

HOME  
(HTTP://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/)

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

BIODATA DIRI  
(HTTP://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA)

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

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

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

SUMBANG BUKU  
(HTTP://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/SUMBANGBUKU)

KASUS & NOTIFIKASI  
(HTTP://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/NOTIFIKASI)

BEBAS PERPUSTAKAAN  
(HTTP://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/BEBAAPERPUSTAKAAN)

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

E-PAPER  
(HTTP://LIBRARY.GUNADARMA.AC.ID/DEPC

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
----------------------------	----------------	------------	--------------	---------------	-----------------	-------	--------	---------	--------	-------	----------

## 4. HASIL DAN PEMBAHASAN

### 4.1. Business Understanding

Penyelesaian masalah dalam memprediksi harga rumah dibutuhkan sebuah algoritma regresi yang dapat mengolah nilai kontinyu dari kasus prediksi. Kriteria yang dibutuhkan untuk proses prediksi ini adalah kota di Jawa Barat, nama rumah, kamar tidur, kamar mandi, garasi, luas tanah, luas bangunan, dan harga rumah. Data diperoleh secara eksternal menggunakan teknik web *scraping* dengan bantuan *library beautifulsoup* pada bahasa pemrograman *Python*. Data yang diperlukan berjumlah sekitar 1000 lebih data agar *machine learning* dapat mempelajari pola data sehingga menghasilkan prediksi yang baik. Data yang sudah dikumpulkan akan dilatih dan diuji menggunakan algoritma *Random Forest Regression* untuk memprediksi harga rumah sedangkan untuk hasil rekomendasi rumah akan menggunakan algoritma *Cosine Similarity*. Proses selanjutnya agar hasil prediksi dan rekomendasi dapat berguna di lingkungan masyarakat umum maka akan dilakukan pembuatan aplikasi berbasis *website* dalam pengimplementasian prediksi harga rumah yang akan diteliti. Aplikasi ini diharapkan dapat memudahkan para calon pembeli rumah dalam menemukan rumah impian dan dapat melihat perkiraan harga rumah yang akan dibeli.

### 4.2. Data Understanding

Pada tahap ini akan menghasilkan beberapa proses seperti hasil pengumpulan data, hasil data *validation*, hasil menelaah data, dan melakukan eksplorasi data.

#### 4.2.1. Pengumpulan Data

Proses pengumpulan data ini menggunakan web *scraping* pada *website* rumah123.com untuk mengambil data tersebut sekilas *source code* yang digunakan untuk web *scraping* sebagai berikut:

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

```
for it in items:
    name = it.find('a', 'ui-organisms-card-r123-
featured_middle-section_title').text
    daerah = it.find('p', 'ui-organisms-card-r123-
featured_middle-
section_address').text.strip().split(', ')[1]
    harga = ' '.join(it.find('div', 'ui-organisms-
card-r123-featured_middle-
section_harga').text.strip().replace('Rp ', ''),
'').replace('JutaCicilan',
'Juta').replace('MiliarCicilan',
'Miliar').split()[0:2])
```

Berdasarkan *source code* di atas setelah *user agent* mendapatkan respon 200 maka *website* tersebut sudah dapat diakses untuk pengambilan data. Pada *code* di atas pertama-tama temukan semua elemen html lalu akses *class* pada elemen tersebut sehingga berisi keperluan data yang dibutuhkan kemudian ambil teks pada elemen tersebut dengan fungsi *.text()*. Lalu ganti setiap teks yang tidak diperlukan sesuai dengan keinginan misalnya disini peneliti mengganti karakter JutaCicilan menjadi Juta.

Tabel 4. 1 Hasil Pengumpulan Data

name	province	price	bedroom	bathroom	parking	surfaceare	buildingarlink
Bonus Fur Depok	791 Juta	3	2	1	72	89	<a href="https://www.rumah123.com/properti/depok/hos8408068/">https://www.rumah123.com/properti/depok/hos8408068/</a>
Rumah 2 Lt Depok	793 Juta	3	2	1	72	88	<a href="https://www.rumah123.com/properti/depok/hos888647/">https://www.rumah123.com/properti/depok/hos888647/</a>
Rumah 2 Lt Depok	790 Juta	3	2	1	72	90	<a href="https://www.rumah123.com/properti/depok/hos9375718/">https://www.rumah123.com/properti/depok/hos9375718/</a>
Rumah Mi Depok	870 Juta	3	2	1	72	61	<a href="https://www.rumah123.com/properti/depok/hos9495151/">https://www.rumah123.com/properti/depok/hos9495151/</a>
Rumah De Depok	690 Juta	3	2	1	64	65	<a href="https://www.rumah123.com/properti/depok/hos6370047/">https://www.rumah123.com/properti/depok/hos6370047/</a>

Berdasarkan Tabel 4.1 di atas terdapat beberapa kolom pada tabel. Keterangan dari masing-masing kolom dijabarkan sebagai berikut:

- Name (*string*) : nama-nama rumah yang ada pada *website* rumah123.com.
- Daerah (*string*) : kota di Provinsi Jawa Barat.
- Harga (*string*) : harga rumah.
- Kamar tidur (*integer*) : jumlah kamar tidur.
- Kamar mandi (*integer*) : jumlah kamar mandi.

-  **HOME**  
([HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/](https://library.gunadarma.ac.id/DEPC SYSTEM/))
  -  **PEMBAYARAN**  
([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](https://library.gunadarma.ac.id/DEPC SYSTEM/ANGGOTA))
  -  **PEMINJAMAN BUKU**  
([HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA/PEMINJAMAN](https://library.gunadarma.ac.id/DEPC SYSTEM/ANGGOTA/PEMINJAMAN))
  -  **PENULISAN ILMIAH**  
([HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA/PENULISAN](https://library.gunadarma.ac.id/DEPC SYSTEM/ANGGOTA/PENULISAN))
  -  **FOTOCOPY**  
([HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA/FOTOCOPY](https://library.gunadarma.ac.id/DEPC SYSTEM/ANGGOTA/FOTOCOPY))
  -  **SUMBANG BUKU**  
([HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA/SUMBANGBUKU](https://library.gunadarma.ac.id/DEPC SYSTEM/ANGGOTA/SUMBANGBUKU))
  -  **KASUS & NOTIFIKASI**  
([HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA/NOTIFIKASI](https://library.gunadarma.ac.id/DEPC SYSTEM/ANGGOTA/NOTIFIKASI))
  -  **BEBAS PERPUSTAKAAN**  
([HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA/BEBASPERPUSTAKAAN](https://library.gunadarma.ac.id/DEPC SYSTEM/ANGGOTA/BEBASPERPUSTAKAAN))
  -  **KUNJUNGAN**  
([HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/KUNJUNGAN](https://library.gunadarma.ac.id/DEPC SYSTEM/KUNJUNGAN))
  -  **E-PAPER**  
([HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/E-PAPER](https://library.gunadarma.ac.id/DEPC SYSTEM/E-PAPER))

- Garasi (*integer*) : jumlah kendaraan yang dapat diisi pada satu garasi.
  - Luas tanah (*integer*) : luas tanah.
  - Luas bangunan (*integer*) : luas bangunan.
  - Link (*string*) : link dari masing-masing profil rumah.

Untuk data kamar mandi dan kamar tidur berisi jumlah dari kamar tidur dan kamar mandi di masing-masing rumah hanya sekitar 1 sampai 7 kamar mandi dan kamar tidur. Data garasi menunjukkan jumlah garasi yaitu sekitar 0 sampai 5 garasi pada masing-masing rumah. Pada data luas tanah menunjukkan luas tanah yaitu hanya 29 m<sup>2</sup> sampai 2760 m<sup>2</sup> sedangkan luas bangunan data ini berkisar antara 15 m<sup>2</sup> sampai 714 m<sup>2</sup>. Pada data harga rumah menggunakan maksimal harga 5 miliar dikarenakan untuk harga rumah melebihi 5 miliar akan menunjukkan sebuah bangunan kos yang dijual pada *website* ini. Data yang didapatkan berjumlah 1004 baris dan 9 kolom yang diambil dengan periode waktu 1 Maret 2021 sampai 5 Maret 2021 pada *website* rumah123.com.

#### 4.2.2. Data Validation

Pada hasil dari proses ini akan berupa hasil dalam pemahaman data secara statistika baik dalam pengubahan data yang tidak relevan maupun melihat statistika deskriptif. Hasil pengumpulan data dapat dilihat bahwa nama pada setiap kolom masih berupa bahasa inggris, maka dari itu diperlukan untuk pengubahan nama kolom dengan *function rename* pada *Python*.

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

Berdasarkan *source code* di atas proses pengubahan nama kolom dilakukan dengan mudah menggunakan *Python*. Proses ini diawali dengan menggunakan *function rename* kemudian tulis nama kolom terdahulu dan tulis nama kolom terbaru yang dipisahkan dengan tanda titik dua (:).

Jika dilihat dari hasil pengumpulan data isi dari *field* harga masih merupakan *string* maka dari itu harus dilakukan pengubahan tipe data dari *field* tersebut menggunakan *source code* berikut:

```
#convert kolom harga
data['harga'] = data['harga'].replace({
    ',' : '.',
    ' Juta' : '*1E6',
    ' Miliar' : '*1E9'
}, regex = True).map(pd.eval).astype(int)
```

Berdasarkan *source code* di atas pertama-tama melakukan pengubahan tipe data pada kolom harga menggunakan *library* *regex* untuk menukar nilai mata uang pada data tersebut lalu mengubah tipe data menjadi *integer*. Untuk pengubahan tipe data pada nilai “juta” maka akan dikalikan dengan \*1E6, sedangkan untuk “miliar” akan dikalikan dengan \*1E9. Kemudian *evaluate* hasil dari perkalian *string* menggunakan fungsi *pd.eval* lalu ubah tipe data dari *string* menjadi *integer*. Jika di perlukan maka lakukan proses untuk mengganti nama kolom untuk memudahkan dalam mengolah data pada proses selanjutnya.

	nama daerah	harga	kamar tidur	kamar mandi	garasi	luas tanah	luas bangunan	link
0	Bonus Furnished - Rumah 2 Lantai Murah Terdekat	Depok 790000000	3	2	1	72	69	<a href="https://www.rumah123.com/properti/depok/hos840">https://www.rumah123.com/properti/depok/hos840</a>
1	Rumah 2 Lantai Murah Nyaman Terdekat Stasiun K.	Depok 700000000	3	2	1	72	68	<a href="https://www.rumah123.com/properti/depok/hos839">https://www.rumah123.com/properti/depok/hos839</a>
2	Rumah 2 Lantai Termurah Terdekat Stasiun Kereta	Depok 750000000	3	2	1	72	69	<a href="https://www.rumah123.com/properti/depok/hos837">https://www.rumah123.com/properti/depok/hos837</a>
3	Rumah Mewah Strategis Pinggir Jalan Bebas Bang	Depok 870000000	3	2	1	72	61	<a href="https://www.rumah123.com/properti/depok/hos848">https://www.rumah123.com/properti/depok/hos848</a>
4	Rumah Dekat Tol Duri	Depok 600000000	3	2	1	64	60	<a href="https://www.rumah123.com/properti/depok/hos831">https://www.rumah123.com/properti/depok/hos831</a>

Gambar 4. 1 Hasil Pengubahan Data

Berdasarkan pada Gambar 4.1 di atas pengubahan data terletak pada data kolom harga menjadi bilangan numerik menggunakan *function regex*. Pengubahan data ini dibutuhkan untuk menghasilkan dataset yang baik sebelum ke tahap selanjutnya.

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

Lakukan penghapusan pada data yang tidak dibutuhkan, pada kasus ini adalah memastikan jumlah kamar mandi dan kamar tidur antara 1 sampai 7 maka dari itu harus di lihat apakah ada data yang memiliki kamar tidur dan kamar mandi lebih dari jumlah ataupun kurang dari jumlah tersebut.

	nama	daerah	harga	kamar tidur	kamarmandi	garasi	luastanah	luastbangunan	link
236	Rumah mewah bisa untuk kost-kostan di daerah setibab	Bandung	2250000000	10	4	0	214	218	<a href="https://www.rumah123.com/properti/bandung/hos1218">https://www.rumah123.com/properti/bandung/hos1218</a>
275	rumah siap huni plus fasilitas di dalamnya	Bandung	2620000000	12	12	2	300	200	<a href="https://www.rumah123.com/properti/bandung/hos1200">https://www.rumah123.com/properti/bandung/hos1200</a>
846	Rumah Mewah, Bers Ruang, Lokasi Strategis d...	Bandung	2000000000	9	3	0	171	200	<a href="https://www.rumah123.com/properti/bandung/hos1200">https://www.rumah123.com/properti/bandung/hos1200</a>
625	Rumah Hook Siap Bangun Coklat Untuk Kehidupan d...	Bekasi	3600000000	6	9	0	396	600	<a href="https://www.rumah123.com/properti/bekasi/hos1600">https://www.rumah123.com/properti/bekasi/hos1600</a>
508	RUMAH ASRI DENGAN HALAMAN 1000M2 BANDUNG	Bandung	2000000000	5	10	0	1200	200	<a href="https://www.rumah123.com/properti/bandung/hos1200">https://www.rumah123.com/properti/bandung/hos1200</a>
962	rumah cantik beradaan resang dengan view indah...	Bandung	2500000000	8	7	0	372	400	<a href="https://www.rumah123.com/properti/bandung/hos1400">https://www.rumah123.com/properti/bandung/hos1400</a>
971	TURUN HARGA DARI 3 MILYAR JADI 2.7 MILYAR LANG...	Bandung	2700000000	8	2	1	176	276	<a href="https://www.rumah123.com/properti/bandung/hos1276">https://www.rumah123.com/properti/bandung/hos1276</a>

Gambar 4. 2 Data Yang Tidak Diperlukan

Berdasarkan Gambar 4.2 di atas dapat dilihat bahwa masih ada beberapa data yang tidak memenuhi kriteria yang dibutuhkan jumlah data yang tidak diperlukan ini adalah 7 data sedangkan untuk data yang memiliki nilai kamar mandi dengan jumlah 0 ada 3 data. Rata-rata data yang ditemukan adalah data berupa bangunan rumah kos-kosan yang dijual pada *website* ini. Lakukan penghapusan pada data tersebut agar dapat mencegah adanya ketidakstabilan hasil prediksi. Setelah dilakukan penghapusan tersebut jumlah data yang memenuhi kriteria adalah 994 data. Langkah selanjutnya pada tahap data validasi ini adalah melihat deskripsi statistika pada data untuk melihat karakteristik data.

#### 4.2.3. Menelaah Data

Pada proses ini akan melihat statistika deskriptif dari dataset, kemudian melihat korelasi antar kolom pada dataset.

	nama	daerah	harga	kamar tidur	kamarmandi	garasi	luastanah	luastbangunan	link
count	994	994	9.540000e+02	994.000000	994.000000	994.000000	994.000000	994.000000	994
unique	965	15	Nan	Nan	Nan	Nan	Nan	Nan	964
top	OCC Tipe Rumah Bandung	Bandung	Nan	Nan	Nan	Nan	Nan	Nan	<a href="https://www.rumah123.com/properti/bogor/hos1025">https://www.rumah123.com/properti/bogor/hos1025</a>
freq	3	427	Nan	Nan	Nan	Nan	Nan	Nan	3
mean	Nan	Nan	1.430202e+09	2.988734	2.135815	0.799959	135.795775	120.344064	Nan
std	Nan	Nan	1.175626e+09	0.962568	1.024702	0.807293	110.846998	91.618462	Nan
min	Nan	Nan	3.500000e+08	1.000000	1.000000	0.000000	29.000000	21.000000	Nan
25%	Nan	Nan	5.300000e+08	2.000000	1.000000	0.000000	70.000000	50.000000	Nan
50%	Nan	Nan	1.000000e+09	3.000000	2.000000	1.000000	102.000000	90.000000	Nan
75%	Nan	Nan	1.977500e+09	3.000000	3.000000	1.000000	156.750000	167.000000	Nan
max	Nan	Nan	5.000000e+09	7.000000	7.000000	5.000000	1000.000000	660.000000	Nan

Gambar 4. 3 Statistika Deskriptif

HOME  
(HTTP://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/)

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

BIODATA DIRI  
(HTTP://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA)

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

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

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

SUMBANG BUKU  
(HTTP://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/SUMBANGBUKU)

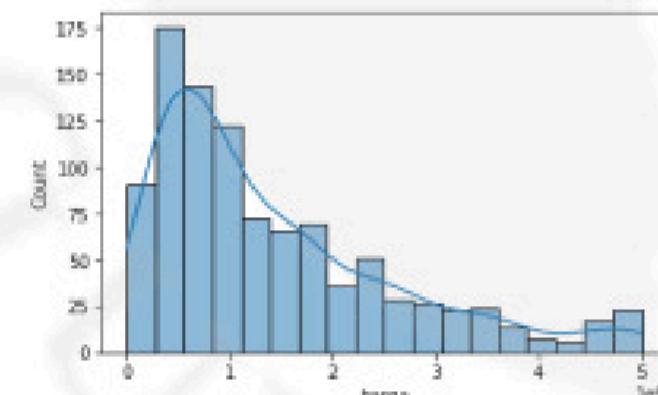
KASUS & NOTIFIKASI  
(HTTP://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/NOTIFIKASI)

BEBAS PERPUSTAKAAN  
(HTTP://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/BEBAZPERPUSTAKAAN)

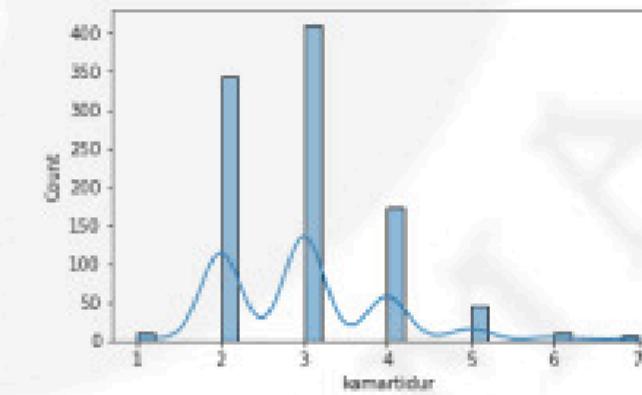
KUNJUNGAN  
(HTTP://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/KUNJUNGAN)

E-PAPER  
(HTTP://LIBRARY.GUNADARMA.AC.ID/DEPC

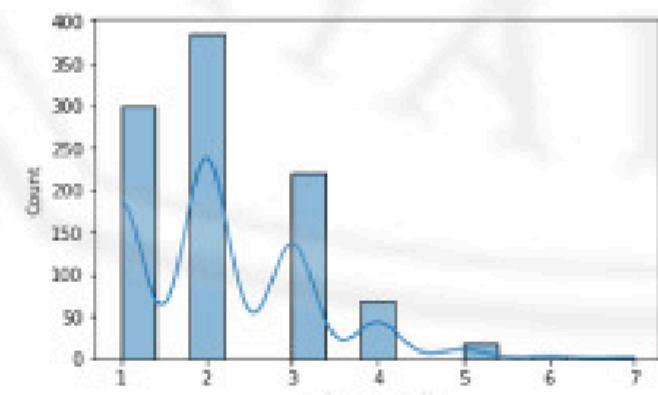
Berdasarkan Gambar 4.3 di atas untuk setiap atribut pada data memiliki ukuran yang sama yaitu 994. Pada kolom kota memiliki 15 kota dengan data terbanyak berada pada kota bandung sebanyak 427 kota bandung mendominasi pada kolom kota. Jika dilihat dari nilai maksimal pada kolom luas tanah, luas bangunan, dan kolom harga memiliki rentang yang cukup jauh dengan nilai minimum yang dimiliki hal ini dapat disimpulkan bahwa terdapat nilai pencilan (*outlier*) pada kolom tersebut. Nilai standar deviasi pada kolom garasi lebih tinggi dari pada nilai rata-rata nya artinya persebaran data yang terjadi pada kolom ini bersifat heterogen atau menyebar secara luas, sedangkan pada kolom lainnya memiliki nilai standar deviasi yang lebih kecil dibandingkan nilai rata-rata maka persebaran data yang terjadi bersifat homogen atau tidak menyebar luas. Apabila nilai standar deviasi dan rata-rata memiliki rentang yang jauh seperti pada kolom luas tanah, luas bangunan dan harga maka pada kolom ini dapat disimpulkan bahwa terjadi *skewness* pada data atau pengumpulan data yang tidak stabil. Statistika deskriptif ini dapat memungkinkan untuk mengambil metode pencegahan seperti apa setelah melihat karakteristik dari masing-masing atribut. Karena jika salah memilih metode pendekatan untuk memaksimalkan *performance* model pada data akan mengakibatkan kurangnya tingkat evaluasi yang dihasilkan.



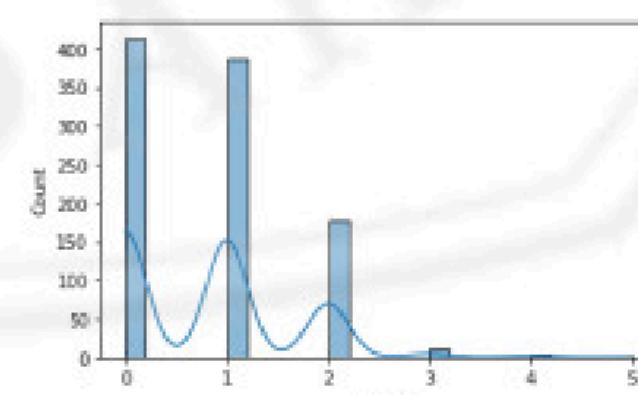
(a) Kolom harga



(b) Kolom kamar tidur



(c) Kolom kamar mandi



(d) Kolom garasi

[HOME  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/\)](https://library.gunadarma.ac.id/DEPC/system/)

[PEMBAYARAN  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/PEMBAYARAN\)](https://library.gunadarma.ac.id/DEPC/system/PEMBAYARAN)

[BIODATA DIRI  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA)

[PEMINJAMAN BUKU  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/PEMINJAMAN\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/PEMINJAMAN)

[PENULISAN ILMIAH  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/PENULISAN\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/PENULISAN)

[FOTOCOPY  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/FOTOCOPY\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/FOTOCOPY)

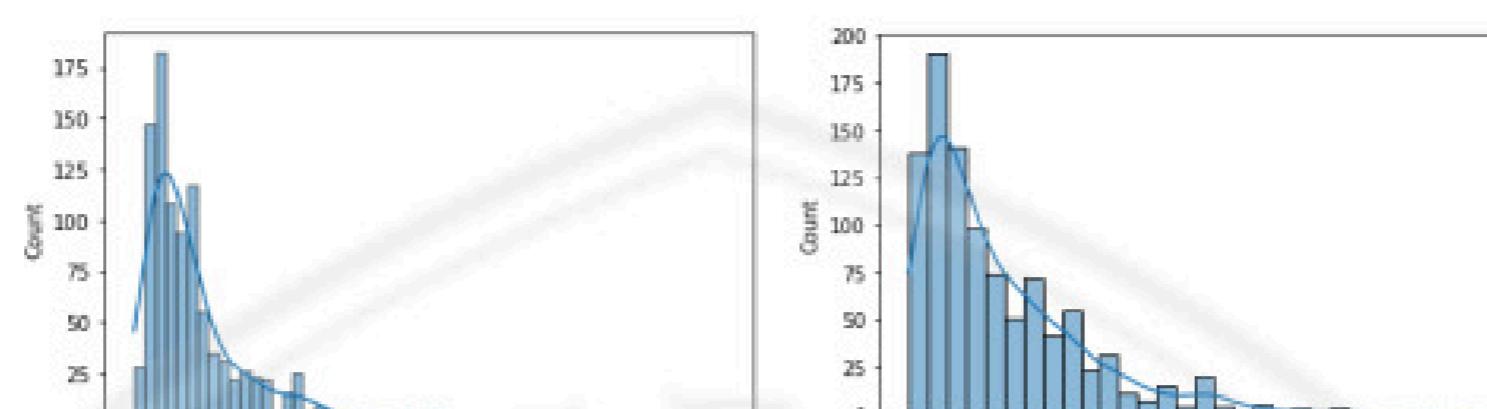
[SUMBANG BUKU  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/SUMBANGBUKU\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/SUMBANGBUKU)

[KASUS & NOTIFIKASI  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/NOTIFIKASI\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/NOTIFIKASI)

[BEBAS PERPUSTAKAAN  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/BEBASPERPUSTAKAAN\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/BEBASPERPUSTAKAAN)

[KUNJUNGAN  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/KUNJUNGAN\)](https://library.gunadarma.ac.id/DEPC/system/KUNJUNGAN)

[E-PAPER  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC](https://library.gunadarma.ac.id/DEPC/system/E-PAPER)



(e) Kolom luas tanah (f) Kolom luas bangunan

Gambar 4. 4 Histogram Masing-Masing Kolom

Berdasarkan Gambar 4.4 di atas untuk kolom garasi, kamar tidur dan kamar mandi memiliki struktur distribusi data yang cukup normal dilihat dari garis yang terbentuk mengalami stabilitas yang cukup baik. Sedangkan untuk kolom harga, luas tanah dan luas bangunan struktur distribusi data tidak normal yang mengalami skew ke kanan karena data yang terkumpul pada suatu titik sesuai dengan hasil standar deviasi pada statistika deskriptif. Setelah mengetahui distribusi data pada masing-masing kolom maka dapat menentukan metode perhitungan korelasi yang akan digunakan. Jika semua atribut pada data memiliki distribusi normal maka disarankan untuk menggunakan metode korelasi *Pearson*.

```
correlation = data.corr(method='spearman')
plt.figure(figsize=(10,10))
plt.title('Correlation of Data House Pricing in Indonesia', fontsize=24, fontweight='bold', pad = 70)
sns.heatmap(correlation, annot=True, cmap='bwr',
            linewidths=1, linecolor='k',
            square=True, cbar_kws={"orientation": "vertical"}, cbar=True, mask=False)
```

Berdasarkan *source code* di atas data yang sudah dilakukan validasi dan telah dilihat karakteristik melalui statistika deskriptif langkah selanjutnya melihat korelasi antar kolom. Melihat korelasi dengan *Python* dapat menggunakan *function corr()* dan untuk metode *default* yang diset pada *Python* adalah korelasi Pearson. Maka dari itu harus diubah metode nya menjadi *Spearman* dengan menggunakan parameter *method*. Hasil yang akan diberikan pada *function corr()* berupa tabel

[HOME  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/\)](https://library.gunadarma.ac.id/DEPC/system/)

[PEMBAYARAN  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/PEMBAYARAN\)](https://library.gunadarma.ac.id/DEPC<br/>SYSTEM/PEMBAYARAN)

[BIODATA DIRI  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA\)](https://library.gunadarma.ac.id/DEPC<br/>SYSTEM/ANGGOTA)

[PEMINJAMAN BUKU  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/PEMINJAMAN\)](https://library.gunadarma.ac.id/DEPC<br/>SYSTEM/ANGGOTA/PEMINJAMAN)

[PENULISAN ILMIAH  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/PENULISAN\)](https://library.gunadarma.ac.id/DEPC<br/>SYSTEM/ANGGOTA/PENULISAN)

[FOTOCOPY  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/FOTOCOPY\)](https://library.gunadarma.ac.id/DEPC<br/>SYSTEM/ANGGOTA/FOTOCOPY)

[SUMBANG BUKU  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/SUMBANGBUKU\)](https://library.gunadarma.ac.id/DEPC<br/>SYSTEM/ANGGOTA/SUMBANGBUKU)

[KASUS & NOTIFIKASI  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/NOTIFIKASI\)](https://library.gunadarma.ac.id/DEPC<br/>SYSTEM/ANGGOTA/NOTIFIKASI)

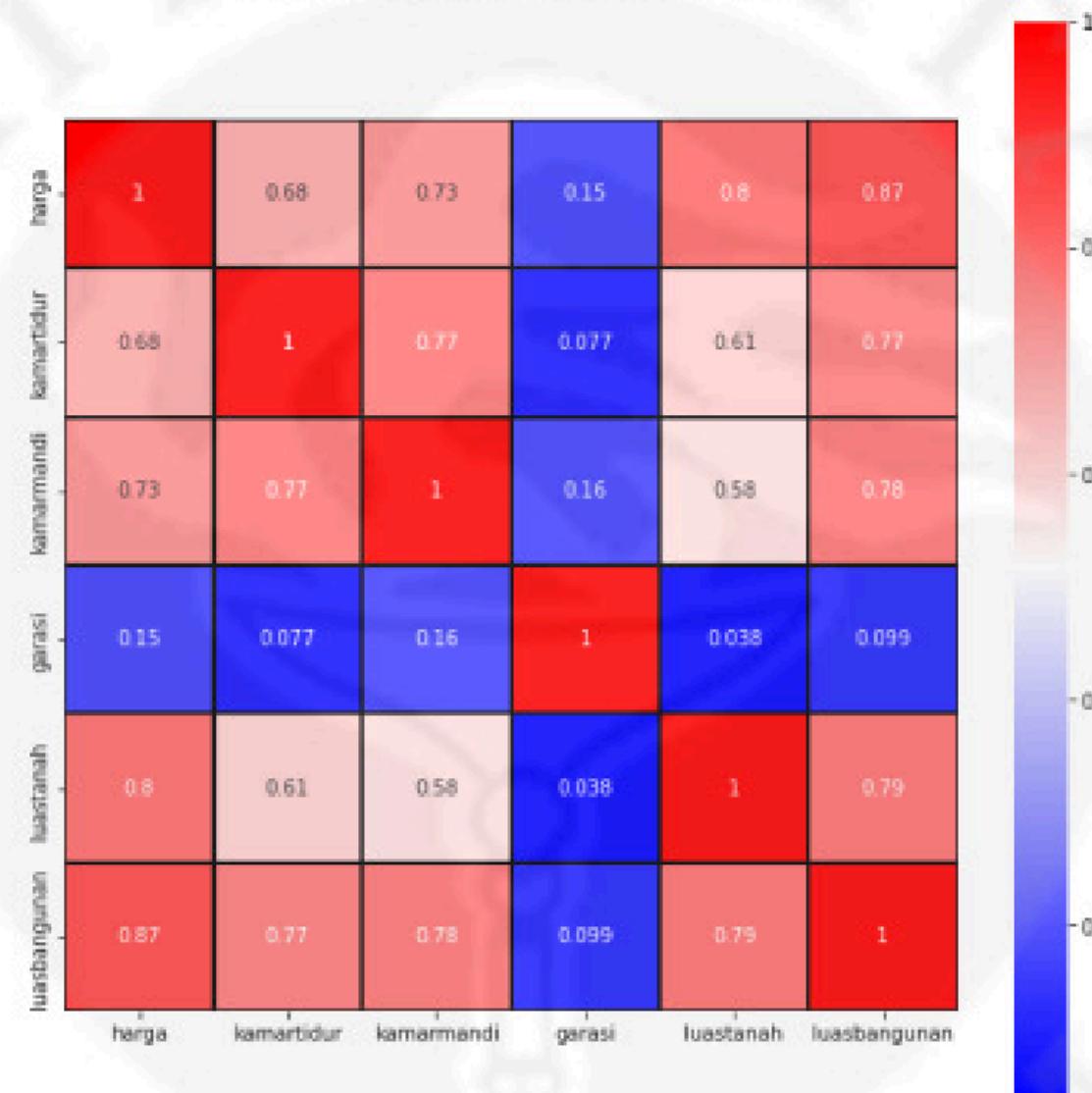
[BEBAS PERPUSTAKAAN  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/BEBASPERPUSTAKAAN\)](https://library.gunadarma.ac.id/DEPC<br/>SYSTEM/ANGGOTA/BEBASPERPUSTAKAAN)

[KUNJUNGAN  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/KUNJUNGAN\)](https://library.gunadarma.ac.id/DEPC<br/>SYSTEM/KUNJUNGAN)

[E-PAPER  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC\)](https://library.gunadarma.ac.id/DEPC)

array 2 dimensi, maka dari itu untuk memudahkan pembacaan korelasi tersebut digunakan visualisasi *heat map* yang akan menampilkan informasi semakin merah kolom tersebut maka semakin besar korelasi yang dihasilkan. Gunakan parameter *annot = ‘true’* karena *default* yang akan diberikan adalah *‘false’*, *annot* ini digunakan untuk memperlihatkan angka korelasi yang menjadi label dari hasil *function corr*.

### Korelasi Data Prediksi Harga Rumah Provinsi Jawa Barat



Gambar 4. 5 Hasil Korelasi *Spearman* pada Dataset

Berdasarkan Gambar 4.5 di atas terlihat bahwa harga rumah memiliki korelasi paling tinggi dengan luas bangunan yaitu 0.87 yang berarti bahwa jika harga rumah mengalami peningkatan maka hal tersebut sangat dipengaruhi oleh luas bangunan. Hubungan korelasi harga dengan jumlah kamar mandi, jumlah kamar tidur dan luas tanah memiliki korelasi yang sedang yang berarti ketiga atribut tersebut memiliki pengaruh tidak sebesar luas bangunan. Pada hubungan korelasi antara harga rumah dengan jumlah garasi adalah korelasi yang kecil akan tetapi

[HOME  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/\)](https://library.gunadarma.ac.id/DEPC/system/)

[PEMBAYARAN  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/PEMBAYARAN\)](https://library.gunadarma.ac.id/DEPC/system/PEMBAYARAN)

[BIODATA DIRI  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA)

[PEMINJAMAN BUKU  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/PEMINJAMAN\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/PEMINJAMAN)

[PENULISAN ILMIAH  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/PENULISAN\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/PENULISAN)

[FOTOCOPY  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/FOTOCOPY\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/FOTOCOPY)

[SUMBANG BUKU  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/SUMBANGBUKU\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/SUMBANGBUKU)

[KASUS & NOTIFIKASI  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/NOTIFIKASI\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/NOTIFIKASI)

[BEBAS PERPUSTAKAAN  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/BEBASPERPUSTAKAAN\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/BEBASPERPUSTAKAAN)

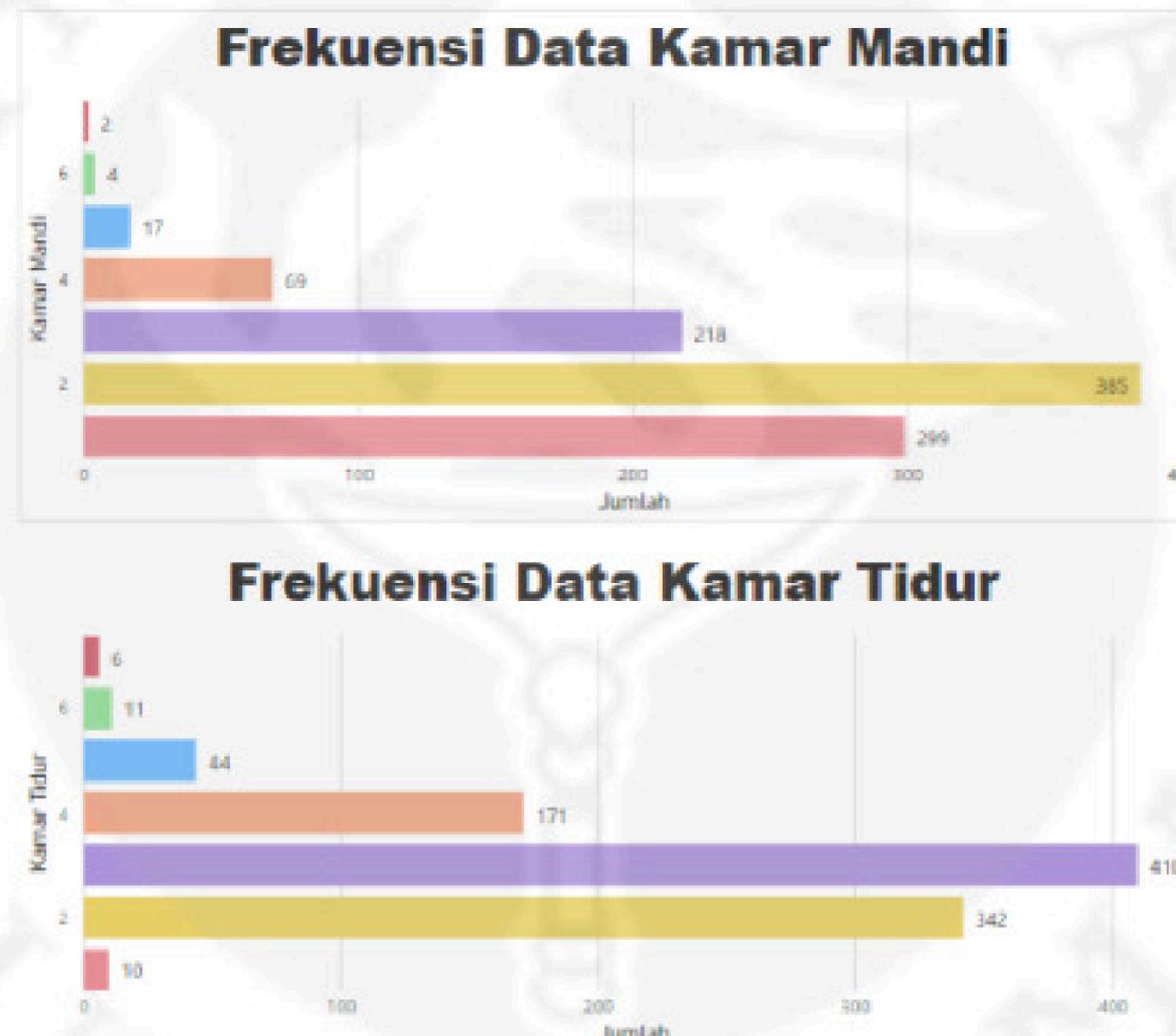
[KUNJUNGAN  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/KUNJUNGAN\)](https://library.gunadarma.ac.id/DEPC/system/KUNJUNGAN)

[E-PAPER  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC\)](https://library.gunadarma.ac.id/DEPC)

memiliki nilai yang positif berarti harga rumah sulit untuk di pengaruhi oleh jumlah garasi dari rumah tersebut. Korelasi dari semua atribut pada data yang akan digunakan memiliki korelasi yang positif sehingga setiap peningkatan yang terjadi antar atribut dapat mempengaruhi nilai atribut lain sesuai dengan nilai korelasi yang didapatkan.

#### 4.2.4. Eksplorasi Data dengan Visualisasi

Pada hasil dari proses ini akan menggunakan perangkat lunak khusus untuk mempermudah visualisasi data yaitu Microsoft Power BI.



Gambar 4. 6 Visualisasi 1

Pada Gambar 4.6 di atas dapat dilihat bahwa frekuensi data rumah memiliki 2 kamar mandi dan 3 kamar tidur dengan hasil masing-masing 385 kamar mandi dan 410 kamar tidur. Kedua fasilitas ini data dengan rumah yang memiliki 7 kamar mandi dan kamar tidur adalah data yang paling sedikit. Trend yang diberikan pada kedua frekuensi fasilitas di atas dapat disimpulkan bahwa penduduk jawa barat lebih dominan tertarik untuk memiliki rumah dengan jumlah kamar mandi yang relatif sedikit antara 1 sampai 4 sedangkan untuk jumlah kamar tidur antara 2

[HOME  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/\)](https://library.gunadarma.ac.id/DEPC/system/)

[PEMBAYARAN  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/PEMBAYARAN\)](https://library.gunadarma.ac.id/DEPC/system/pembayaran)

[BIODATA DIRI  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA\)](https://library.gunadarma.ac.id/DEPC/system/anggota/biodata)

[PEMINJAMAN BUKU  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/PEMINJAMAN\)](https://library.gunadarma.ac.id/DEPC/system/anggota/peminjaman)

[PENULISAN ILMIAH  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/PENULISAN\)](https://library.gunadarma.ac.id/DEPC/system/anggota/penulisan)

[FOTOCOPY  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/FOTOCOPY\)](https://library.gunadarma.ac.id/DEPC/system/anggota/fotocopy)

[SUMBANG BUKU  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/SUMBANGBUKU\)](https://library.gunadarma.ac.id/DEPC/system/anggota/sumbangbuku)

[KASUS & NOTIFIKASI  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/NOTIFIKASI\)](https://library.gunadarma.ac.id/DEPC/system/anggota/notifikasi)

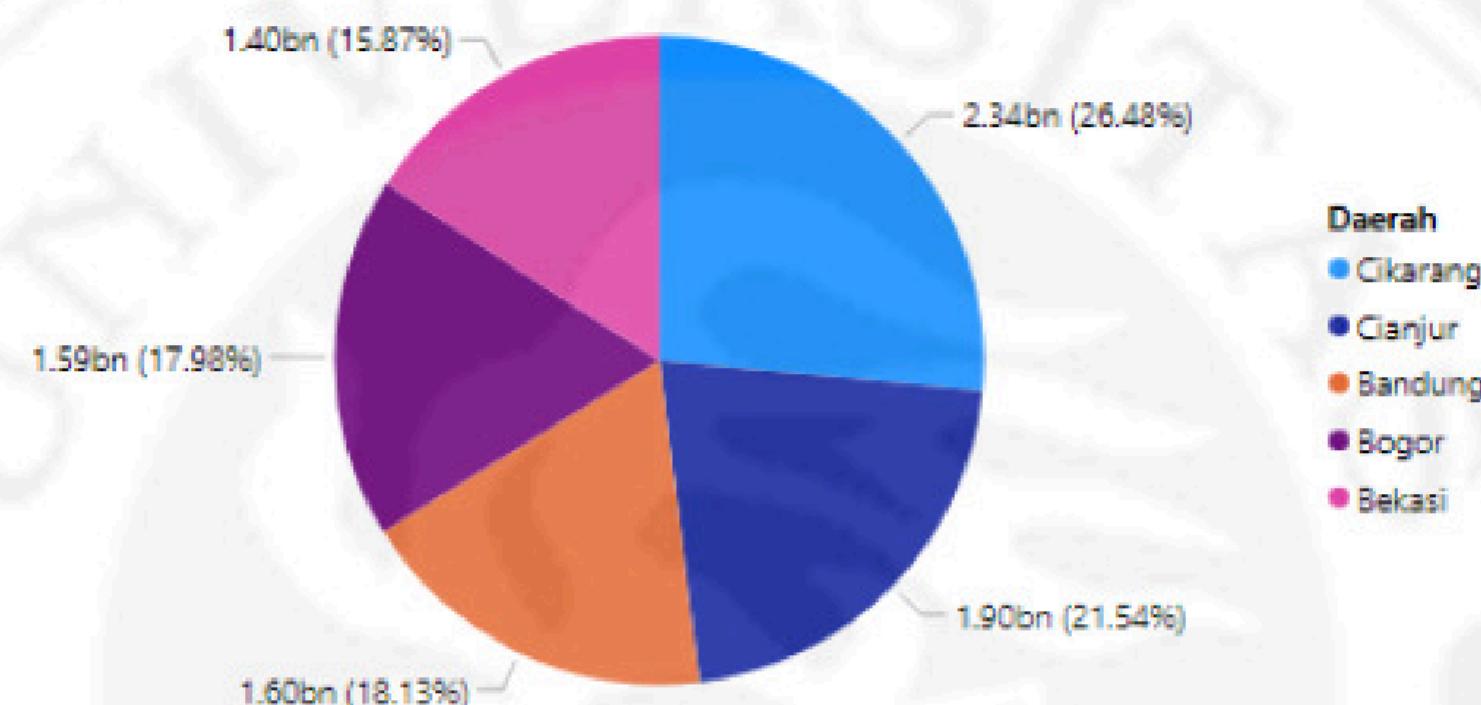
[BEBAS PERPUSTAKAAN  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/BEBASPERPUSTAKAAN\)](https://library.gunadarma.ac.id/DEPC/system/anggota/bebasperpustakaan)

[KUNJUNGAN  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/KUNJUNGAN\)](https://library.gunadarma.ac.id/DEPC/system/kunjungan)

[E-PAPER  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC](https://library.gunadarma.ac.id/DEPC/e-paper)

sampai 4 kamar tidur. Kemungkinan bahwa pemilik rumah akan menyesuaikan kebutuhan dari jumlah kamar tidur dan kamar mandi yang akan dibangun pada rumahnya sehingga hal ini dapat disesuaikan dengan luas tanah atau luas bangunan maupun keinginan pemilik rumah.

## 5 Daerah Rata-Rata Harga Rumah Tertinggi



Gambar 4. 7 Visualisasi 2

Berdasarkan Gambar 4.7 di atas menjelaskan bahwa terdapat 5 wilayah teratas yang memiliki jumlah rumah dijual paling banyak. Pada visualisasi tersebut wilayah Cikarang yang merupakan kota dengan jumlah harga rumah dijual paling besar dengan mendapatkan rata-rata harga rumah yang paling tinggi diantara 4 wilayah lain. Dilihat bahwa rata-rata dari masing-masing kota tersebut memiliki rentang yang cukup kecil hanya sekitar 1%-4% perbedaan. Meskipun pada data daerah paling banyak adalah wilayah Bandung bukan berarti pada wilayah ini akan menjadi harga rumah paling tinggi, bisa dikatakan apabila harga rumah yang dijual pada wilayah Cikarang mahal dengan jumlah yang sedikit maka akan mengakibatkan nilai harga rumah di Cikarang akan mahal, sehingga apabila harga rumah di wilayah Bandung mahal akan tetapi dengan jumlah yang banyak maka rata-rata dari seluruh harga rumah akan lebih murah dibandingkan harga di Cikarang. Sehingga dapat disimpulkan bahwa harga rumah pada setiap daerah di provinsi Jawa Barat cukup berpengaruh besar terhadap wilayah yang dipilih karena

[HOME  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/\)](https://library.gunadarma.ac.id/DEPC/system/)

[PEMBAYARAN  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/PEMBAYARAN\)](https://library.gunadarma.ac.id/DEPC/system/PEMBAYARAN)

[BIODATA DIRI  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA)

[PEMINJAMAN BUKU  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/PEMINJAMAN\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/PEMINJAMAN)

[PENULISAN ILMIAH  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/PENULISAN\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/PENULISAN)

[FOTOCOPY  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/FOTOCOPY\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/FOTOCOPY)

[SUMBANG BUKU  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/SUMBANGBUKU\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/SUMBANGBUKU)

[KASUS & NOTIFIKASI  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/NOTIFIKASI\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/NOTIFIKASI)

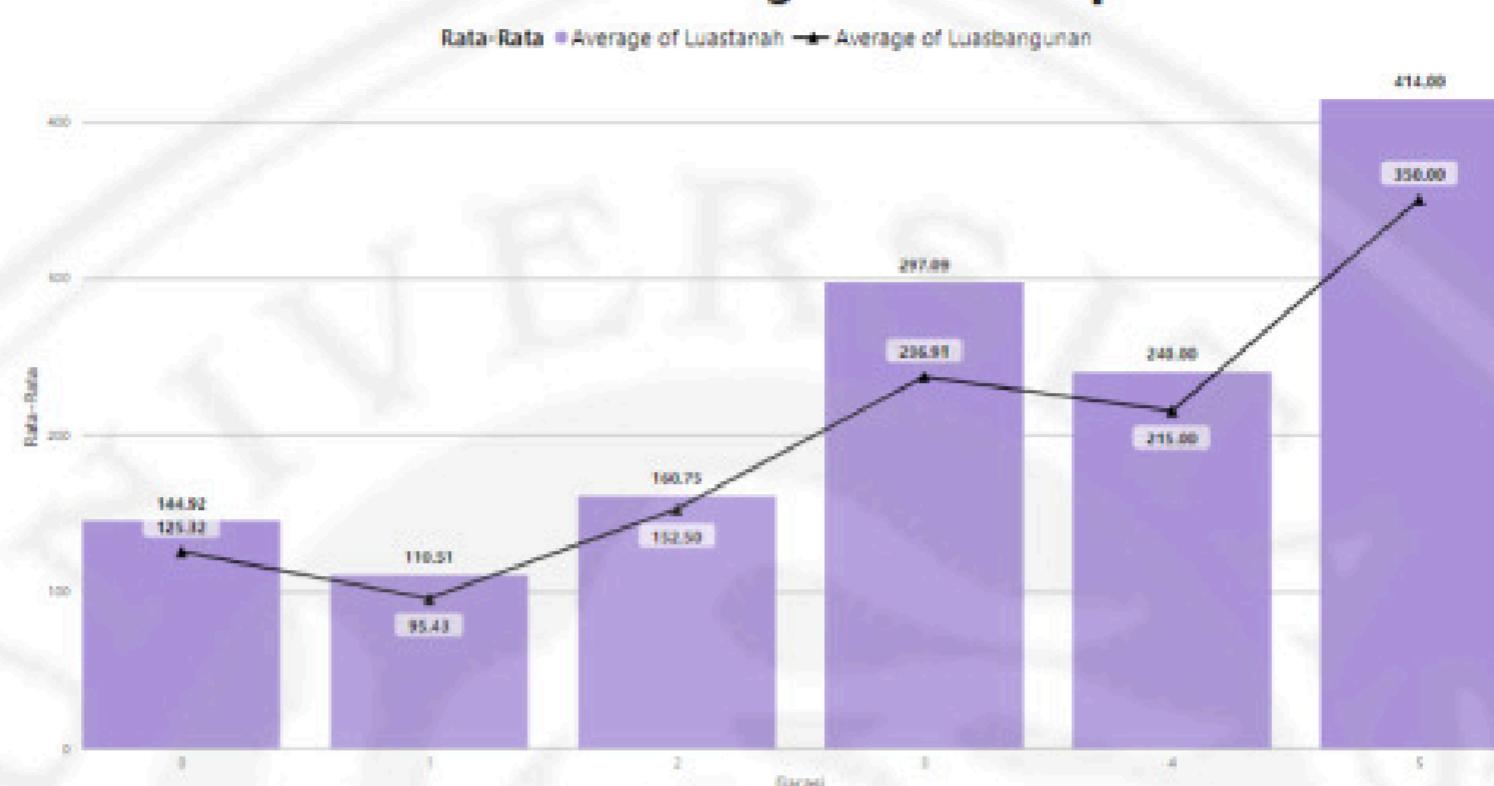
[BEBAS PERPUSTAKAAN  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/BEBASPERPUSTAKAAN\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/BEBASPERPUSTAKAAN)

[KUNJUNGAN  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/KUNJUNGAN\)](https://library.gunadarma.ac.id/DEPC/system/KUNJUNGAN)

[E-PAPER  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC](https://library.gunadarma.ac.id/DEPC/system/E-PAPER)

memiliki rata-rata harga jual rumah yang disesuaikan dengan jumlah rumah di daerah tersebut dan harga rumahnya.

### Rata-Rata Luas Tanah dan Bangunan Terhadap Jumlah Garasi



Gambar 4. 8 Visualisasi 3

Berdasarkan Gambar 4.8 di atas dapat dilihat bahwa semakin besar luas tanah maka semakin besar juga luas yang dimiliki oleh bangunan tersebut. Begitupun sebaliknya jika luas tanah kecil maka luas bangunan yang dimiliki juga kecil. Berdasarkan data jumlah muatan kendaraan dalam garasi tidak mempengaruhi rata-rata luas tanah dan luas bangunan hal ini terlihat ketika jumlah muatan 1 kendaraan dalam garasi memiliki luas tanah dan luas bangunan yang relatif lebih kecil dibandingkan dengan jumlah muatan 0 kendaraan dalam garasi. Hal ini mungkin bisa disebabkan ada beberapa pengiklanan pada website rumah123.com yang kurang dalam menulis informasi tentang muatan kendaraan pada garasi.

Visualisasi data merupakan tahap terakhir dari proses data *understanding* pada bagian menelaah data. Tahap selanjutnya setelah mengetahui karakteristik dari masing-masing data maka akan dilakukan proses data *preparation*.

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

### 4.3. Data Preparation

Pada proses ini akan melakukan beberapa tahap untuk meningkatkan kualitas pada data yang sudah dilihat karakteristik pada proses sebelumnya. Proses ini memiliki pengaruh besar terhadap tingkat akurasi yang dihasilkan oleh model.

#### 4.3.1. Seleksi Data

Seleksi fitur digunakan untuk mendapatkan fitur yang mempunyai pengaruh lebih banyak dengan target *output*. Terdapat dua cara yaitu dengan *feature selection* dan *feature importance*, disini akan menggunakan *feature selection* karena tipe target yang akan digunakan merupakan suatu yang harus di prediksi. *Feature selection* ini tidak disarankan untuk metode masalah pada kasus klasifikasi. Pertama-tama tentukan data yang diperlukan dan yang tidak diperlukan. Pada kasus prediksi harga rumah ini akan menghapus data pada kolom nama dan link karena tidak memiliki relevan terhadap data. Pisahkan antara data *independent* dan data *dependent* untuk data *dependent* akan menggunakan data harga sebagai target, sedangkan untuk data *independent* akan menggunakan data kamar mandi, kamar tidur, garasi, luas tanah dan luas bangunan.

```
#apply selectkbest with chi square function
bestfeature = SelectKBest(score_func=f_regression,
                           k=5)
fit = bestfeature.fit(x,y)
dfscore = pd.DataFrame(fit.scores_)
dfcolumn = pd.DataFrame(x.columns)

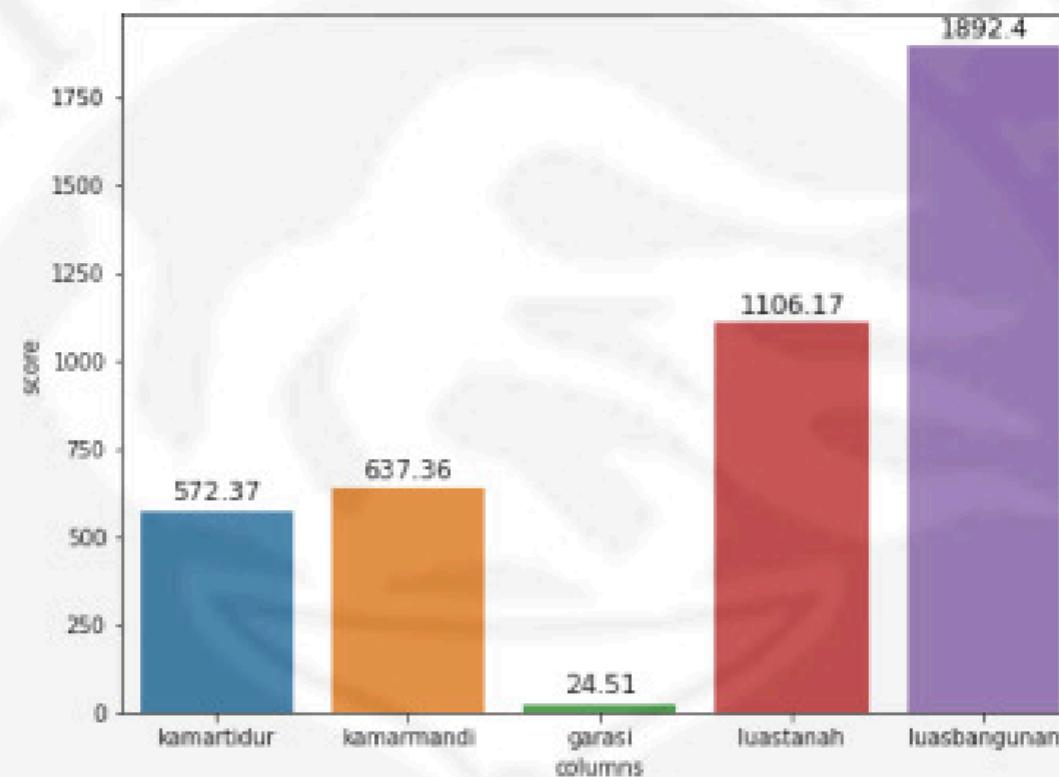
#concat into dataset
features = pd.concat([dfcolumn, dfscore], axis=1)
features.columns = ['columns', 'score']
```

Pada penggunaan *feature selection* menggunakan metode *SelectKBest* yang hanya mengubah pada bagian parameter ‘*score\_func*’ untuk dapat menghitung fitur terbaik pada saat pelatihan data terhadap model. Fungsi *score* secara *default* menggunakan perhitungan anova yang akan menghitung nilai uji F pada masalah

- [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  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/E-PAPER\)](#)

klasifikasi. Pada masalah regresi dapat menggunakan fungsi *score f\_regression* yang akan menghitung nilai uji F pada label untuk target regresi. Masih banyak fungsi *score* yang dapat digunakan pada *feature selection* ini sehingga dapat disesuaikan dengan permasalahan yang akan dipecahkan tersebut. Parameter selanjutnya yang penting untuk dimasukan pada *feature selection* ini adalah jumlah fitur yang akan diperlukan, jumlah *default* adalah 10. Pada kasus ini karena fitur yang digunakan hanya 5 maka nilai k untuk jumlah fitur adalah 5.

**Seleksi Fitur Data Prediksi Harga Rumah**



Gambar 4. 9 Seleksi Fitur pada Data

Berdasarkan Gambar 4.9 di atas dapat dilihat bahwa fitur yang paling mempengaruhi harga rumah adalah luas bangunan. Hasil ini dinyatakan sama dengan hasil korelasi *Spearman*. Fitur yang kurang penting pada kasus ini adalah data garasi dimana mendapatkan *score* paling kecil diantara semua fitur dengan begitu data garasi ini dapat dibuang atau tidak untuk mempercepat proses pelatihan data. Memilih fitur terbaik adalah suatu proses yang sangat penting untuk membantu menghilangkan data yang kurang penting pada saat pelatihan dengan model.

#### 4.3.2. Pembersihan Data

Pada proses ini akan melakukan perbaikan, penghapusan dan pengabaian dari *noise* yang ada pada data. Diperlukan untuk menentukan pendekatan yang tepat

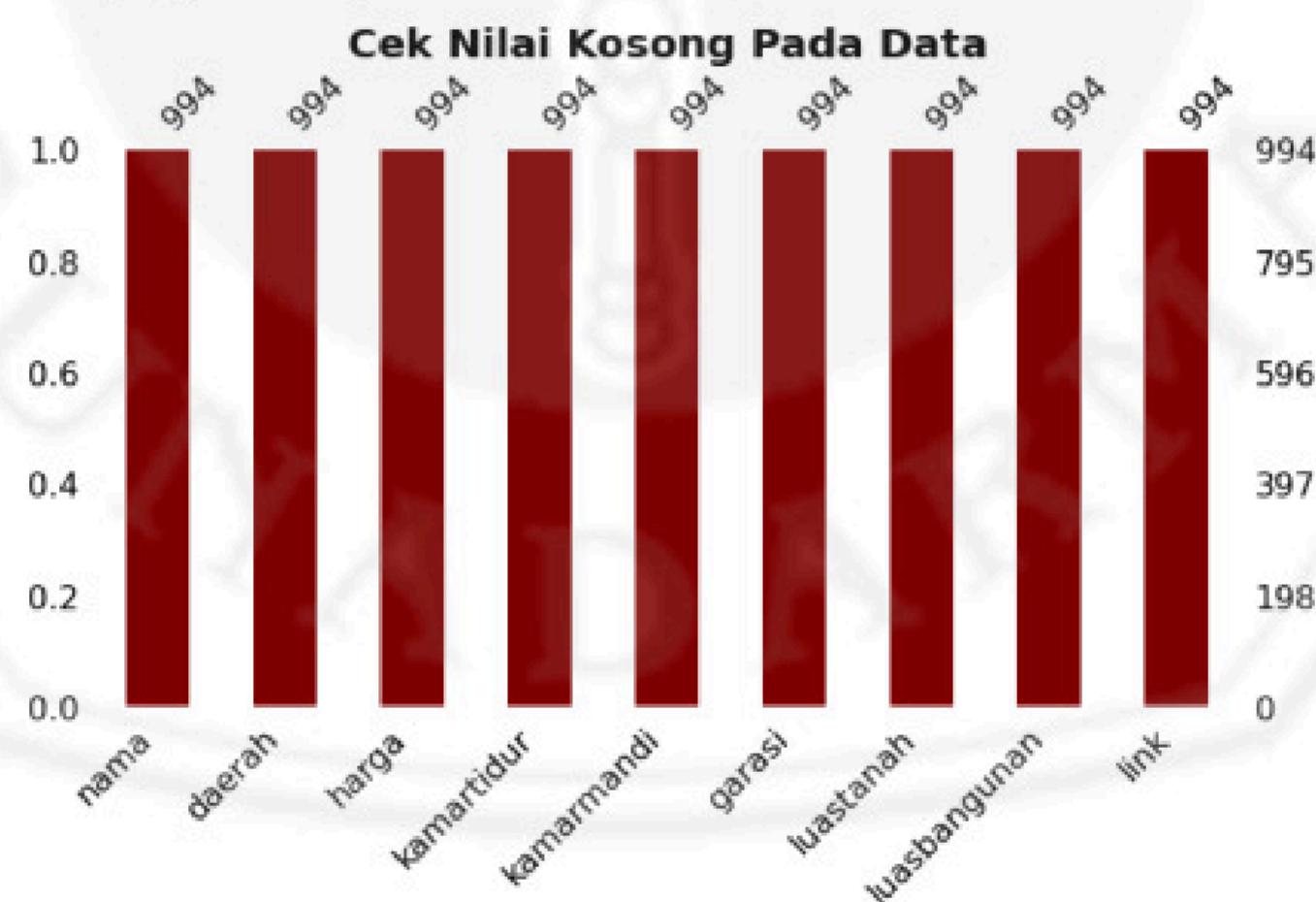
- [HOME  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/\)](#)
- [PEMBAYARAN  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/PEMBAYARAN\)](#)
- [BIODATA DIRI  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA\)](#)
- [PEMINJAMAN BUKU  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA/PEMINJAMAN\)](#)
- [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  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA/BEBAZPERPUSTAKAAN\)](#)
- [KUNJUNGAN  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/KUNJUNGAN\)](#)
- [E-PAPER  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/E-PAPER\)](#)

pada setiap nilai-nilai khusus seperti *outlier*, nilai yang hilang (*missing value*), dan nilai duplikat. Pendekatan dilakukan dengan melihat hasil dari karakteristik data pada proses sebelumnya sehingga mudah untuk menangani agar data siap untuk dilakukan pemodelan.

Pertama-tama lihat jumlah *missing value* yang ada pada data. Agar dapat melihat jumlah *missing value* ini dapat menggunakan 2 cara yaitu menggunakan *print out text* atau menggunakan visualisasi dari *missing value* tersebut. Kedua cara ini disediakan pada bahasa pemrograman *Python*.

```
#looking missing value in plot
msno.bar(data, figsize=(10,6), color='maroon')
plt.title('Check Missing value in Dataset',
          fontsize=20, fontweight='bold')
plt.show()
```

Cara yang paling cepat digunakan untuk mengetahui nilai hilang pada data adalah dengan visualisasi. Pada bahasa pemrograman *Python* terdapat *library* *missingno* yang akan memperlihatkan *missing value* dengan menggunakan visualisasi. Pada *code* di atas menggunakan visualisasi dengan jenis *bar plot* pada data baru yang sudah dipisahkan antara data numerik dan non-numerik.



Gambar 4. 10 Cek Nilai Hilang pada Data

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

Berdasarkan Gambar 4.10 di atas bahwa pada dataset yang akan digunakan tidak memiliki nilai yang hilang. Semua jumlah nilai pada setiap atribut memiliki jumlah yang sama yaitu masing-masing berjumlah 994 data sehingga tidak dibutuhkan untuk melakukan penanganan terhadap nilai hilang. Jika pada data terdapat nilai yang hilang maka harus segera ditangani sesuai dengan karakteristik data tersebut. Pada setiap penanganan nilai hilang memiliki beberapa kriteria yang dapat disesuaikan dengan data yang digunakan.

```
#duplicate data
print('Duplicate      in      Dataset      :      ')
data.duplicated().sum()

print('Ukuran data dengan duplikat : ')
print('Kolom: ', data.shape[1])
print('Baris: ', data.shape[0])
#drop duplicate data
data.drop_duplicates(inplace=True)
print('Ukuran data tanpa duplikat : ')
print('Kolom: ', data.shape[1])
print('Baris: ', data.shape[0])
```

Source code di atas berfungsi untuk melihat data duplikat menggunakan *function* `duplicated.sum()` yang akan mengeluarkan outlier jumlah data duplikat. Penanganan yang paling tepat untuk data duplikat adalah dengan cara dihapus sehingga model tidak salah atau terjadi bias dalam mempelajari pola pada data saat pelatihan. Lihat ukuran data dengan duplikat dan ukuran data tanpa duplikat menggunakan shape dari baris dan kolom masing-masing dataset.

```
Ukuran data dengan duplikat :
Kolom: 9
Baris: 994
Ukuran data tanpa duplikat :
Kolom: 9
Baris: 990
```

Gambar 4. 11 Cek Data Duplikat

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

Pada Gambar 4.11 terlihat bahwa ada data duplikat sebanyak 4 data sehingga data tersebut harus dihapus. Sehingga ukuran dari data berubah menjadi 990 dengan data tanpa duplikat. Proses selanjutnya adalah melihat *outlier* pada dataset.

```
#check outlier in dataset
#use IQR
q1 = datanew.kuartil(0.25)
q3 = datanew.kuartil(0.75)
iqr = q3-q1
outlier = ((datanew < (q1 - 1.5 * iqr)) | (datanew >
(q3 + 1.5 * iqr))).sum()
#check outlier with boxplot
plt.figure(figsize=(20,10))
plt.subplot(2,3,1)
sns.boxplot(x=datanew['kamarmandi'])
plt.subplot(2,3,2)
sns.boxplot(x=datanew['kamartidur'])
plt.subplot(2,3,3)
sns.boxplot(x=datanew['garasi'])
plt.subplot(2,3,4)
sns.boxplot(x=datanew['luastanah'])
plt.subplot(2,3,5)
sns.boxplot(x=datanew['luasbangunan'])
plt.subplot(2,3,6)
sns.boxplot(x=datanew['harga'])
plt.show()
```

Berdasarkan *source code* di atas dilakukan pengecekan *outlier* dengan menggunakan IQR. Pada *Python* perhitungan IQR dilakukan secara manual dengan memberikan beberapa rumus karena *Python* tidak menyediakan *library* untuk mengecek *outlier* secara otomatis. Visualisasi yang dapat digunakan untuk

[HOME  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/\)](https://library.gunadarma.ac.id/DEPC/system/)

[PEMBAYARAN  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/PEMBAYARAN\)](https://library.gunadarma.ac.id/DEPC/system/PEMBAYARAN)

[BIODATA DIRI  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA)

[PEMINJAMAN BUKU  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/PEMINJAMAN\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/PEMINJAMAN)

[PENULISAN ILMIAH  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/PENULISAN\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/PENULISAN)

[FOTOCOPY  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/FOTOCOPY\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/FOTOCOPY)

[SUMBANG BUKU  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/SUMBANGBUKU\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/SUMBANGBUKU)

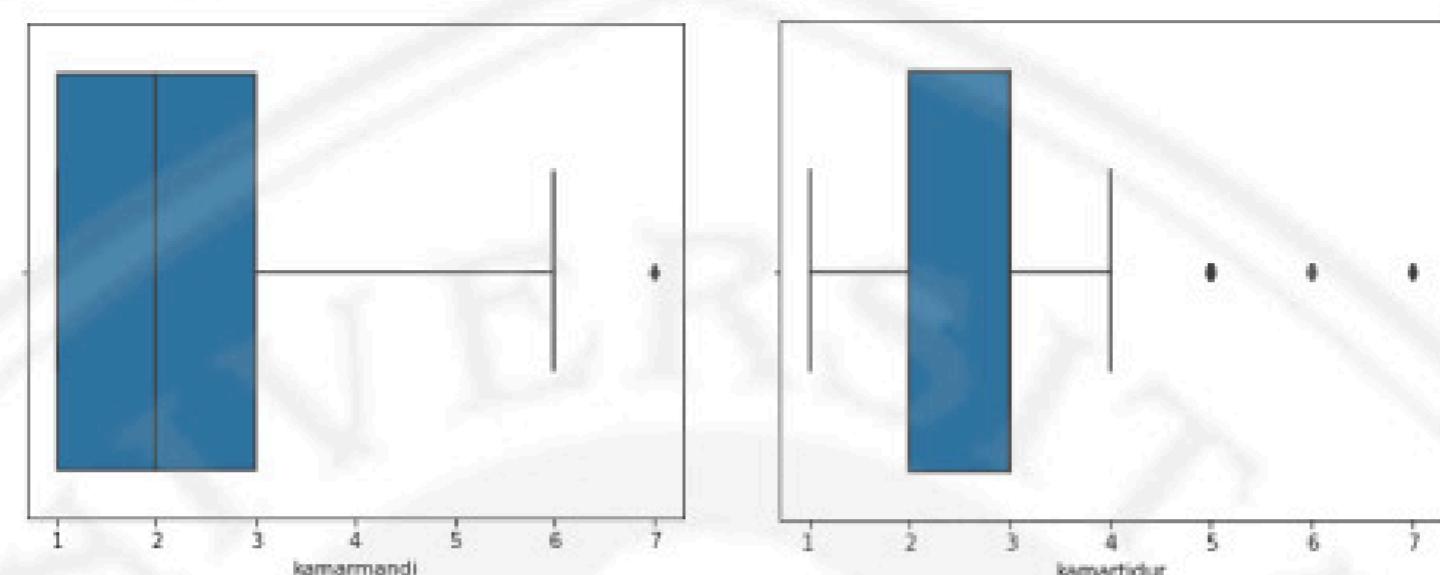
[KASUS & NOTIFIKASI  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/NOTIFIKASI\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/NOTIFIKASI)

[BEBAS PERPUSTAKAAN  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/BEBASPERPUSTAKAAN\)](https://library.gunadarma.ac.id/DEPC/system/ANGGOTA/BEBASPERPUSTAKAAN)

[KUNJUNGAN  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/KUNJUNGAN\)](https://library.gunadarma.ac.id/DEPC/system/KUNJUNGAN)

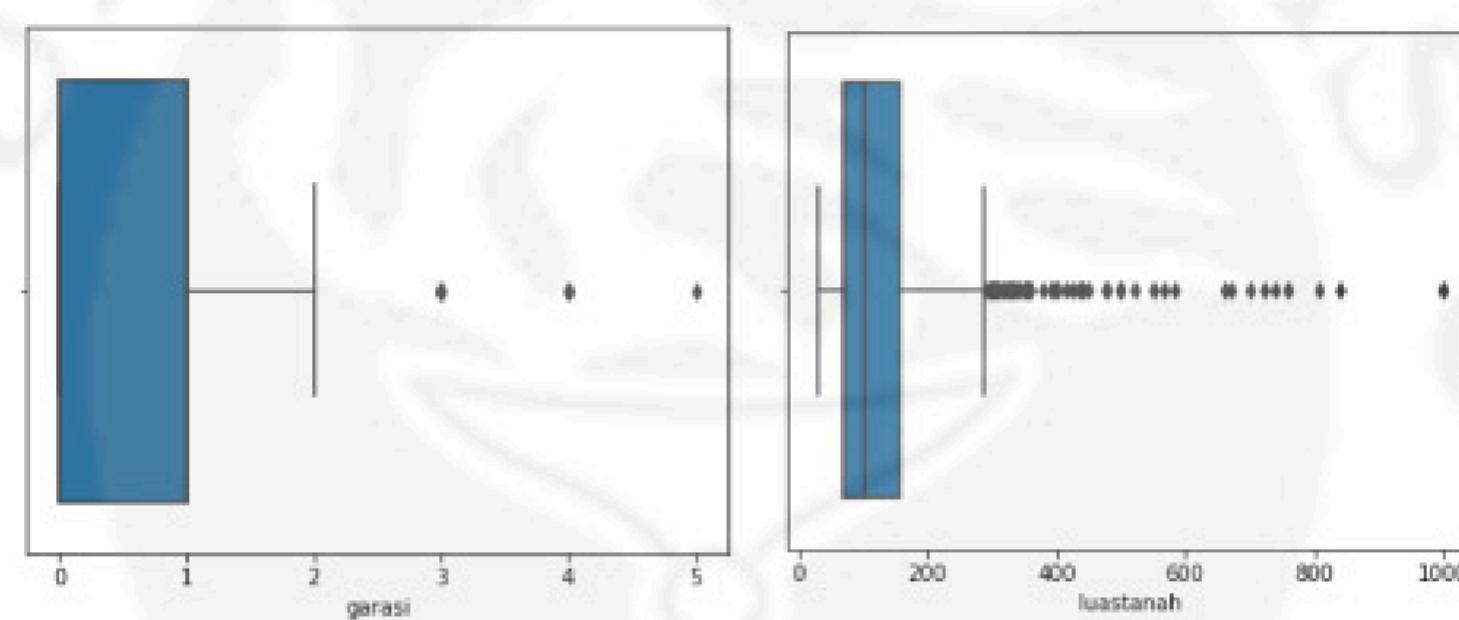
[E-PAPER  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC\)](https://library.gunadarma.ac.id/DEPC)

pengecekan *outlier* pada seluruh dataset akan menggunakan bahasa pemrograman *Python* dengan visualisasi *boxplot*.



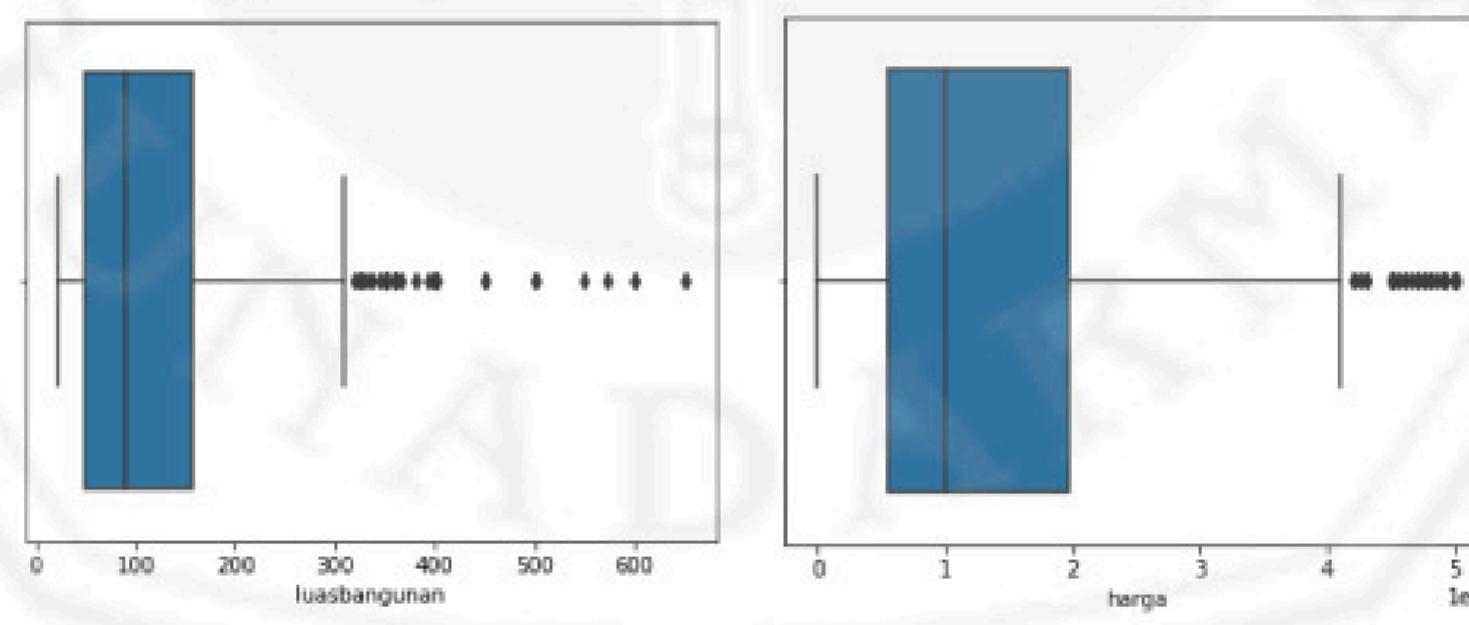
(a) Outlier kamar mandi

(b) Outlier kamar tidur



(c) Outlier garasi

(d) Outlier luas tanah



(e) Outlier luas bangunan

(f) Outlier harga

Gambar 4. 12 Cek *Outlier* pada Seluruh Data

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

Berdasarkan Gambar 4.12 di atas dapat dilihat bahwa banyak *outlier* yang ada pada data terutama pada data luas tanah. Penanganan yang akan dilakukan terhadap *outlier* ini adalah dengan cara di hapus karena data prediksi memiliki sifat yang sangat sensitif untuk memberikan hasil jadi apabila ada data pencilan atau titik data yang jauh dari data yang lain akan menyebabkan kesalahan terhadap prediksi yang cukup fatal. Menghapus data ini akan menyebabkan nilai *mean* yang mengecil dan jumlah data yang berubah.

	harga	kamartidur	kanarmandi	garasi	luastanah	luasbangunan
count	8.480000e+02	848.000000	848.000000	848.000000	848.000000	848.000000
mean	1.116580e+09	2.744104	1.963443	0.752358	105.145047	95.895047
std	8.247419e+08	0.712074	0.886649	0.714164	50.918519	58.905380
min	3.500000e+06	1.000000	1.000000	0.000000	29.000000	21.000000
25%	4.650000e+08	2.000000	1.000000	0.000000	69.000000	50.000000
50%	6.655000e+08	3.000000	2.000000	1.000000	90.000000	80.000000
75%	1.600000e+09	3.000000	2.000000	1.000000	126.000000	130.000000
max	3.950000e+09	4.000000	6.000000	2.000000	287.000000	308.000000

Gambar 4. 13 Data Sudah Bersih

Berdasarkan Gambar 4.13 di atas adalah jumlah data yang sudah bersih yaitu 848 data. Hasil data ini akan dilakukan proses konstruksi untuk mempersiapkan data ke tahap *modeling*.

#### 4.3.3. Konstruksi Data

Tahap ini akan melabelkan data yang sudah dibersihkan. Pada kasus prediksi harga rumah ini tidak melakukan proses *scaling* dikarenakan karena data yang digunakan adalah data real untuk proses prediksi maka dari itu tidak akan mengubah skala pada data. Pada tahap ini hanya melakukan pelabelan data pada kolom daerah dengan mengubah menjadi angka menggunakan metode *LabelEncoder()*.

```
le = LabelEncoder()
data_nooutlier['daerah'] =
le.fit_transform(data_nooutlier['daerah'])
print(le.classes_)
print(np.sort(data_nooutlier['daerah'].unique()))
```

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

Pada *code* di atas akan melabelkan data pada *field* daerah dengan menggunakan *library* LabelEncoder(). Pada *library* ini akan memberikan label berupa angka terhadap nilai unik data berdasarkan abjad.

```
['Bandung' 'Banjar' 'Bekasi' 'Bogor' 'Cianjur' 'Cikarang' 'Cimahi'  
'Cirebon' 'Depok' 'Indramayu' 'Karawang' 'Kuningan' 'Purwakarta' 'Subang'  
'Sumedang']  
[ 0  1  2  3  4  5  6  7  8  9 10 11 12 13 14]
```

Gambar 4. 14 Hasil Pelabelan Data

Berdasarkan Gambar 4.14 di atas dapat dilihat bahwa sekarang data kota sudah berubah menjadi angka yang diurutkan berdasarkan abjad daerah tersebut. Proses pelabelan ini sangat penting dilakukan karena *machine learning* tidak dapat mengolah data dengan sifat *string* atau *object*. Semua data yang akan dilakukan pemodelan harus berupa angka *integer* atau *float*.

```
scaler = MinMaxScaler()  
col = ['daerah', 'harga', 'kamartidur',  
       'kamarmandi', 'garasi', 'luastanah',  
       'luasbangunan']  
data_nor = scaler.fit_transform(data[col])
```

Berdasarkan *source code* di atas merupakan proses *scaling* data menggunakan *MinMaxScaler*. Langkah pertama inisialisasi kan metode yang akan digunakan, kemudian pisahkan data numerik dan data non-numerik pada kasus ini akan menggunakan variabel col untuk pemisahan data. Data akan di fit dengan metode yang dipilih dan menggunakan data numerik berdasarkan kolom yang sudah di pisahkan.

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/BEBAСПERPUSTAKAAN)

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

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

	daerah	harga	kamartidur	kamarmandi	garasi	luastanah	luasbangunan
0	0.571429	0.199544	0.666667	0.2	0.5	0.166667	0.236934
1	0.571429	0.200051	0.666667	0.2	0.5	0.166667	0.233449
2	0.571429	0.199291	0.666667	0.2	0.5	0.166667	0.240418
3	0.571429	0.219562	0.666667	0.2	0.5	0.166667	0.139373
4	0.571429	0.173952	0.666667	0.2	0.5	0.135659	0.153310
...	...	...	...	...	...	...	...
843	0.571429	0.181553	0.333333	0.0	0.5	0.236434	0.118467
844	0.571429	0.201824	0.666667	0.2	0.5	0.178295	0.170732
845	0.000000	0.174712	0.666667	0.2	1.0	0.124031	0.101045
846	0.000000	0.097681	0.666667	0.2	0.5	0.116279	0.066202
847	0.142857	0.379197	0.666667	0.2	0.0	0.383721	0.623693

Gambar 4. 15 Hasil Normalisasi Data

Berdasarkan Gambar 4.15 di atas hasil data yang telah dinormalisasi sudah memiliki rentang yang sama antar kolom. Jika sudah melakukan proses data konstruksi maka data sudah siap untuk dimasukan ke dalam tahap model dengan algoritma *Random Forest Regression*.

#### 4.4. Modeling

Pada proses ini merupakan proses yang paling penting untuk melihat hasil dari pengolahan data yang sudah dilakukan dengan beberapa proses sebelumnya.

##### 4.4.1. Pemisahan Data

Pada proses pemodelan algoritma *Random Forest Regression* akan menggunakan 3 kriteria pemisahan data yaitu 80:20, 75:25, 70:30. Pisahkan data *training* dan data *testing* yang akan digunakan pada masing-masing fungsi data tersebut. Data *training* akan berisi seluruh data kecuali data harga, sedangkan pada data *testing* akan berisi hanya data harga sebagai target dari model.

```
#data training
X = data_nor.drop(columns='harga').values
#data testing
y = data_nor['harga'].values
```

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

Berdasarkan *source code* di atas data latih akan mengambil 6 kolom dengan menghapus data harga yaitu daerah, garasi, kamar mandi, kamar tidur, luas bangunan, dan luas tanah. Pada data uji akan mengambil data harga sebagai target dari penelitian ini.

```
#split 80% data training 20% data testing
X_train80, X_test80, y_train80, y_test80 =
train_test_split(X, y, test_size=0.2, random_state=42)
#split 70% data training 30% data testing
X_train70, X_test70, y_train70, y_test70 =
train_test_split(X, y, test_size=0.3, random_state=42)
#split 75% data training 25% data testing
X_train75, X_test75, y_train75, y_test75 =
train_test_split(X, y, test_size=0.25,
random_state=42)
```

Berdasarkan *source code* di atas merupakan proses yang dilakukan untuk membagi data ke dalam 3 kriteria pemisahan. Pada masing-masing kriteria memiliki *random\_state* yang sama yaitu 42 berarti bahwa setiap kriteria akan menghasilkan nilai acak yang sama sehingga tidak merubah hasil apabila dilakukan *run* ulang pada program. Penentuan jumlah banyaknya pemisahan data terletak pada parameter *test\_size* yang membagi data *testing* dan sisanya sebagai data *training*.

Tabel 4. 2 Hasil Pemisahan Data

Kriteria	Data Latih	Data Uji
80:20	659	165
75:25	618	206
70:30	577	247

Berdasarkan Tabel 4.2 di atas adalah hasil dari pemisahan data menggunakan 3 kriteria. Masing-masing kriteria ini akan dilakukan pelatihan terhadap algoritma *Random Forest Regression*.

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/BEBASTERPUSTAKAAN)

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

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

#### 4.4.2. Modeling Algoritma Random Forest Regression

Setelah proses pemisahan data sudah selesai lalu akan dilakukan proses *modeling* untuk melatih data *training* yang sudah di pisahkan pada masing-masing kriteria dengan algoritma *Random Forest Regression*.

```
forest80 = RandomForestRegressor(n_estimators=300,  
random_state=42)  
forest80.fit(X_train80, y_train80)  
forest75 = RandomForestRegressor(n_estimators=300,  
random_state=42)  
forest75.fit(X_train75, y_train75)  
forest70 = RandomForestRegressor(n_estimators=300,  
random_state=42)  
forest70.fit(X_train70, y_train70)
```

Source code di atas merupakan pelatihan algoritma *Random Forest Regression* dengan masing-masing pemisahan data. Parameter yang akan ditentukan hanya jumlah pohon yang terbentuk yaitu 300 sedangkan untuk parameter lain seperti *bootstrap*, *max\_features* dan lainnya akan ditentukan secara *default*. Pada masing-masing skenario akan dilakukan pelatihan dengan algoritma *Random Forest Regression* dengan parameter yang sama.

```
y_pred_for80 = Forest80.predict(X_test80)  
y_pred_for70 = Forest70.predict(X_test70)  
y_pred_for75 = Forest75.predict(X_test75)
```

Gunakan data uji untuk melakukan pengujian prediksi pada model berdasarkan pemisahan data yang digunakan. Data hasil prediksi ini akan digunakan untuk menghitung nilai *error* pada saat proses validasi model.

#### 4.4.3. Tuning Parameter

Pada proses ini dilakukan untuk meningkatkan hasil akurasi yang pada model. Pada tahap *modeling* dengan menggunakan algoritma *Random Forest Regression* hanya menggunakan parameter *n\_estimators*, sedangkan untuk

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

parameter lain di gunakan secara *default*. Tuning parameter ini akan memanfaatkan library *GridSearchCV* dalam memberikan parameter terbaik.

```
n_estimators = [100, 200, 300, 500, 650, 600, 750, 1000] #  
number of trees in the Random Forest  
max_features = [2, 3, 'auto', 'sqrt'] # number of  
features in consideration at every split  
max_depth = [int(x) for x in np.linspace(5, 450, num =  
9)] # maximum number of levels allowed in each decision  
tree  
min_samples_split = [2, 6, 10] # minimum sample number  
to split a node  
min_samples_leaf = [1, 2, 3, 4] # minimum sample number  
that can be stored in a leaf node  
criterion = ['squared_error', 'absolute_error',  
'poisson']  
bootstrap = [True, False] # method used to sample data  
points  
random_state = [3, 12, 42] # random value which make  
the result not change  
  
Random_grid = {'n_estimators': n_estimators,  
'max_features': max_features,  
'max_depth': max_depth,  
'min_samples_split': min_samples_split,  
'min_samples_leaf': min_samples_leaf,  
'criterion': criterion,  
'bootstrap': bootstrap}  
  
#input into gridsearch  
model = RandomForestRegressor()
```

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/BEBASTERPUSTAKAAN)

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

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

```
gsearch = GridSearchCV(model, param_grid=Random_grid,
cv=10, n_jobs=-1, verbose=2)
```

Pada *source code* di atas pertama-tama definisikan parameter yang akan digunakan untuk *tuning*. Pada kasus ini akan menggunakan semua parameter yang ada pada algoritma *Random Forest Regression*. Jika sudah di definisikan lalu masukan semua parameter ke dalam *param\_grid*, kemudian masukan model yang akan digunakan. Tentukan nilai *k* untuk *cross validation* pada *tuning parameter* disini akan menggunakan 10 *k-fold* yang berarti akan membagi data menjadi 10 bagian dengan data latih. Langkah selanjutnya menggunakan *verbose* = 2 untuk menampilkan progress bar pada saat melakukan *cross validation*. Gunakan *n\_jobs* = -1 karena pada penelitian ini akan menggunakan semua prosesor pada saat menjalankan *tuning*.

```
Best Parameters (80:20): {n_estimators: 500, min_samples_split: 2, min_samples_leaf: 4, max_features: 3,
max_depth: 450, criterion: squared_error, bootstrap: True, random_state: 42}
Best Parameters (75:25): {n_estimators: 300, min_samples_split: 2, min_samples_leaf: 2, max_features: 3,
max_depth: 375, criterion: squared_error, bootstrap: True, random_state: 42}
Best Parameters (70:30): {n_estimators: 300, min_samples_split: 2, min_samples_leaf: 2, max_features: 2,
max_depth: 200, criterion: squared_error, bootstrap: True, random_state: 42}
```

Gambar 4. 16 Hasil *Tuning parameter*

Berdasarkan Gambar 4.16 di atas merupakan parameter terbaik yang berhasil dipilih pada proses *tuning parameter*. Jika sudah lakukan pengecekan terhadap perbedaan akurasi yang dihasilkan dari model sebelum *tuning* dan setelah *tuning*.

### Akurasi Model



Gambar 4. 17 Akurasi Data Uji Model Setelah *Tuning parameter*

Berdasarkan Gambar 4.17 di atas akurasi data uji tertinggi didapatkan oleh kriteria 75% data latih dan 25% data uji dengan hasil prediksi menggunakan parameter metode Grid Search CV naik sekitar 4.99%. Dapat disimpulkan bahwa

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

model saat mempelajari data latih dan memprediksi menggunakan data uji cukup baik dilakukan dengan jarak akurasi yang tidak begitu besar sehingga model tidak mengalami *overfitting*.



Gambar 4. 18 Hasil *Random Forest* Setelah *Tuning parameter*

Berdasarkan Gambar 4.18 di atas merupakan hasil pohon yang berhasil dibentuk oleh model dengan split data 75% data latih dan 25% data uji setelah menggunakan parameter *gridsearchCV*. Hasil dari *tuning parameter* dengan 75% data latih dan 25% data uji akan digunakan untuk proses evaluasi selanjutnya dan proses *deployment*.

#### 4.4.4. Modeling Algoritma Cosine Similarity

Pada proses pemodelan algoritma sistem rekomendasi ini akan menggunakan data split dengan kriteria 75% data latih dan 25% data uji. Data latih

- [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  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/BEBASTERPUSTAKAAN\)](#)
- [KUNJUNGAN  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/KUNJUNGAN\)](#)
- [E-PAPER  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC](#)

akan digunakan pada saat pelatihan sedangkan data uji akan digunakan untuk prediksi hasil rekomendasi.

```
daerah= 2
kamar tidur = 2
kamar mandi = 1
garasi = 1
luas tanah = 60
luas bangunan = 54
code = [[daerah, kamar tidur, kamar mandi, garasi, luas
tanah, luas bangunan]]
predict = model.predict(code)
predcode = [[daerah, kamar tidur, kamar mandi, garasi,
luas tanah, luas bangunan, int(predict)]]
```

Pada *source code* di atas pertama-tama masukan data input yang akan di prediksi oleh model. Data input tersebut dimasukan ke dalam *array* 2 dimensi lalu di prediksi menggunakan model *Random Forest Regression* yang telah di simpan sebelumnya. Selanjutnya satukan hasil prediksi dengan data input untuk menjadi sebuah variabel dengan kumpulan *array*.

Predict : Rp.577,593,200

Gambar 4. 19 Hasil Prediksi Data Rekomendasi

Berdasarkan Gambar 4.19 merupakan hasil dari prediksi data rekomendasi yang dimasukan pada *code* di atas. Kemudian hasil dari harga beserta atribut lainnya akan dihitung dengan algoritma *Cosine Similarity* untuk mendapatkan hasil rekomendasi berdasarkan data latih.

```
rec = cosine_similarity(predcode, datar)
recidx = rec.argsort()[0, 0:10]
recommend = data.loc[recidx]
dtrecomendation = recommend[['name','harga','daerah',
'kamar tidur', 'kamar mandi', 'garasi', 'luas
tanah','luas bangunan','link']].reset_index()
```

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

Pada *code* di atas merupakan pembentukan data untuk memberikan hasil rekomendasi berdasarkan nilai *Cosine Similarity* antara data latih dengan data input yang dimasukan pengguna. Urutkan index dari hasil *Cosine Similarity* menggunakan *method* argsort() dan ambil sebanyak 10 data. Tampilkan data berdasarkan index dari urutan data *Cosine Similarity* dengan urutan kolom yaitu nama rumah, harga, kota, kamar tidur, kamar mandi, garasi, luas tanah, luas bangunan dan link. Data yang direkomendasikan ini berdasarkan data bersih yang digunakan pada penelitian, jadi masih ada kemungkinan terjadi tidak ada rekomendasi berdasarkan data yang diproses.

```
def recomendation2(data):
    minprice = int(predict2) - 100000000
    maxprice = int(predict2) + 100000000
    minsur = int(sur) - 10
    maxsur = int(sur) + 10
    data = data[(data['daerah'] == (daerah)) &
                (data['harga'].between(minprice,
                                      maxprice)) &
                (data['luastanah'].between(minsur,
                                           maxsur))]
    data['harga'] = data['harga'].map('{:, .0f}'.format)
    data = data.drop(columns = ['index']).head()

    return data
```

Berdasarkan *source code* di atas merupakan proses content based *filtering* sebagai sebuah metode tambahan untuk pelengkap dari hasil rekomendasi rumah yang diberikan. Fitur yang akan digunakan adalah 3 fitur yang memiliki kontribusi penting bagi pengguna dalam memilih harga rumah. Fitur pertama adalah daerah yang harus mengeluarkan hasil sama dengan inputan yang diberikan oleh pengguna pada saat proses prediksi. Fitur kedua adalah luas tanah sebagai elemen penting, pada fitur ini memiliki kesesuaian pada luas bangunan jadi fitur ini bisa menjadi salah satu fitur penting dalam menentukan rekomendasi rumah yang diberikan yaitu

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

memiliki rentang luas tanah di antara nilai minimal dan maksimal luas tanah yang sudah di hitung. Fitur terakhir adalah harga memiliki rentang antara minimal dan maksimal berdasarkan perhitungan, pada nilai minimal harga maka harga prediksi yang dihasilkan akan dikurangkan oleh 100 juta sedangkan pada nilai maksimal akan ditambahkan oleh 100 juta dari harga hasil prediksi.

	nama	harga	daerah	kamar_tidur	kamarmandi	garasi	luastanah	luasbangunan	link
28	Tipe Uptown Villa di Uptown Estate Bekasi	889.000.000	Bekasi	2	2	0	60	66	<a href="https://www.rumah123.com/properti/bekasi/hos66">https://www.rumah123.com/properti/bekasi/hos66</a>
31	Step Hill Smart Home Mewah Jatiuwit	880.000.000	Bekasi	2	1	1	60	40	<a href="https://www.rumah123.com/properti/bekasi/hos40">https://www.rumah123.com/properti/bekasi/hos40</a>
41	Rumah Murah Banglo Nyaman Batu Renovasi , Blt	579.000.000	Bekasi	2	1	0	65	45	<a href="https://www.rumah123.com/properti/bekasi/hos45">https://www.rumah123.com/properti/bekasi/hos45</a>
72	Rumah Mezzanine Design Exclusive Di Jatimulya,	539.000.000	Bekasi	2	1	1	60	46	<a href="https://www.rumah123.com/properti/bekasi/hos46">https://www.rumah123.com/properti/bekasi/hos46</a>
86	Step Hill Smart Home Mewah	500.000.000	Bekasi	2	1	1	60	40	<a href="https://www.rumah123.com/properti/bekasi/hos40">https://www.rumah123.com/properti/bekasi/hos40</a>

Gambar 4. 20 Data Hasil Rekomendasi Menggunakan *Python*

Berdasarkan Gambar 4.20 hasil yang diberikan oleh perhitungan *Cosine Similarity* ini cukup dekat dengan data input yang dimasukan oleh pengguna. Terlihat dari kota yang dipilih, luas bangunan yang hampir mendekati data input, dan sebagainya.

#### 4.5. Evaluasi

Pada proses ini akan dilakukan dengan 2 cara yaitu menggunakan perhitungan manual dengan data sampel dan menggunakan bahasa pemrograman *Python*.

##### 4.5.1. MSE (*Mean Squared Error*)

Saat proses *modeling* selesai maka dilakukan perhitungan untuk melihat *error* yang dihasilkan pada model. Proses penggunaan pada bahasa pemrograman *Python* sama dengan MSE yaitu menggunakan data uji dan data prediksi.

Menghitung nilai MSE pada semua dataset akan menggunakan bahasa pemrograman *Python*.

```
MSEfor_75 = mean_squared_error(y_test75, y_pred_for75)
gscvMSEfor_75      =      mean_squared_error(y_test75,
y_pred_for_gscv75)

MSEcosine = mean_squared_error(dtrec_test, dtrecpred)
```

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

```
print('MSE Model Random Forest Sebelum Tuning: ',  
      MSEfor_75)  
print('MSE Model Random Forest Setelah Tuning: ',  
      gscvMSEfor_75)  
print('MSE Model Cosine Similarity: ', MSEcosine)
```

Pada *source code* di atas perhitungan MSE pada *Python* akan menggunakan data uji dan data prediksi pada model *Random Forest Regression*. Pada *Cosine Similarity* akan menggunakan harga pada data uji dan data prediksi.

```
MSE Model Random Forest Sebelum Tuning: 4.37  
MSE Model Random Forest Setelah Tuning: 1.76  
MSE Model Cosine Similarity: 6.26
```

Gambar 4.21 Hasil Evaluasi MSE

Berdasarkan Gambar 4.21 di atas hasil dari nilai evaluasi MSE menggunakan *Python* pada seluruh model yang sudah dilakukan pelatihan. Pada model *Random Forest Regression* menggunakan kriteria 75:25 sebagai kriteria pemisahan data paling baik. Pada model *Cosine Similarity* mengikuti data yang digunakan pada *Random Forest Regression* karena model *Cosine Similarity* terbentuk setelah model prediksi dilakukan pelatihan.

#### 4.5.2. MAE (*Mean Absolute Error*)

Perhitungan evaluasi secara keseluruhan akan menggunakan bahasa pemrograman *Python* dengan seluruh dataset.

```
MAEfor_75 = mean_absolute_error(y_test75,  
                                y_pred_for75)  
gscvMAEfor_75 = mean_squared_error(y_test75,  
                                    y_pred_for_gscv75)  
MAEcosine = mean_absolute_error(dtrec_test, dtrecpred)  
  
print('MAE Model Random Forest Sebelum Tuning: ',  
      MAEfor_75)
```

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

```
print('MAE Model Random Forest Setelah Tuning: ',  
gscvMAEfor_75)  
print('MAE Model Cosine Similarity: ', MAEcosine)
```

Pada perintah di atas merupakan perhitungan MAE menggunakan *Python* pada seluruh data uji dan data prediksi dengan model *Random Forest Regression*. Pada algoritma *Cosine Similarity* akan menggunakan data uji dan data prediksi yang dihasilkan.

```
MAE Model Random Forest Sebelum Tuning: 6.25  
MAE Model Random Forest Setelah Tuning: 2.55  
MAE Model Cosine Similarity: 2.05
```

Gambar 4. 22 Hasil Evaluasi MAE

Berdasarkan Gambar 4.22 di atas merupakan hasil dari evaluasi MAE menggunakan *Python* dengan pemisahan data 75:25 untuk setiap model yang digunakan. Model pada *Random Forest Regression* mengalami penurunan *error* saat menggunakan parameter *tuning*.

#### 4.5.3. *R-Squared*

Pada perhitungan evaluasi *R-Squared* ini akan menghasilkan seberapa besar variabel dependen dapat mempengaruhi variabel independen. Proses penggunaan pada bahasa pemrograman *Python* sama dengan MSE dan MAE yaitu menggunakan data uji dan data prediksi.

```
R2for_75 = r2_score(y_test75, y_pred_for75)  
gscvR2for_75      =      mean_squared_error(y_test75,  
y_pred_for gscv75)  
R2cosine = r2_score(dtrec_test, dtrecpred)  
  
print('R2 Model Random Forest Sebelum Tuning: ',  
R2for_75)  
print('R2 Model Random Forest Setelah Tuning: ',  
gscvR2for_75)  
print('R2 Model Cosine Similarity: ', R2cosine)
```

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/BEBAСПЕРПУСТАКААН)

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

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

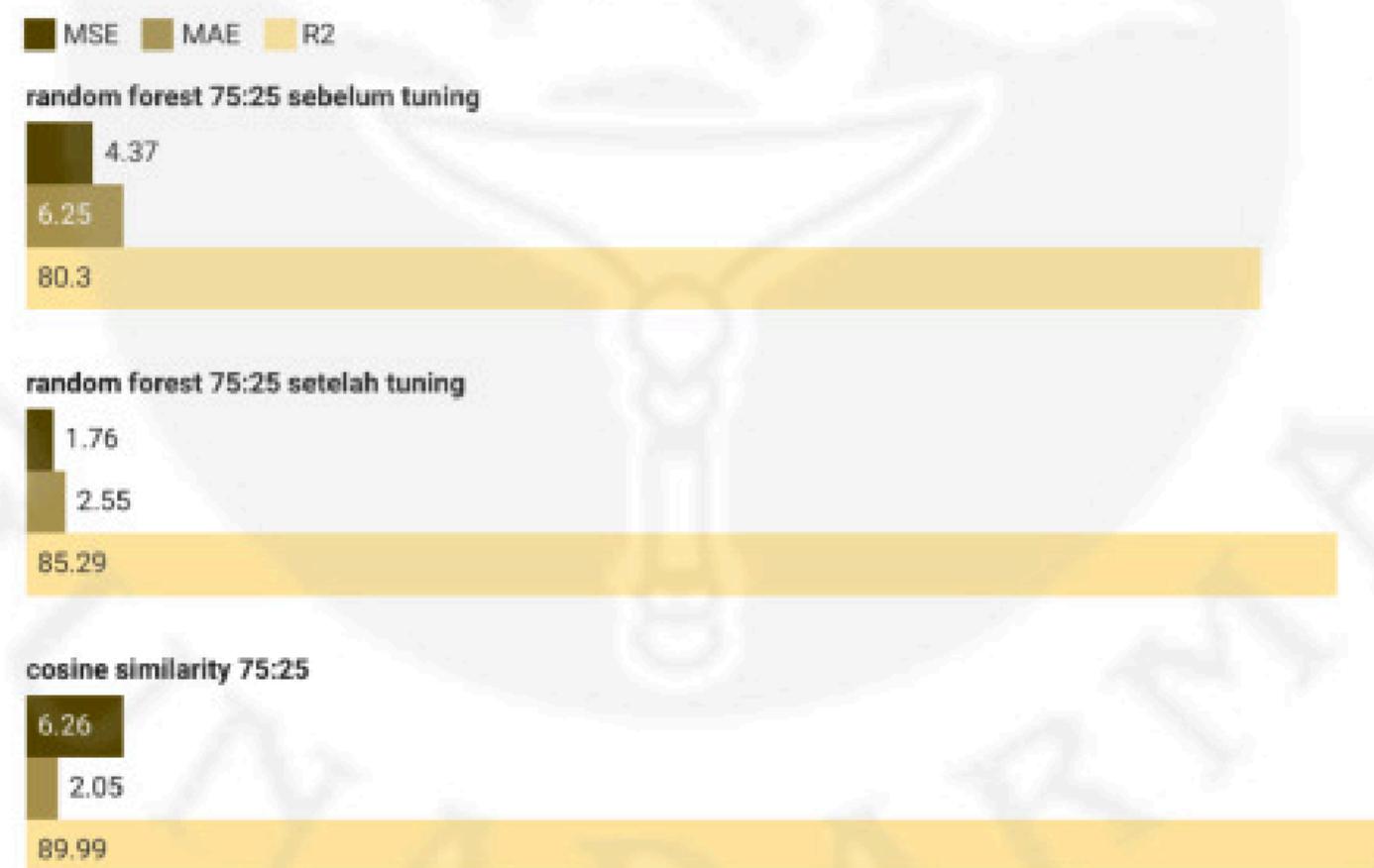
Pada *source code* di atas perhitungan *R-Squared* menggunakan *Python* dengan algoritma *Random Forest Regression* akan menggunakan data uji dan data prediksi pada model. Perhitungan *R-Squared* algoritma *Cosine Similarity* akan menggunakan data uji dan data prediksi dari hasil rekomendasi.

```
R2 Model Random Forest Sebelum Tuning: 80.3
R2 Model Random Forest Setelah Tuning: 85.29
R2 Model Cosine Similarity: 89.99
```

Gambar 4. 23 Hasil Evaluasi *R-Squared*

Berdasarkan Gambar 4.23 di atas merupakan hasil evaluasi pada metrik *R-Squared* menggunakan *Python*. Pada semua model menggunakan pemisahan data yang sama yaitu 75:25. Dapat dilihat bahwa perkembangan yang cukup baik dihasilkan pada model *Random Forest Regression* setelah menggunakan parameter *tuning*.

### Evaluasi Model dengan Python



Gambar 4. 24 Perbandingan Nilai Evaluasi

Berdasarkan Gambar 4.24 di atas dengan menggunakan *Python* pada seluruh dataset yang digunakan untuk pelatihan *random forest* setelah *tuning parameter* mengalami penurunan *error* baik pada metrik MSE maupun MAE hal ini membuktikan bahwa model yang diberikan setelah *tuning* sudah cukup

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

maksimal pada dataset yang digunakan. Hasil yang diperoleh adalah nilai MSE dan MAE  $< 10$  sehingga dapat dikatakan bahwa model algoritma *Random Forest Regression* sudah sangat baik dalam memprediksi data yang dihasilkan. Pada sistem rekomendasi dengan menggunakan algoritma *Cosine Similarity* mendapatkan hasil 6.26 sehingga dapat dikatakan bahwa rekomendasi yang dihasilkan cukup baik. Hasil yang diberikan evaluasi *R-Squared* pada algoritma *Random Forest Regression* mengalami peningkatan yang cukup baik setelah melakukan proses *tuning parameter* dengan hasil *R-Squared* sebesar 85.29% dan pada algoritma *Cosine Similarity* sebesar 89.99%. Hasil dari evaluasi hampir mendekati 100% maka dapat dikatakan bahwa hasil prediksi dan rekomendasi yang diberikan sudah baik.

Model dengan menggunakan parameter pemilihan *tuning* memiliki kriteria yang baik untuk dilakukan proses *deployment* karena memenuhi syarat yaitu nilai *error* yang dihasilkan pada MSE dan MAE mengalami penurunan dan nilai pada *R-Squared* mengalami penaikan *score*. Parameter yang digunakan adalah 300 *n\_estimators* yang berarti bahwa pohon yang terbentuk terdiri dari 300 pohon *decision tree* yang membangun hutan dalam model. *Criterion* yang digunakan pada model terbaik adalah *squared\_error* yang berarti bahwa pengukuran untuk kualitas *split* menggunakan nilai MSE sebagai kriteria pendukung untuk memperoleh informasi. Pada parameter *max\_depth* = 375 yang berarti bahwa panjang *root* sampai daun yang terbentuk mencapai ke dalaman 375 daun. Parameter *min\_samples\_split* yang terbaik adalah 2 yang berarti bahwa jumlah sampel minimum yang dibutuhkan untuk membelah *split* pada pohon sebanyak 2 data. Parameter *min\_samples\_leaf* yang terbaik adalah 2 yang berarti bahwa minimum sampel yang diperlukan pada daun pohon yang terbentuk minimal 2 data. Pada parameter *max\_features* mendapatkan nilai terbaik 3 yang berarti bahwa jumlah maksimal fitur pada data dalam mempertimbangkan pemisahan *split* pohon terbaik adalah 3. Parameter terakhir adalah *random\_state* = 42 yang berarti bahwa nilai acak keluaran mesin akan sama jika terjadi *run* ulang pada program sehingga tidak mengakibatkan nilai prediksi yang berbeda.

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

#### 4.6. Deployment

Pada proses pembuatan *website* sebagai implementasi dari hasil analisis ini akan menggunakan *framework Flask* yang ada pada bahasa pemrograman *Python* sebagai *backend*, lalu menggunakan *HTML* dan *CSS* sebagai *frontend* dari *website* yang akan dibangun.

```
app = Flask(__name__)

@app.route('/')
def index():
    return render_template('index.html')

@app.route('/about')
def about():
    return render_template('about.html')

@app.route('/predict', methods=['POST', 'GET'])
def predict():
    if request.method == 'POST':

        with open('rfr_best_jabar', 'rb') as r:
            model = pickle.load(r)

            daerah = request.form.get("daerah")
            kamar_tidur = request.form.get("kamar tidur")
            kamar_mandi = request.form.get("kamar mandi")
            garasi = request.form.get("garasi")
            luas_tanahface = request.form.get("luas tanahface")
            luas_bangunanding = request.form.get("luas bangunanding")
```

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

```
        datas = np.array((daerah, kamar tidur, kamar
mandi, garasi, luas tanahface, luas bangunanding))
        datas = np.reshape(datas, (1, -1))

        pricepred = model.predict(datas)

        return render_template('result.html', finalData =
        'Rp.           (:,.0f}'.format(int(pricepred)),
        daerah=getdaerah(daerah), kamar mandi=kamar mandi,
        kamar tidur=kamar tidur, luas tanahface=luas
        tanahface, garasi=garasi, luas bangunanding=luas
        bangunanding,
        col=heading,
        column_names=dtrec2.columns.values,
        row_data=list(dtrec2.values.tolist()), zip=zip,
        name=name,           daerahrec=daerahrec,
        garasiingarea=garasiingarea,           price=price,
        dtrec2=[dtrec2.to_html()], kamar tidurroom=kamar
        tidurroom, kamar mandiroom=kamar mandiroom, luas
        tanahfacearea=luas      tanahfacearea,     luas
        bangunandingarea=luas bangunandingarea, link=link

    else:
        return render_template('predict.html')
```

Berdasarkan penggalan *source code* di atas proses implementasi model ke dalam *website* menggunakan *Flask*. Pada variabel app akan menyimpan sebuah halaman *website* kosong dengan menggunakan `return __name__`, lalu untuk *route* awal akan diarahkan ke halaman index.html atau beranda. Apabila pengguna menekan halaman tentang maka akan mengembalikan halaman about.html. Pada halaman prediksi akan diberikan sebuah *form* yang menangkap data berupa post yaitu data inputan dari pengguna. Apabila *form* sudah diisi oleh pengguna maka

- [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  
\(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/E-PAPER\)](#)

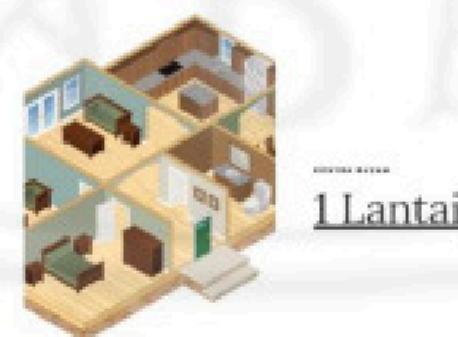
sistem akan membuka model dengan menggunakan *library* pickle.load yang disimpan dengan nama rfr\_best\_jabar sebagai model terbaik. *Form* tersebut akan diubah menjadi bentuk *array* dengan menggunakan *library* numpy dan di reshape ukurannya menjadi 2 dimensi. Data yang sudah tersimpan menjadi *array* akan di prediksi dengan model dan halaman hasil akan mengembalikan hasil prediksi dan rekomendasi berdasarkan data inputan pengguna.

#### 4.6.1. Tampilan Halaman Beranda

Pada halaman ini merupakan halaman pertama yang akan dilihat pada website prediksi harga rumah. Halaman ini juga akan berisi informasi mengenai rumah dan tipe-tipe dari rumah yang akan dibangun sesuai dengan keinginan pengguna.



## TIPE-TIPE RUMAH



Gambar 4. 25 Tampilan Halaman Beranda

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/BEBASTERPUSTAKAAN)

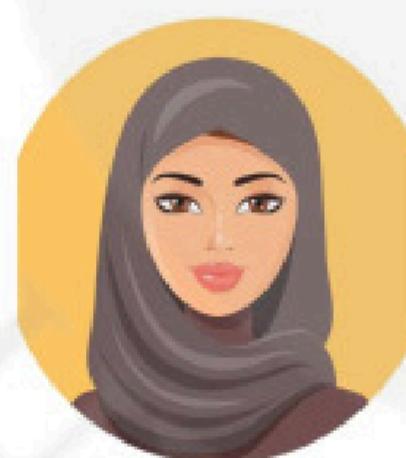
KUNJUNGAN  
(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/KUNJUNGAN)

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

Berdasarkan Gambar 4.25 di atas merupakan tampilan halaman beranda pada *website* prediksi. Pada halaman ini akan menampilkan informasi tentang tipe-tipe rumah mulai dari rumah 1 lantai hingga 5 lantai.

#### 4.6.2. Tampilan Halaman Tentang

Halaman tentang merupakan halaman yang akan mendeskripsikan penulis, dan deskripsi singkat mengenai *website* prediksi.



#### Ega Sri Lestari, Hobi Menulis

Saya mahasiswa semester 8 pada Universitas Gunadarma Jurusan Sistem Informasi.  
Saya sangat tertarik untuk mengembangkan diri saya pada bidang data science.  
Saya adalah orang yang menghargai waktu, disiplin, dapat bekerja secara tim  
maupun individu, dan berlanggung jawab. Saya mempunyai hobi di luar bidang  
programmer yaitu menulis.

Gambar 4. 26 Tampilan Halaman Tentang

Berdasarkan Gambar 4.26 di atas pada tampilan halaman tentang akan menampilkan informasi singkat mengenai deskripsi penulis dan deskripsi penggunaan *website* prediksi untuk pada pengguna aplikasi.

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/BEBASTERPUSTAKAAN)

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

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

#### 4.6.3. Tampilan Halaman Prediksi

Halaman prediksi merupakan halaman yang akan digunakan untuk memberikan input terhadap kriteria rumah yang diinginkan sehingga akan menghasilkan prediksi harga rumah sesuai kriteria.

Gambar 4. 27 Tampilan Halaman Prediksi

Berdasarkan Gambar 4.27 di atas merupakan hasil tampilan halaman prediksi yang bertujuan untuk memberikan input kriteria rumah yang akan diprediksi. Pada pemilihan daerah akan ditampilkan inputan *dropdown* yang berisi kota dan kabupaten di Jawa Barat. Pada inputan jumlah kamar mandi, kamar tidur, dan garasi berupa *number input* dengan *range* sesuai dengan data yang digunakan pada model. Inputan selanjutnya pada luas tanah dan luas bangunan menggunakan

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/BEBASTERPUSTAKAAN)

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

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

*slider input* agar memudahkan memilih luas tanah dan luas bangunan bagi pengguna.

#### 4.6.4. Tampilan Halaman Hasil

Pada halaman ini merupakan hasil ketika pengguna klik tombol prediksi sebagai hasil dari proses prediksi dan rekomendasi yang akan diberikan.



#### Hasil Prediksi

Prediksi Harga Rumah untuk daerah Bekasi dengan jumlah 2 kamar tidur, 1 kamar mandi, 1 garasi, 60 m<sup>2</sup> luas tanah, 40 m<sup>2</sup> luas bangunan

Rp. 527,479,500

#### Rekomendasi Rumah Berdasarkan Hasil Prediksi

- Tipe Uptown Villa di Uptown Estate Bekasi  
Rp. 589,000,000  
Bekasi □ 2 □ 2 □ 0 □ 60 □ 55  
[informasi lengkap](#)
- Rumah Smart Home Mewah Jatislo  
Rp. 580,000,000  
Bekasi □ 2 □ 1 □ 1 □ 60 □ 40  
[informasi lengkap](#)
- Rumah Murah Banget, Nyaman Baru Renovasi , Siap Pakai di Harapan Indah Bekasi  
Rp. 570,000,000  
Bekasi □ 2 □ 1 □ 0 □ 65 □ 45  
[informasi lengkap](#)
- Selangkah Puri Gading, Bebas Banjir, Akses 3 Mobil di Cluster Jatislo  
Rp. 560,000,000  
Bekasi □ 2 □ 1 □ 1 □ 60 □ 40  
[informasi lengkap](#)
- Rumah Mezzanine Design Exclusive Di Jatimulya Bekasi Dekat Tol dan LRT Bekasi Timur  
Rp. 530,000,000  
Bekasi □ 2 □ 1 □ 1 □ 60 □ 45  
[informasi lengkap](#)

[Selengkap](#)

Gambar 4. 28 Tampilan Halaman Hasil Prediksi

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

Berdasarkan Gambar 4.28 di atas merupakan tampilan hasil prediksi harga rumah. Pada halaman ini akan ditampilkan informasi kriteria yang dipilih beserta hasil prediksi harga rumah. Hasil tersebut akan diberikan 10 rekomendasi rumah berdasarkan prediksi harga rumah, rekomendasi ini akan disesuaikan dengan data yang digunakan pada pelatihan model.

#### 4.6.5. *Hosting Website*

Setelah membuat tampilan pada *website* langkah selanjutnya adalah *hosting website* agar dapat diakses oleh masyarakat secara *online*. Dalam publikasi *website* akan menggunakan server heroku yang menyediakan layanan *hosting* secara gratis yang dapat diakses seumur hidup. Ada beberapa hal yang harus dilakukan sebelum melakukan *hosting* di heroku sebagai berikut:

- Menginstall GitBash.
- Membuat akun heroku.
- Siapkan file Requirement.txt.

File *requirement* ini digunakan untuk menginstall semua *library* yang digunakan pada *source code website*.

```
click==8.0.3
colorama==0.4.4
Flask==2.0.2
Flask-SQLAlchemy==2.5.1
greenlet==1.1.2
gunicorn==20.1.0
itsdangerous==2.0.1
Jinja2==3.0.3
joblib==1.1.0
MarkupSafe==2.0.1
numpy==1.22.1
pandas==1.4.1
scikit-learn==1.0.2
scipy==1.7.3
```

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

```
sklearn==0.0
SQLAlchemy==1.4.31
threadpoolctl==3.1.0
Werkzeug==2.0.2
```

Berdasarkan *source code* di atas merupakan *library* yang digunakan pada proses pembuatan *website* prediksi ini, lalu setelah menulis *library* yang digunakan maka *save file* dalam bentuk txt dan masukkan ke dalam folder *project*.

- Siapkan file Procfile.

File ini digunakan untuk heroku dalam mengenal aplikasi yang akan di deploy dan bahasa yang digunakan.

```
web: gunicorn app:app
```

Berdasarkan *source code* di atas dapat diketahui bahwa aplikasi yang akan dibuat berupa aplikasi *website* dan menggunakan gunicorn untuk memberitahu kepada heroku bahwa yang bahasa yang digunakan adalah *Python* dan *source code* utama berada pada file *app.py*.

- Mempunyai akun GitHub (optional).
- Menginstall heroku pada GitBash (optional).

Pada saat melakukan *hosting* pada server heroku hal yang wajib di lakukan adalah menginstall GitBash pada laptop/pc/device yang digunakan, lalu membuat akun heroku bisa langsung menggunakan email atau membuat secara manual. Ada 3 cara untuk *hosting* aplikasi yang sudah dibuat yaitu menggunakan GitBash dengan menginstall heroku pada CMD GitBash, menghubungkan akun GitHub sehingga harus mempunyai akun GitHub. Pada penelitian ini akan menggunakan cara dengan menghubungkan heroku dengan akun GitHub peneliti sehingga dapat melakukan *hosting* aplikasi.

Langkah-langkah yang diperlukan untuk *hosting website* heroku menggunakan GitHub sebagai berikut:

1. Pertama-tama yang harus dilakukan adalah membuat akun GitHub. Setelah membuat akun masukkan akun yang sudah dibuat pada menu *Sign-in*.