

BAB I

PENDAHULUAN

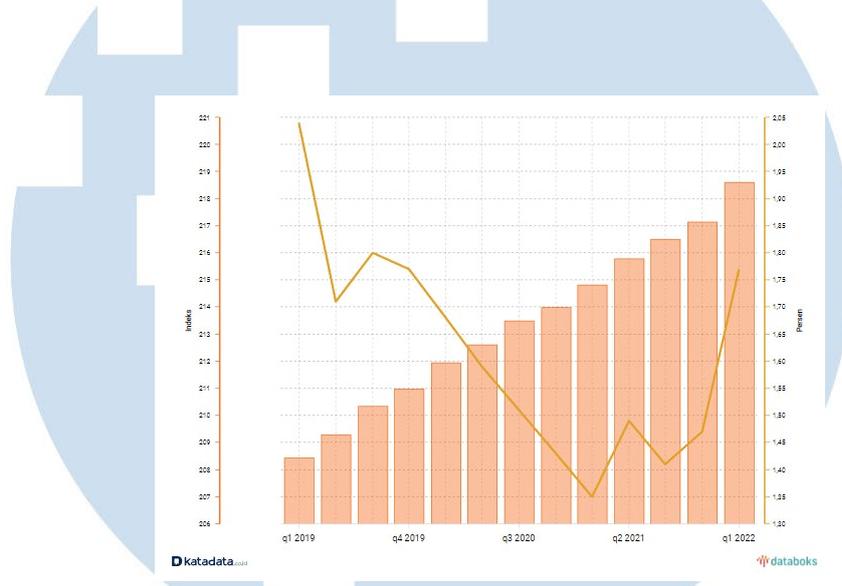
1.1 Latar Belakang

Melihat era modern dimana dunia dikelilingi dengan *Fourth Industrial Revolution* atau yang biasa disebut Revolusi Industri 4.0, membuat perkembangan teknologi semakin meningkat dan juga jumlah pengadopsian teknologi semakin menyebar. Banyaknya fenomena disrupsi pada berbagai industri ataupun perusahaan menjadi salah satu dari mulainya revolusi ini. Disrupsi atau disruptif yang dimaksud ini adalah sebuah kondisi yang mengharuskan sebuah bisnis atau perusahaan untuk melakukan inovasi menyesuaikan dengan perkembangan zaman yang membuat bisnis atau perusahaan tidak hanya berjalan untuk memenuhi kebutuhan pada waktu sekarang ini tetapi bisa melakukan antisipasi pada kebutuhan kedepannya [1].

Salah satu industri yang mengalami disrupsi adalah industri *real estate* atau properti, khususnya rumah. Rumah merupakan salah satu dari kebutuhan hidup terpenting bagi semua orang. Menjadi salah satu kebutuhan primer untuk manusia, Rumah memiliki fungsi sebagai tempat tinggal. Tangerang Selatan menjadi salah satu lokasi yang sangat diincar oleh masyarakat dimana dengan harga yang lebih murah dibandingkan di Jakarta dengan fasilitas dan infrastruktur yang cukup baik membuat kota ini menjadi pilihan masyarakat [2]. Menurut data dari lamudi, Tangerang Selatan menjadi kota dengan nilai investasi yang tinggi dimana mendapatkan pertumbuhan sebesar 3,5% dari bulan April sampai Mei 2022 [3].

Menurut data dari Survei Harga Properti Residensial Bank Indonesia, harga properti residensial pada primer bertumbuh dan mengalami kenaikan pada Q1 2022 dan menurut Indeks Harga Properti Residensial pada Q1 2022 mengalami pertumbuhan tahunan sebesar 1,77%. Pertumbuhan tersebut menjadi pertumbuhan terbesar semenjak pandemi COVID-19 bermula dimana pada Q1 2020 terdapat 1.68% pertumbuhan harga *property* melalui Indeks Harga Properti Residensial [4]. Dari survei terbaru yaitu Q4 tahun 2022, Bank Indonesia mencatat kenaikan indeks

harga properti residential (IHPR) sebesar 2% secara *year on year*, angka ini melebihi IHPR pada kuartal sebelumnya yang mendapatkan sebesar 1.94% [5]. Bank Indonesia juga mencatat penurunan penjualan tahunan sebesar 4,54% pada Q4 2022 dan menurun 13,58% jika dibandingkan Q3 2022.



Gambar 1.1 Grafik Survei Harga Properti Resedensial Bank Indonesia

Sumber : [3]

Mengikuti perkembangan zaman dan digitalisasi, sektor properti juga menjadi salah satu industri yang sudah memanfaatkan internet dan membuat platform jual-beli atau sewa-menyewa melalui *website*. Beberapa diantaranya adalah rumah.com, rumah123, lamudi, olx dll. Platform digital jual-beli rumah menjadi salah satu cara untuk meningkatkan industri properti. Platform tersebut dapat memudahkan pembeli properti untuk dapat mengakses dan mencari properti yang mereka idamkan. *Digital marketing* menjadi opsi yang sangat efektif dan efisien pada industri ini dimana sekarang rumah dapat dipasarkan dengan melakukan segmentasi pada dunia digital untuk lebih mengoptimalkan target pasar yang sesuai [6]. Melihat masalah penurunan penjualan dan kenaikan harga sebelumnya, fitur prediksi harga rumah dapat membantu pembeli untuk mendapatkan perkiraan tempat lokasi dan spesifikasi yang sesuai

dengan kemampuan dari pembeli properti. Bagi penjual yang ingin menjual rumah juga dapat terbantu dari fitur prediksi ini yang nantinya bisa digunakan penjual untuk menentukan harga pasar yang sesuai dengan rumah yang ingin mereka jual. Dengan membantunya pembeli dan penjual dalam masalah harga tersebut, fitur prediksi ini dapat membuat penyesuaian harga dan guna untuk meningkatkan kembali industri pasar properti rumah.

Penelitian ini dimaksudkan untuk mendapatkan model prediksi terbaik yang digunakan untuk membantu para pembeli untuk dapat memperkirakan harga pasar rumah dengan spesifikasi dan area rumah yang mereka inginkan. Data didapat dari platform jual-beli rumah lamudi yang meliputi banyak fitur menjadi penentu harga prediksi rumah.

Random Forest merupakan salah satu algoritma yang dapat dipakai untuk melakukan prediksi pada penelitian ini. *Random Forest* sering digunakan dalam melakukan penghitungan valuasi rumah atau *appraisal*. *Random Forest* merupakan teknik *regression tree ensemble* yang dapat membuat prediksi dari rata-rata nilai per *tree*-nya. Keunggulan dari *Random Forest* dimana terdapat sedikit *hyperparameters* dengan potensi yang kuat dalam hasil performanya dan dapat disebut menjadi salah satu teknik terbaik untuk melakukan *appraisal*. Algoritma lainnya yang dipakai pada penelitian ini merupakan *Gradient Boosting* yang merupakan salah satu teknik *boosting* dari *ensemble*. *Gradient Boosting* memiliki digunakan sebagai *learning tools* dari *ensemble* yang dapat digunakan pada klasifikasi maupun regresi. *Gradient Boosting* memiliki tingkat *r-squared* yang tinggi pada hasil masalah prediksi harga rumah. Algoritma terakhir yang dipakai pada penelitian ini merupakan *AdaBoost* yang adalah algoritma *boosting ensemble*. *AdaBoost* merupakan algoritma regresi yang dapat membentuk *classifier* kuat dengan kombinasi *classifier* lemah dan sederhana.

Pada penelitian terdahulu berjudul *A House Price Valuation Based On The Random Forest Approach: The Mass Appraisal Of Residential Property In South Korea* yang ditulis oleh Jengei Hong, Choi Heeyoul , Kim Woo-sung,

mereka melakukan penelitian untuk melakukan *appraisal* dengan menggunakan metode *random forest* pada properti-properti di Korea Selatan, lebih tepatnya di Gangnam, Seoul [7]. Hasil yang didapat pada penelitian tersebut adalah *random Forest* memiliki hasil yang lebih baik dibandingkan dengan prediksi OLS based dimana *random Forest* dapat mendapatkan hasil prediksi pada data yang kompleks dengan *r-squared* sebesar 0.97.

Penelitian lain dengan judul *House Price Prediction Using Regression Techniques: A Comparative Study* yang dikerjakan oleh C.H. Raga Madhuri, Anuradha G, dan M. Vani Pujitha [8]. Pada penelitian ini dilakukan perbandingan beberapa algoritma seperti *Multiple Linear, Ridge, Lasso, Elastic Net, AdaBoost* dan juga *Gradient Boosting Regression*. Penelitian tersebut menghasilkan bahwa *Gradient Boosting* memiliki akurasi yang paling besar yang sangat jauh jika dibandingkan dengan algoritma lainnya, *Gradient Boosting* mendapatkan hasil akurasi sebesar 0.917 dengan RMSE sebesar 10,971,390,390.

Pada penelitian lain dengan judul *House Price Prediction using Random Forest Machine Learning Technique* yang dilakukan oleh Abigail Bola Adetunji dkk [9]. Penelitian ini dilakukan untuk melakukan prediksi harga rumah yang dilakukan dengan algoritma *random Forest*. Penelitian ini menggunakan data dari *dataset Boston housing* yang mengemukakan beberapa hasil prediksi harga dengan membandingkannya dengan harga yang sebenarnya dari data tersebut. Hasil dari penelitian ini adalah algoritma *random Forest* dapat melakukan prediksi harga rumah pada *dataset Boston housing* dimana pada perbandingan harga hanya memiliki nilai *difference* kurang lebih sebesar 5.

Penelitian lain yang membahas tentang prediksi harga rumah merupakan penelitian dengan judul *Housing Prices Prediction with a Deep Learning and RandomForest Ensemble* yang ditulis oleh Bruno Klaus de Aquino Afonso dkk dari Brazil [10]. Penelitian ini menggunakan *dataset* yang cukup besar yaitu sekitar 2 juta data. *Dataset* tersebut meliputi 24 fitur seperti jumlah lantai,

jumlah ruangan, tanggal bangun, *longitude*, *latitude*, lokasi, harga dll. Penelitian ini menghasilkan bahwa *random Forest* dengan teknik *Ensemble* dapat bekerja dengan baik dimana menghasilkan RMSLE sebesar 0.23847 dibandingkan dengan teknik lainnya.

Penelitian terdahulu lainnya merupakan penelitian yang membahas tentang perbandingan algoritma *linear regression* dan *random Forest* pada harga rumah dengan judul *A new machine learning approach to house price estimation* oleh Changchun Wang dan Hui Wu dari Amerika Serikat [11]. *Dataset* yang digunakan dari penelitian ini adalah data harga rumah *single-family* di Arlington, Virginia, Amerika Serikat pada 2015 yang berjumlah sekitar 27649 data dengan fitur luas tanah dan tahun bangun dari rumah tersebut. Hasil yang didapat dari penelitian ini adalah algoritma *random Forest* jauh lebih efektif dan memiliki akurasi lebih besar jika dibandingkan dengan algoritma *linear regression* dimana pada *random Forest* memiliki akurasi sebesar 0.7 dan pada *linear regression* sekitar 0.4 pada model yang berisikan fitur *latitude-longitude*, tahun bangun dan juga luas tanah. Penelitian lainnya dalam penggunaan *cross validation* oleh Rasi Rahwali dkk mendapatkan tingkat *error* sebesar 16.5% dan juga menggunakan *cross-validation* pada penelitiannya [12]. Penelitian terakhir diambil dari penelitian Predicting the Case of COVID-19 in Indonesia using Neural Prophet Model oleh Efraim Yahya dan Alethea Suryadibrata dalam memprediksi kasus COVID 19 yang mendapatkan hasil RMSE sebesar 5728 [13].

Penelitian ini menggunakan data *listing* dari platform lamudi fitur seperti luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, jumlah *carport*, keamanan, jumlah lantai, halaman belakang, taman, balkon, harga, lokasi. Pemilihan algoritma Random Forest dipilih karena Random Forest sudah dipakai pada banyak studi terdahulu dan terbukti dapat melakukan prediksi harga rumah dengan baik. Pada algoritma Gradient Boosting dan AdaBoost juga dipilih melihat pada penelitian terdahulunya dimana kedua model dibuktikan memiliki hasil akurasi yang paling baik diantara algoritma

lainnya. Penelitian ini membandingkan algoritma yang *Random Forest*, *Gradient Boosting*, dan *AdaBoost* untuk mendapatkan hasil performa dari ketiga algoritma tersebut. Setelah itu dilakukan pembuatan *website* untuk melakukan prediksi sesuai dengan parameter yang *user* inginkan.

1.2 Rumusan Masalah

Berdasar dari latar belakang yang sudah dituliskan diatas, berikut merupakan rumusan masalah dari penelitian ini:

1. Bagaimana hasil performa prediksi model yang didapatkan dari proses algoritma *Random Forest*, *Gradient Boosting*, dan *AdaBoost* terhadap data rumah daerah Tangerang Selatan ?
2. Apa faktor yang paling mempengaruhi prediksi harga rumah di Tangerang Selatan?
3. Bagaimana mengimplementasikan model prediksi algoritma dalam bentuk prototipe dengan *framework flask*?

1.3 Batasan Masalah

Batasan masalah yang terdapat pada penelitian ini sebagai berikut:

1. Implementasi estimasi berbasis *website* dengan *framework Flask*.
2. Data properti yang diambil untuk estimasi adalah data properti jenis rumah yang berlokasi di daerah Tangerang dan Tangerang Selatan.
3. Data properti yang diambil didapat dari platform lamudi
4. Indikator yang dipakai dalam penelitian ini adalah data luas tanah, luas bangunan, jumlah *car spaces*, jumlah kamar tidur, jumlah kamar mandi, area, *multiple_floor*, *issecurity*, *isgarden*, *iscourtyard*, dan *isbalcony*.
5. Sistem akan menampilkan hasil berupa harga rumah yang merupakan hasil estimasi dari parameter yang dimasukkan oleh *user*.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Berikut adalah tujuan penelitian dari penelitian ini:

1. Memberikan hasil informasi performa prediksi algoritma *Random Forest*, *Gradient Boosting*, dan *AdaBoost* pada data rumah di Tangerang Selatan.
2. Menemukan faktor paling mempengaruhi estimasi harga rumah di Tangerang Selatan
3. Memberikan estimasi harga rumah sesuai *input* dari *user* yang ada pada *web app* dengan *output* harga rumah.

1.4.2 Manfaat Penelitian

Berikut adalah manfaat penelitian yang dapat diambil dari penelitian ini:

1. Memberi wawasan terhadap performa algoritma Random Forest, Gradient Boosting, AdaBoost pada dataset rumah.
2. Menjadi inspirasi untuk melakukan penelitian kedepannya untuk bantu meningkatkan properti Indonesia
3. Memberikan fitur prediksi harga rumah sesuai *input* dari *user* yang ada pada prototipe web yang dapat dikembangkan lagi kedepannya.

1.5 Sistematika Penulisan

BAB 1 PENDAHULUAN

Bab ini memberikan penjelasan fenomena dari permasalahan penelitian, menuliskan penelitian-penelitian terkait, menuliskan tujuan penelitian, manfaat penelitian, rumusan masalah dan batasan masalah

BAB II LANDASAN TEORI

Bab ini memberikan penjelasan terhadap dasar-dasar teori yang dipakai pada penelitian ini yang diantaranya adalah rumah, lamudi, *Random Forest*, *Gradient Boosting*, *AdaBoost*, *CRISP-DM*, *flask*, *python*, *cross-validation*, *jupyter*, *r-squared*, *RMSE*, dan Penelitian Terdahulu.

BAB III METODOLOGI

Penjelasan terhadap metode yang digunakan pada penelitian serta gambaran umum proses penelitian yang dilakukan.

BAB IV ANALISIS DAN HASIL PENELITIAN

Penjabaran hasil-hasil penelitian yang sudah dikerjakan dengan proses yang sudah dijelaskan pada BAB III. Bab ini akan membahas alur penelitian secara dalam dari proses pengambilan data sampai prediksi harga rumah menggunakan ketiga algoritma pada penelitian ini.

BAB V SIMPULAN DAN SARAN

Menyimpulkan hasil dari penelitian dan juga memberikan saran dan juga informasi terhadap penelitian kedepannya

