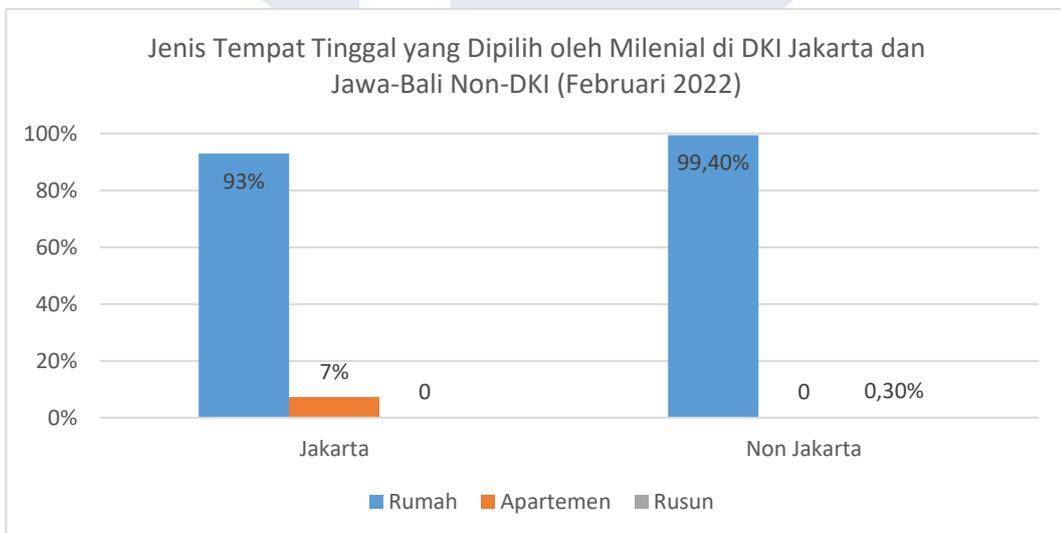
Berdasarkan pemaparan pada paragraf sebelumnya, kebutuhan papan merupakan kebutuhan manusia akan tempat tinggal yang layak. Salah satu contoh dari kebutuhan papan adalah rumah. Menurut artikel [2], rumah merupakan kebutuhan pokok manusia berupa bangunan yang berfungsi sebagai tempat tinggal yang nyaman, wadah pembinaan keluarga, simbol harkat dan martabat penghuninya, serta aset berharga bagi pemiliknya. Di Indonesia, ada delapan jenis rumah dan hunian, yaitu rumah tapak, *town house*, *cluster*, apartemen, rumah susun, kondotel, rumah kantor, dan kos-kosan [3].

Jenis rumah yang paling umum ditemukan di Indonesia adalah rumah tapak. Pernyataan tersebut didukung oleh survei yang dilakukan oleh Rumah.com pada 2021 lalu. Hasil dari penelitian tersebut menyatakan bahwa dari 1031 responden, 85% diantaranya lebih memilih untuk membeli rumah tapak ketimbang apartemen [3]. Beberapa alasan dari 85% responden tadi lebih memilih rumah tapak diantaranya: akan mendapat luas lebih besar, tidak suka tinggal di gedung tinggi, tidak bisa diperluas di masa depan, status kepemilikan dibatasi waktu, biaya bulanan yang lebih besar, tidak suka lingkungan yang padat, dan kurang privasi [3].



Gambar 1.1 Perbandingan Responden Membeli Rumah/Tanah/Apartemen

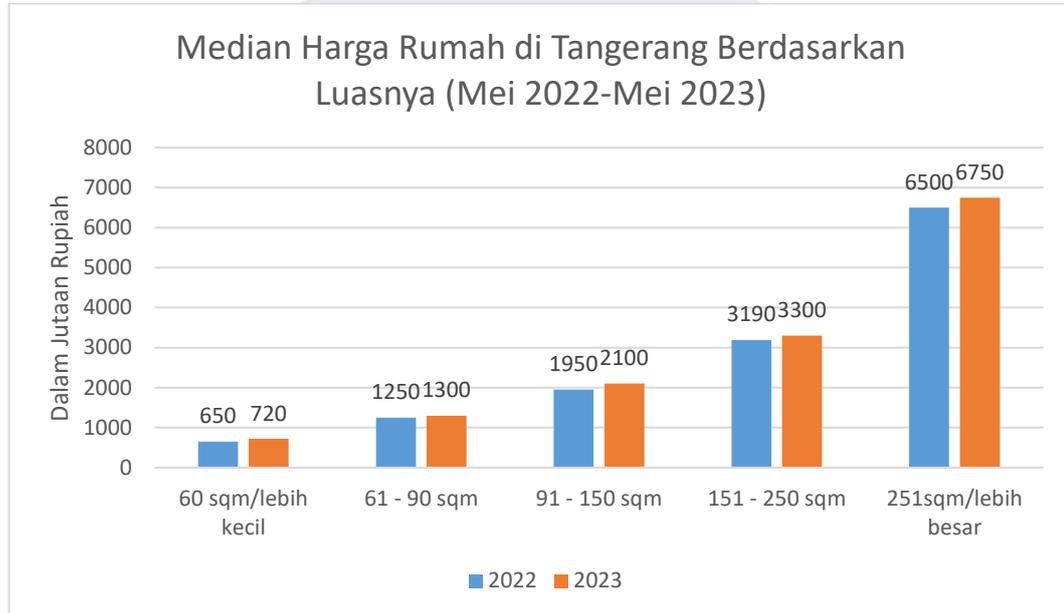
Selain itu, pernyataan sebelumnya juga didukung dengan adanya temuan serupa yang dikemukakan oleh Telkomsel melalui survei yang dilakukan pada 2022 lalu [4]. Dari 390 responden, sebanyak 99,4% warga Jawa-Bali lebih memilih rumah tapak dibandingkan apartemen, sedangkan sebanyak 93% warga Jakarta juga memilih rumah tapak dibandingkan apartemen [4].



Gambar 1.2 Jenis Tempat Tinggal yang Dipilih oleh Milenial (Februari 2022)

Salah satu lokasi yang banyak dicari untuk dijadikan tempat tinggal adalah Kabupaten Tangerang. Pemilihan lokasi dari objek penelitian di Kabupaten Tangerang diperkuat oleh survei yang dilakukan oleh Rumah123.com yang menyatakan bahwa harga rumah di Tangerang cenderung mengalami kenaikan setiap tahunnya, berdasarkan luasnya [4]. Berdasarkan data tersebut, rumah dengan

luas tanah 60 m<sup>2</sup> hingga 251 m<sup>2</sup> terlihat mengalami kenaikan harga dibandingkan dengan tahun 2022 [5]. Untuk rinciannya, dapat dilihat pada diagram di bawah.



Gambar 1.3 Median Harga Rumah di Tangerang (Mei 2022-Mei 2023)

Perkembangan teknologi yang pesat membuat disrupsi pada beberapa sektor di industri. Salah satu sektor yang terdampak dari adanya disrupsi teknologi ini adalah sektor properti atau *real estate*. Di masa kini, orang dapat mencari rumah yang dijual secara daring (*online*). Adapun, *platform* jual beli rumah yang ada di Indonesia antara lain adalah rumah123.com, lamudi.co.id, realoka.com, dan sebagainya. Orang-orang dapat dengan mudah mencari harga, spesifikasi, dan lokasi dari rumah yang akan calon pembeli beli pada situs tersebut. Akan tetapi, karena banyaknya variasi dari jenis rumah yang dijual, membuat calon pembeli merasa bimbang apakah uang yang calon pembeli keluarkan sepadan dengan apa yang calon pembeli dapatkan. Hal tersebut dapat terjadi karena tidak adanya harga yang menjadi harga patokan ketika calon pembeli ingin membeli rumah [6]. Oleh karena itu, diperlukanlah sebuah algoritma *machine learning* yang dapat melakukan prediksi dari harga rumah berdasarkan spesifikasi dan lokasi dari rumah yang akan dibeli oleh calon pembeli.

Secara sederhananya, *machine learning* (ML) merupakan algoritma dan/atau model statistik yang digunakan oleh komputer untuk melakukan tugas tertentu

tanpa perlu untuk diprogram secara eksplisit oleh *programmer* [7]. Salah satu penerapan dari *machine learning* adalah untuk melakukan prediksi. Jenis prediksi yang dilakukan bisa digunakan untuk masalah klasifikasi ataupun regresi [7]. Penelitian ini akan berfokus pada permasalahan regresi, karena hasil dari luaran penelitian ini berupa prediksi dari harga rumah di wilayah Kabupaten Tangerang. Adapun, algoritma yang digunakan dalam permasalahan regresi antara lain: regresi linear [8], *Random Forest* [8], *Decision Tree* [8], *K-Nearest Neighbor* (KNN) [8], dan *XGBoost* [9].

Algoritma *XGBoost* merupakan algoritma yang dalam penerapannya menggunakan *gradient boosting machine* untuk melakukan tugas regresi maupun klasifikasi [9]. Sementara itu, regresi linear adalah algoritma yang digunakan untuk menganalisis estimasi nilai variabel dependen terhadap variabel independen [10]. Berbeda dengan algoritma *XGBoost*, algoritma *Random Forest* merupakan sekumpulan dari banyak pohon keputusan (*Decision Tree*) yang dilatih secara paralel, dan akan mengambil keputusan berdasarkan hasil mayoritas dari pohon yang dilatih [11]. Algoritma *Decision Tree* memiliki bentuk menyerupai pohon yang di tiap *internal node*-nya merepresentasikan atribut dan cabang dari hasil yang diuji [12]. Algoritma KNN akan menghitung jarak terdekat dari variabel target dengan data yang ada di dataset [11].

Sudah banyak penelitian terdahulu yang telah mengkaji performa dari algoritma *machine learning* dalam tugas memprediksi harga rumah. Dalam penelitian “Analisis Perbandingan Metode Regresi Linear, *Random Forest Regression*, dan *Gradient Boosted Trees Regression Method* untuk Prediksi Harga Rumah”, performa algoritma *Random Forest* memberikan akurasi sebesar 81,5% [10]. Sementara itu, algoritma KNN yang dianalisis dan diteliti dalam penelitian “*House Price Prediction using Multiple Linear Regression and KNN*”, memberikan akurasi yang lebih rendah, yaitu sebesar 63% [11]. Pada penelitian lain yang berjudul “*House Price Prediction Using Hedonic Pricing Model and Machine Learning Techniques*”, algoritma *XGBoost* memberikan akurasi sebesar 84,1% [13]. Selain itu, pada penelitian yang berjudul “*Decision Tree Regressor Compared with Random Forest Regressor for House Price Prediction in Mumbai*”, performa

algoritma *Decision Tree* memberikan akurasi sebesar 72,45% memprediksi harga rumah [14]. Terakhir, pada penelitian yang berjudul “*Analisis Prediksi Harga Rumah Sesuai Spesifikasi Menggunakan Multiple Linear Regression*”, algoritma *Multiple Linear Regression* memberikan nilai akurasi sebesar 66% [15]. Untuk detailnya, dapat dilihat pada tabel 1.1 di bawah.

Tabel 1.1 Komparasi Performa Algoritma *Machine Learning*

Algoritma	Akurasi
<i>Random Forest Regressor</i> [10]	81,5%
<i>K-Nearest Neighbor (KNN)</i> [11]	63%
<i>XGBoost</i> [13]	84,1%
<i>Decision Tree Regressor</i> [14]	72,45%
<i>Multiple linear regression</i> [15]	66%

Berdasarkan tabel di atas, performa algoritma *XGBoost* memiliki akurasi tertinggi, diikuti oleh *Random Forest*, *Decision Tree*, dan *Multiple Linear Regression*, sementara KNN menunjukkan akurasi paling rendah.

Data yang akan digunakan dalam penelitian ini adalah data iklan rumah yang tayang pada situs rumah123.com. Adapun fitur yang digunakan antara lain, data harga rumah, lokasi rumah, kamar tidur, kamar mandi, luas tanah, luas bangunan, sertifikat, daya listrik, jumlah lantai, kondisi properti, *carport*, kamar tidur pembantu, dan kamar mandi pembantu. Algoritma *Multiple Linear Regression*, *Random Forest*, dan *XGBoost* akan digunakan dalam penelitian ini karena ketiga algoritma ini banyak dipakai pada penelitian terdahulu dan mampu memberikan performa yang cukup baik dalam kasus prediksi, termasuk prediksi harga rumah. Selain itu, akan dilakukan perbandingan algoritma antara algoritma *Multiple Linear Regression*, *Random Forest*, dan *XGBoost* untuk mendapatkan algoritma dengan performa terbaik dalam memprediksi harga rumah. Algoritma terbaik akan di *deploy* ke dalam *website* untuk melakukan prediksi harga rumah sesuai dengan spesifikasi yang pengguna masukkan ke dalam *website*.

## 1.2 Rumusan Masalah

Berdasarkan pemaparan latar belakang sebelumnya, berikut ini merupakan rumusan masalahnya:

1. Bagaimana cara memilih algoritma *machine learning* terbaik untuk prediksi harga rumah di Kabupaten Tangerang?
2. Bagaimana cara mengimplementasikan model prediksi harga rumah di Kabupaten Tangerang dalam bentuk *website* dengan *streamlit*?
3. Bagaimana cara mengukur keakuratan dari model prediksi harga rumah di Kabupaten Tangerang?

## 1.3 Batasan Masalah

Berikut merupakan batasan masalah yang diterapkan dalam penelitian ini:

1. Data harga rumah yang dipasang di situs rumah123.com pada bulan Januari – Februari 2025.
2. Wilayah rumah di Kabupaten Tangerang.
3. Algoritma yang digunakan dalam penelitian ini adalah algoritma *Multiple Linear Regression*, *Random Forest*, *Extreme gradient boosting* (XGBoost).
4. Metrik akurasi yang digunakan adalah  $R^2$ , *Mean Absolute Percentage Error* (MAPE), *Root Mean Squared Error* (RMSE).

## 1.4 Tujuan dan Manfaat Penelitian

### 1.4.1 Tujuan Penelitian

Berikut ini merupakan tujuan dari penelitian yang berjudul “Optimasi dan Pemilihan Algoritma *Machine Learning* untuk Prediksi Harga Rumah di Kabupaten Tangerang”:

1. Mengetahui cara memilih algoritma *machine learning* terbaik untuk prediksi harga rumah di Kabupaten Tangerang.
2. Mengetahui cara mengimplementasikan model prediksi harga rumah di Kabupaten Tangerang dalam bentuk *website* dengan *streamlit*.
3. Mengetahui cara mengukur keakuratan dari model prediksi harga rumah di Kabupaten Tangerang.

#### 1.4.2 Manfaat Penelitian

Berikut ini merupakan manfaat dari penelitian yang berjudul “Optimasi dan Pemilihan Algoritma *Machine Learning* untuk Prediksi Harga Rumah di Kabupaten Tangerang”:

1. Memberikan wawasan mengenai perbandingan dari performa algoritma *Multiple Linear Regression*, *Random Forest*, dan *Extreme Gradient Boosting (XGBoost)* dalam prediksi harga rumah di Kabupaten Tangerang, baik dalam ruang lingkup Universitas Multimedia Nusantara ataupun masyarakat umum.
2. Dapat dijadikan sebagai acuan dalam penelitian sejenis di masa yang akan datang.

#### 1.5 Sistematika Penulisan

##### BAB I PENDAHULUAN

Bab I akan menjelaskan mengenai fenomena yang ada dari permasalahan dalam penelitian, menuliskan penelitian-penelitian terkait, menuliskan rumusan masalah, batasan masalah, tujuan penelitian, dan manfaat penelitian.

##### BAB II LANDASAN TEORI

Bab II akan menjelaskan teori dasar yang dipakai pada penelitian ini. Teori yang dijelaskan adalah teori mengenai rumah, *framework* CRISP-DM, algoritma yang digunakan dalam penelitian, teknik evaluasi performa model, dan penelitian terdahulu.

##### BAB III METODOLOGI

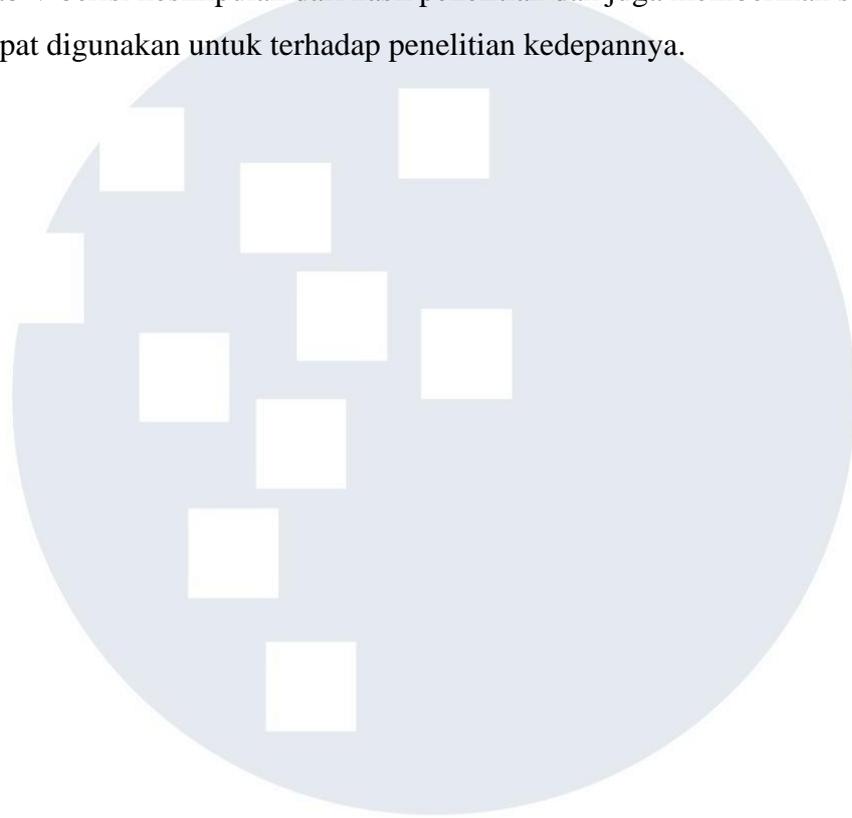
Bab III berisi penjelasan mengenai metode yang digunakan pada penelitian ini serta memberikan gambaran umum terkait dengan proses penelitian yang akan dilakukan dalam penelitian.

##### 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.



# UMMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA