

🏠

HOME

(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/)

💰

PEMBAYARAN

(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/PEMBAYARAN)

👤

BIODATA DIRI

(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA)

📖

PEMINJAMAN BUKU

(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/PEMINJAMAN)

📄

PENULISAN ILMIAH

(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/PENULISAN)

📎

FOTOCOPY

(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/FOTOCOPY)

📚

SUMBANG BUKU

(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/SUMBANGBUKU)

📬

KASUS & NOTIFIKASI

(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/NOTIFIKASI)

👥

BEBAS PERPUSTAKAAN

(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/BEBASPERPUSTAKAAN)

📍

KUNJUNGAN

(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/KUNJUNGAN)

📄

E-PAPER

(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/E-PAPER)

FILKOM/KA/SKRIPSI/5209/2022

IMPLEMENTASI MACHINE LEARNING MENGGUNAKAN ALGORITMA RANDOM FOREST REGRESSION UNTUK PREDIKSI HARGA JUAL RUMAH

EGA SRI LESTARI | 12118163 | 2022 | S1 - SKRIPSI | Sistem Informasi - Ilmu Komputer dan Teknologi Informasi

Pembimbing: Ida Astuti | Promotor:

Table Of Contents

ABSTRAKSI (BAHASA INGGRIS)	KATA PENGANTAR	DAFTAR ISI	DAFTAR TABEL	DAFTAR GAMBAR	DAFTAR LAMPIRAN	BAB I	BAB II	BAB III	BAB IV	BAB V	DAFTAR P
----------------------------	----------------	------------	--------------	---------------	-----------------	-------	--------	---------	--------	-------	----------

## 1. PENDAHULUAN

### 1.1 Latar Belakang

Secara geografis Provinsi Jawa Barat ini memiliki luas daerah sekitar 35.378 km<sup>2</sup>. Provinsi Jawa Barat merupakan provinsi yang memiliki jumlah penduduk terbanyak dibandingkan dengan provinsi lain di Indonesia dan juga memiliki tingkat kepadatan penduduk yang cukup tinggi berdasarkan data yang ada. Provinsi Jawa Barat ini terdiri dari 26 kabupaten/kota, sedangkan jumlah kecamatan ada sebanyak 625, dan ada 5.899 desa/kelurahan (SIAK, 2017). Cukup banyak daerah kota besar di Provinsi Jawa Barat yang memiliki kawasan strategis dan juga merupakan kawasan untuk dijadikan pusat pemukiman, pariwisata, dan perkebunan.

Tabel 1. 1 Pertumbuhan Penduduk Pulau Jawa Tahun 2016-2020 (BPS, 2022)

Pulau Jawa	2016	2017	2018	2019	2020
DKI Jakarta	10.265,3	10.348,3	10.428,0	10.504,1	10.576,4
Jawa Barat	47.365,8	47.922,8	48.475,5	49.023,2	49.565,2
Jawa Tengah	33.946,4	34.156,4	34.358,5	34.552,5	34.738,2
DI Yogyakarta	3.718,5	3.768,2	3.818,3	3.868,6	3.919,2
Jawa Timur	39.041,4	39.287,3	39.521,9	39.744,8	39.955,9
Banten	12.157,2	12.345,0	12.530,8	12.714,3	12.895,3

Berdasarkan data tahun 2016 sampai 2020 pada Tabel 1.1 di atas dapat dilihat bahwa provinsi Jawa Barat ini merupakan provinsi yang mendapatkan jumlah pertumbuhan penduduk paling tinggi di Pulau Jawa. Pertumbuhan penduduk paling padat terjadi di provinsi Jawa Barat yang memberikan suatu akibat

masyarakat harus memiliki tempat tinggal yang layak dengan harga yang terjangkau sesuai dengan kebutuhannya masing-masing.

Rumah menjadi salah satu komponen penting bagi manusia untuk menjalani hidupnya. Tempat tinggal atau yang biasa disebut dengan rumah merupakan salah satu dari banyaknya kebutuhan primer bagi setiap orang khususnya keluarga dalam mendapatkan sebuah tempat tinggal pribadi demi melindungi penghuni rumah dari berbagai kondisi alam yang ada di sekitarnya seperti hujan, panas terik matahari, angin kencang, dan sebagainya (Nur Ika Setyani, 2021). Seiring dengan berkembangnya waktu, rumah menjadi salah satu investasi yang menarik untuk dimiliki setiap orang dikarenakan pergerakan harganya yang berubah sewaktu-waktu dan semakin banyak orang yang membutuhkan rumah.

Mengetahui harga rumah berdasarkan kriteria yang diinginkan sangat bermanfaat untuk masyarakat dalam menyesuaikan *budget* keuangan yang dimiliki. Selain mengetahui prediksi harga dari rumah tersebut masyarakat juga membutuhkan sebuah referensi rumah yang berkaitan dengan hasil prediksi yang sudah ditentukan. Aplikasi prediksi dengan memberikan rekomendasi sangat cocok digunakan pada masyarakat untuk mendapatkan rumah yang diinginkan dengan cepat dan mudah. Saat ini pengiklanan rumah yang disediakan tidak hanya berupa brosur, dari mulut ke mulut lain atau pamflet di pinggir jalan karena teknik ini masih sulit untuk menyebarkan informasi yang akan disampaikan secara luas kepada masyarakat. Teknologi saat ini sangat membantu untuk memasarkan penjualan yang dilakukan terhadap masyarakat yang membutuhkan. Hal ini sangat berguna karena masyarakat dapat memperoleh informasi penjualan rumah dimanapun dan kapanpun saat masih menggunakan koneksi internet sehingga dapat mempercepat transaksi jual beli yang akan terjadi. Biasanya pengiklanan untuk penjualan rumah tersebut ada pada beberapa *platform* seperti olx.co.id, rumah123.com, dan sebagainya. Sudah banyak situs online yang bergerak di bidang pelayanan barang dan jasa di dalam *platform* tersebut terdapat beberapa informasi yang sudah lengkap di dalamnya seperti informasi luas tanah, luas bangunan, jumlah kamar tidur, dan sebagainya. Informasi seperti ini yang dapat memudahkan para pembeli rumah

untuk menentukan kriteria dari rumah yang di inginkan dan dapat disesuaikan dengan kondisi keuangan.

Pada *website* rumah123.com yang akan menjadi objek dari pengumpulan data pada penelitian ini memiliki kelebihan yaitu pada *website* ini memiliki semua atribut data yang dibutuhkan untuk penelitian dibandingkan dengan *website* pelayanan jual beli rumah lainnya dan pada *website* rumah123.com ini termasuk *website* yang menyediakan informasi *user agent* pada bagian *inspect* sehingga dapat mempermudah dalam proses pengumpulan data. Tidak semua *website* penjualan rumah yang ada pada internet menyediakan *user agent* untuk dapat dilakukan pengambilan data penelitian dengan struktur HTML. Pada *website* rumah123.com juga memiliki kekurangan pada sistem *filtering* yang ada yaitu masih terdapat beberapa kesalahan *filtering* yang dilakukan dengan inputan yang diberikan oleh pengguna *website* tersebut.

Memprediksi harga rumah diperlukan sebuah algoritma *Machine Learning* yang dapat menyelesaikan masalah yang dibutuhkan. Secara sederhana *Machine Learning* ini adalah salah satu cabang ilmu kecerdasan buatan, khususnya dapat mempelajari tentang bagaimana komputer mampu belajar dari data untuk meningkatkan kecerdasannya. Tujuannya untuk mencari pola tertentu dari setiap sekumpulan data yang menentukan kategori khusus dari masing-masing data agar dapat menyimpulkan sebuah aturan yang digunakan untuk melakukan identifikasi, pengelompokan, dan prediksi bagi data dengan model yang digunakan. Menggunakan *Machine Learning* dapat memudahkan para analisis data dalam mengolah datanya (Victor Marudut Mulia Siregar, 2017). Penelitian ini akan difokuskan untuk menggunakan algoritma *Random Forest Regression* yang dapat memprediksi suatu nilai kontinyu menjadi sebuah keputusan. Algoritma *Random Forest* ini merupakan algoritma yang berbasis pohon dengan menggunakan fitur kualitas dari beberapa pohon keputusan untuk membuat keputusan (Nugraha Listiana dan Achmad Udin, 2020). Algoritma *Random Forest Regression* ini dapat mengatasi suatu kelemahan yang dimiliki oleh *Decision Tree Regression* yaitu sering terjadi *over-fitting* saat pemodelan dengan data yang disediakan. *Random*

*Forest* ini adalah sebuah metode *ensemble* atau gabungan dari banyak model CART sehingga dapat meningkatkan akurasi hasil prediksi nya.

Beberapa penelitian sudah pernah dilakukan sebelumnya sehingga dapat dijadikan sebagai referensi pada penelitian ini. Pada tahun 2020, telah dilakukan sebuah penelitian penggunaan algoritma *Random Forest Regression* untuk memprediksi harga sewa apartment yang memiliki tingkat akurasi sebesar 92% dengan peningkatan menggunakan *tuning parameter* sebesar 0.89% (Indira Luthfiana Mulyahati, 2020). Pada tahun yang sama terdapat penelitian dalam penggunaan metode *Moving Average* untuk memprediksi harga rumah di kota Bandung. Akurasi yang di hasilkan untuk pengujian ini 75% dengan nilai *error MSE* 15% (Arif Fadilah, 2020). Pada tahun 2021, telah dilakukan penelitian dengan menggunakan algoritma *Linear Regression* untuk memprediksi harga rumah di daerah Tebet. Hasil akurasi yang didapatkan sebesar 72% tanpa melakukan *tuning parameter* untuk meningkatkan akurasi (Nur Ika Setyani, 2021).

Memberikan sistem rekomendasi kepada pengguna agar memudahkan dalam melakukan pencarian referensi rumah. Algoritma *Cosine Similarity* akan mengukur kemiripan antara data baru dengan data yang sudah ada pada sistem. Pada tahun 2020, Penelitian dalam menggunakan *Cosine Similarity* dalam mencari keyword pada sesuatu yang hilang menggunakan metode TF-IDF menghasilkan akurasi sebesar 88% (Luluk Suryani dan Kasmi Edy, 2020). Kelebihan penggunaan algoritma ini dalam melakukan rekomendasi adalah tidak ada pengaruh dengan panjang pendeknya suatu data karena mementingkan kemiripan yang dimiliki data tersebut sehingga memiliki tingkat akurasi yang tinggi. Penelitian mengenai penggunaan *Cosine Similarity* juga sudah dilakukan pada tahun 2018, dengan penerapan *Cosine Similarity* untuk memberikan rekomendasi dokumen berdasarkan klasifikasi. Memberikan hasil akurasi pada model klasifikasi sebesar 70%, sedangkan pada model rekomendasi untuk *precision* 75%, *recall* 60% dan *f-measure* 66% (Rut Samuel, et.all, 2018).

Berdasarkan beberapa penelitian yang telah dilakukan sebelumnya, peneliti tertarik untuk melakukan penelitian prediksi harga rumah di provinsi Jawa Barat sebagai provinsi dengan jumlah penduduk paling besar menggunakan algoritma

*Random Forest Regression* serta memberikan rekomendasi referensi rumah menggunakan algoritma *Cosine Similarity*. Alasan menggunakan algoritma *Random Forest Regression* ini dikarenakan pada beberapa penelitian terkait algoritma ini memberikan hasil yang paling bagus untuk mengatasi permasalahan prediksi atau regresi lain.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang sudah dijabarkan, maka rumusan masalah yang berhubungan dengan penelitian ini sebagai berikut:

1. Bagaimana mengimplementasikan algoritma *Random Forest Regression* untuk memprediksi harga rumah dan algoritma *Cosine Similarity* untuk rekomendasi rumah di Jawa Barat?
2. Bagaimana merancang aplikasi prediksi harga rumah berbasis *website* berdasarkan hasil model *Machine Learning* menggunakan *Flask* pada *Python*?
3. Bagaimana melakukan pengujian terhadap model hasil prediksi dan model rekomendasi serta pengujian terhadap *website* prediksi harga rumah?

1.3 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang diambil merupakan data hasil *web scraping* menggunakan *BeautifulShop* terhadap *website* rumah123.com pada tanggal 1 Maret 2022 hingga 5 Maret 2022 hanya wilayah provinsi Jawa Barat sebanyak 1004 data dengan atribut data yaitu nama rumah, kota (kota di Jawa Barat), harga, jumlah kamar mandi (1 - 7), jumlah kamar tidur (1 - 7), jumlah mobil dalam garasi (0 – 5), luas tanah (29 m<sup>2</sup> – 2760 m<sup>2</sup>), luas bangunan (15 m<sup>2</sup> – 714 m<sup>2</sup>) dan link.
2. Proses prediksi menggunakan algoritma *Random Forest Regression* dengan rekomendasi menggunakan algoritma *Cosine Similarity*.
3. Proses implementasi hasil prediksi dan rekomendasi menggunakan *framework Flask* pada *Python* berbasis *website*.

4. Hasil rekomendasi yang diberikan kepada pengguna akan berupa link peralihan ke *website* rumah123.com untuk informasi lebih lengkap.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengimplementasikan algoritma *Random Forest Regression* untuk memprediksi harga rumah dan algoritma *Cosine Similarity* untuk rekomendasi rumah di Jawa Barat.
2. Merancang aplikasi prediksi harga rumah berbasis *website* berdasarkan hasil model *Machine Learning* menggunakan *Flask* pada *Python*.
3. Melakukan pengujian terhadap model hasil prediksi dan model rekomendasi serta pengujian terhadap *website* prediksi harga rumah.
4. Membantu masyarakat dalam menentukan harga rumah berdasarkan fitur yang tersedia dan dapat membantu para marketing perumahan dalam memasarkan rumah dengan prediksi harga sesuai fitur yang diinginkan pembeli.

1.5 Sistematika Penulisan

Adapun untuk sistematika penulisan pada penelitian ini dibagi menjadi lima bagian. Pada bagian pertama yaitu pendahuluan. Pada bagian ini berisi latar belakang dari penelitian yang akan dilakukan, rumusan masalah yang akan dihadapi, batasan masalah pada penelitian, tujuan penelitian, dan sistematika dari penulisan penelitian.

Bagian kedua dari sistematika penulisan ini yaitu tinjauan pustaka. Bagian tinjauan pustaka berisi referensi dan teori yang mendukung penelitian ini diantaranya *Artificial Intelligence*, *Machine Learning (Supervised, Unsupervised, Reinforcement)*, *CART (Classification And Regression Tree)*, Sistem Rekomendasi (*Recommender System*), *Random Forest Regression*, *Cosine Similarity*, *CRISP-DM (Cross-Industry Standard Process for Data Mining)*, *Web Scraping*, *Python*, *Flask*, *HTML*, *CSS*, *Google Colaboratory*, *Visual Studio Code*, *Microsoft Power BI*,

PyCharm, *flowchart*, struktur navigasi, *wireframe*, pengujian inferensial, *user acceptance test*, *website*, heroku, rumah, dan penelitian terkait.

Bagian ketiga dari sistematika penulisan untuk penelitian ini yaitu metode penelitian. Pada bagian ini berisi tahapan perencanaan dan alur dari penelitian yang akan dilakukan. Bagian metode penelitian meliputi *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Deployment*.

Pada bagian keempat dari sistematika penulisan untuk penelitian ini adalah hasil dan pembahasan. Pada bagian ini berisi hasil dari implementasi pada bagian ketiga dalam bentuk program berbasis *website*.

Bagian kelima sebagai bagian terakhir dari sistematika penulisan ini adalah penutup. Pada bagian ini berisi kesimpulan dari penelitian yang telah dilakukan serta saran yang akan ditunjukkan untuk pengembangan penelitian ke depannya.