

## ANALISIS PREDIKSI HARGA RUMAH DI JABODETABEK MENGUNAKAN MULTIPLE LINEAR REGRESSION

Linda Uswatun Hasanah<sup>1)</sup>, Inayatul Maula<sup>2)</sup>, Abu Tholib<sup>3)</sup>

<sup>123</sup>Fakultas Teknik Universitas Nurul Jadid

Jl. PP Nurul Jadid, Dusun Tj. Lor, Karanganyar, Kec. Paiton, Kab. Probolinggo,  
Jawa Timur

Email : [lindahasanah69@gmail.com](mailto:lindahasanah69@gmail.com), [inytl.maula18@gmail.com](mailto:inytl.maula18@gmail.com)<sup>2)</sup>,  
[ebuenje@gmail.com](mailto:ebuenje@gmail.com)<sup>3)</sup>

### ABSTRACT

The economic and infrastructure development in the Jabodetabek area has increased the rate of population growth in the region continues to increase. This attracted the attention of property agents in the housing price competition in the area, considering that the demand for housing will also continue to increase. There are several kinds of specificities and house prices in the Jabodetabek area, so an appropriate analysis is needed to get the desired house in this study the researcher implements the Multiple Linear method Regression to predict house prices with the desired specifications. This study used a total dataset of 3553 data with 11 variables obtained from the Kaggle.com. website The data will be through the pre-pronnce process of data before training the model. Furthermore, at the evaluation stage using a test matrix, namely CC (Coefficient of Correlation), MAE (Met Absolute Error), RMSE (Root Met Square Error) is used to assess the performance of the model. Based on the results of the analysis from the dataset were 965 data, where the data was divided into two parts, namely, 80% as training data and 20% as testing data from the analysis results were obtained at an accuracy rate of 0.85 or 85% with an error rate of MAE of 375428909.51 and the RMSE level of 516385.50.

**Keywords :** Price, Multiple Linear Regression, Prediction, House

### ABSTRAK

Perkembangan ekonomi dan infrastruktur di kawasan Jabodetabek membuat laju pertumbuhan penduduk di wilayah tersebut terus meningkat. Hal ini menarik perhatian para agen properti dalam persaingan harga rumah di kawasan tersebut, mengingat permintaan akan rumah juga terus meningkat. Ada beberapa macam spesifikasi dan harga rumah di kawasan Jabodetabek, sehingga diperlukan analisis yang tepat untuk mendapatkan rumah yang diinginkan Pada penelitian ini peneliti mengimplementasikan metode *Multiple Linear Regression* untuk memprediksi harga rumah dengan spesifikasi yang diinginkan. Penelitian ini menggunakan dataset sebanyak 3553 data dengan 11 variabel yang diperoleh dari *website* Kaggle.com. Data tersebut akan melalui proses *pre-processing* data sebelum dilakukannya *training* model. Selanjutnya pada tahap evaluasi menggunakan matrik pengujian, yaitu CC (*Coefficient of Correlation*), MAE (*Mean Absolute Error*), RMSE (*Root Mean Squared Error*) digunakan untuk menilai kinerja model. Berdasarkan hasil analisis dari dataset sebanyak 965 data, dimana data tersebut dibagi dua bagian yaitu, sebanyak 80% sebagai data *training* dan 20% sebagai data *testing* dari hasil analisis tersebut diperoleh tingkat akurasi sebesar 0,85 atau 85% dengan tingkat *error* MAE sebesar 375428909.51 dan tingkat *error* RMSE sebesar 515738165.50.

**Kata Kunci :** Harga, Multiple Linear Regression, Prediksi, Rumah

## 1. PENDAHULUAN

Jabodetabek atau Jakarta, Bogor, Depok, Tangerang, Bekasi merupakan kawasan metropolitan di Indonesia yang terdiri dari beberapa kota dan kabupaten di Provinsi Jawa Barat, Banten, dan DKI Jakarta. Saat ini, Jabodetabek menjadi daerah dengan tingkat kepadatan penduduk tertinggi di Indonesia. Hal ini disebabkan, perkembangan ekonomi di Jabodetabek yang semakin pesat dan infrastruktur yang semakin baik, sehingga banyak masyarakat yang bermigrasi ke wilayah ini dengan harapan untuk mendapatkan kehidupan juga fasilitas yang lebih baik [1]

Pasar properti di wilayah jabodetabek selalu menjadi perhatian bagi para investor dan pembeli properti, mengingat pertumbuhan jumlah penduduk yang terus bertambah di wilayah ini, yang berdampak pada peningkatan permintaan akan rumah [2]. Umumnya, calon pembeli memiliki beberapa kriteria untuk memutuskan melakukan pembelian rumah [3]. Baik dari segi kondisi bangunan, luas tanah, luas bangunan, serta fasilitas [4]. Namun, dengan tingginya permintaan rumah di wilayah ini menyebabkan harga rumah menjadi aspek penting yang perlu diperhatikan dalam membeli sebuah rumah impian [5]. Sehubungan dengan hal tersebut maka dibutuhkanlah sebuah sistem keputusan yang dapat membantu calon pembeli dalam membuat keputusan pembelian rumah yang optimal berdasarkan kriteria dan harga yang telah ditentukan sebelumnya. Salah satu metode yang dapat digunakan dalam melakukan analisis prediksi harga rumah adalah metode *Multiple Linear Regression*.

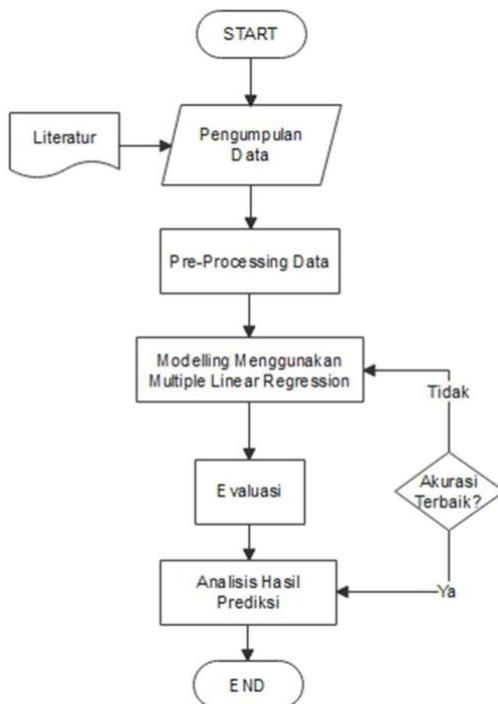
Metode *Multiple Linear Regression* diimplementasikan dalam proses analisis prediksi harga rumah karena metode ini dapat mengestimasi koefisien regresi yang mewakili pengaruh variabel independen terhadap harga rumah, sehingga model dapat digunakan untuk memprediksi harga rumah

baru berdasarkan nilai-nilai variabel independen yang diberikan [6]. sehingga penggunaan metode ini lebih tepat diimplementasikan pada analisis prediksi harga rumah pada kawasan Jabodetabek. Beberapa penelitian terkait metode *Multiple Linear Regression* sebagai sistem penunjang keputusan dilakukan oleh peneliti terdahulu. Penelitian [7] dilakukan tentang prediksi harga kamar kost menggunakan metode *Multiple Linear Regression*. Penelitian ini menggunakan beberapa variabel independen yang berpengaruh dalam proses prediksi harga kamar kost yaitu air panas(1), luas kamar(2), fasilitas dapur(3), internet(4), AC(5), kamar mandi dalam(6), laundry(7) dan kost khusus(8). Hasil dari penelitian ini, ditemukan bahwa model memiliki koefisien korelasi sebesar 0,70. Penelitian [5] dilakukan tentang prediksi harga rumah berdasarkan spesifikasi dengan menggunakan metode *Multiple Linear Regression*. Penelitian ini menggunakan beberapa variabel independen yang berpengaruh dalam proses prediksi harga rumah yaitu jumlah kamar(1), jumlah kamar mandi(2), luas tanah(3), luas bangunan(4), garasi(5) dan kota(6). Hasil dari penelitian ini, ditemukan bahwa model memiliki tingkat akurasi sebesar 66%. Penelitian [6] dilakukan tentang penggunaan *machine learning* untuk memprediksi harga sembako dengan metode regresi linier berganda. Penelitian ini menggunakan beberapa variabel independen yang berpengaruh dalam proses prediksi harga rumah yaitu tanggal(1), komoditas(2) dan pasar(3). Hasil dari penelitian ini, ditemukan bahwa variabel independen memiliki pengaruh terhadap variabel dependen sebesar 84,2%. Penelitian [8] dilakukan tentang prediksi pengeluaran desa pada sistem keuangan desa menggunakan metode *Multiple Linear Regression*. Penelitian menggunakan 2 variabel independen yaitu dana desa dan alokasi dana desa. Hasil dari penelitian ini, ditemukan bahwa model memiliki koefisien korelasi sebesar 0,89.

Pada penelitian ini peneliti mengimplementasikan metode *Multiple Linear Regression* untuk membantu memprediksi kisaran harga dalam membeli rumah di kawasan Jabodetabek. Proses analisis pada penelitian ini menggunakan 9 variabel independen yaitu lokasi, luas lahan, luas bangunan, jumlah kamar, jumlah kamar mandi, lantai, jumlah kamar utama, jumlah kamar mandi utama dan garasi. Sedangkan variabel dependen atau variabel terikat adalah harga rumah. Hasil dari penelitian ini diharapkan dapat meminimalisir risiko terjadinya kesalahan dalam mengambil keputusan untuk membeli sebuah rumah impian di kawasan Jabodetabek [9].

## 2. METODOLOGI

Pada penelitian ini dilakukan beberapa tahapan-tahapan dalam mengimplementasikan metode *Multiple Linear Regression* dalam proses analisis prediksi harga rumah. Tahapan proses pada penelitian ini dimulai dengan pengumpulan data, *pre-processing* data, *modelling* dan evaluasi.



Gambar 2.1. Tahapan penelitian

### 2.1. Pengumpulan Data

Pada penelitian ini dataset di peroleh dari website [kaggle.com](https://www.kaggle.com) [10], dataset yang digunakan merupakan data daftar harga rumah pada daerah Jabodetabek yang diperoleh dari website rumah123.com pada tahun 2022. Dataset tersebut terdiri dari 3553 *row* dan 27 *column*. Dataset yang digunakan dalam penelitian ini mengandung informasi mengenai harga rumah, kota, luas bangunan, luas lahan, jumlah kamar, jumlah kamar mandi, jumlah kamar pembantu, jumlah kamar mandi pembantu, jumlah lantai, garasi, dan tahun pembangunan. Proses selanjutnya yakni mengumpulkan literatur dari beberapa jurnal dan buku elektronik yang erat kaitannya dengan *Multiple Linear Regression*.

### 2.2. Pre-Processing Data

*Pre-processing* merupakan proses yang digunakan untuk membersihkan, mentransformasi, dan mempersiapkan data sebelum diolah lebih lanjut. *Pre-processing* data dilakukan untuk meningkatkan kualitas data, menghilangkan nilai yang tidak lengkap atau tidak valid, mengubah format data, dan menyesuaikan data agar sesuai dengan format yang dibutuhkan oleh model atau algoritma yang akan digunakan [11].

### 2.3. Modelling menggunakan Multiple Linear Regression

*Multiple Linear Regression* merupakan metode statistik yang digunakan untuk menganalisis hubungan linear antara satu atau lebih variabel independen (*x*) dan variabel dependen (*y*) [12]. Variabel independen merupakan faktor yang mempengaruhi, sedangkan variabel dependen merupakan variabel yang dipengaruhi [13]. Persamaan multiple linear regression dapat dituliskan sebagai berikut:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n + \varepsilon$$

[14] (1)

Keterangan:

*y* = variabel dependen

*b* = koefisien regresi

$x$  = variabel independen  
 $\varepsilon$  = error atau deviasi dari nilai prediksi

Data yang telah melalui proses *pre-processing* akan digunakan sebagai *input* model pada algoritma *Multiple Linear Regression* [15], kemudian data tersebut dibagi menjadi 80% data *training* dan 20% data *testing*.

#### 2.4. Evaluasi

Tahap evaluasi model prediksi dilakukan dengan menggunakan metrik pengujian, yaitu CC (*Coefficient of Correlation*), MAE (*Mean Absolute Error*), RMSE (*Root Mean Squared Error*). Pada tahap ini dilakukan penarikan kesimpulan dari hasil data *mining*.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini sebanyak 3553 data dengan 10 variabel. Variabel yang ada dalam dataset ini mencakup informasi penting mengenai harga rumah, kota, luas bangunan, luas lahan, jumlah kamar tidur, jumlah kamar mandi, jumlah kamar pembantu, jumlah kamar mandi pembantu, jumlah lantai, jumlah garasi, dan tahun pembangunan rumah.

Harga rumah merupakan variabel yang ingin diprediksi, sedangkan variabel kota, luas bangunan, luas lahan, jumlah kamar tidur, jumlah kamar mandi, jumlah kamar pembantu, jumlah kamar mandi pembantu, jumlah lantai, garasi, dan tahun pembangunan rumah digunakan sebagai variabel independen dalam analisis.

#### 3.2. Pre-Processing Data

Pada tahap selanjutnya, dilakukan *data cleaning* untuk memastikan kualitas data yang akan digunakan. Pada tahap ini dilakukan pemeriksaan *missing value*,

dalam dataset tersebut terdapat beberapa data yang bernilai *null (missing value)*.

```

harga          0
kota           0
kamar_tidur    34
kamar_mandi    29
luas_lahan     2
luas_bangunan  2
kamar_pembantu 0
kamar_mandi_pembantu 0
lantai         6
tahun          1445
garasi         0
dtype: int64
    
```

Gambar 3.1. *missing value*

```

harga          0
kota           0
kamar_tidur    0
kamar_mandi    0
luas_lahan     0
luas_bangunan  0
kamar_pembantu 0
kamar_mandi_pembantu 0
lantai         0
tahun          0
garasi         0
dtype: int64
    
```

Gambar 3.2 Setelah penanganan *missing value*

Pada gambar 3.1 ditunjukkan bahwa variabel kamar tidur, kamar mandi, luas lahan, luas bangunan, lantai dan tahun mengandung beberapa nilai yang hilang (*missing value*). Dalam penanganan *missing value* dapat dilakukan dengan mengisi nilai variabel yang hilang dengan menggunakan rata-rata dari variabel tersebut. Pada gambar 3.2 menunjukkan dataset yang telah dilakukan penanganan *missing value*.

Selanjutnya periksa tipe data hal ini dilakukan karena model *Multiple Linear Regression* hanya dapat memproses data numerik. Oleh karena itu, variabel kota yang semula memiliki tipe data objek perlu diubah menjadi tipe data numerik sebelum digunakan dalam proses analisis. Untuk melakukan itu, dapat menggunakan kode

angka sebagai pengganti nilai-nilai kota. Berikut adalah keterangan nilai-nilai kota yang digunakan:: Bekasi

- 1 : Bogor
- 2 : Depok
- 3 : Jakarta Barat
- 4 : Jakarta Selatan
- 6 : Jakarta Utara
- 7 : Jakarta Pusat
- 8 : Jakarta Timur
- 9 : Tangerang

Dengan mengubah nilai dari variabel “kota” dengan kode angka di atas, maka variabel “kota” akan berubah menjadi tipe data numerik yang dapat diproses oleh model *Multiple Linear Regression*. Perubahan tipe data pada variable “kota” dapat dilihat pada tabel 3.1 berikut.

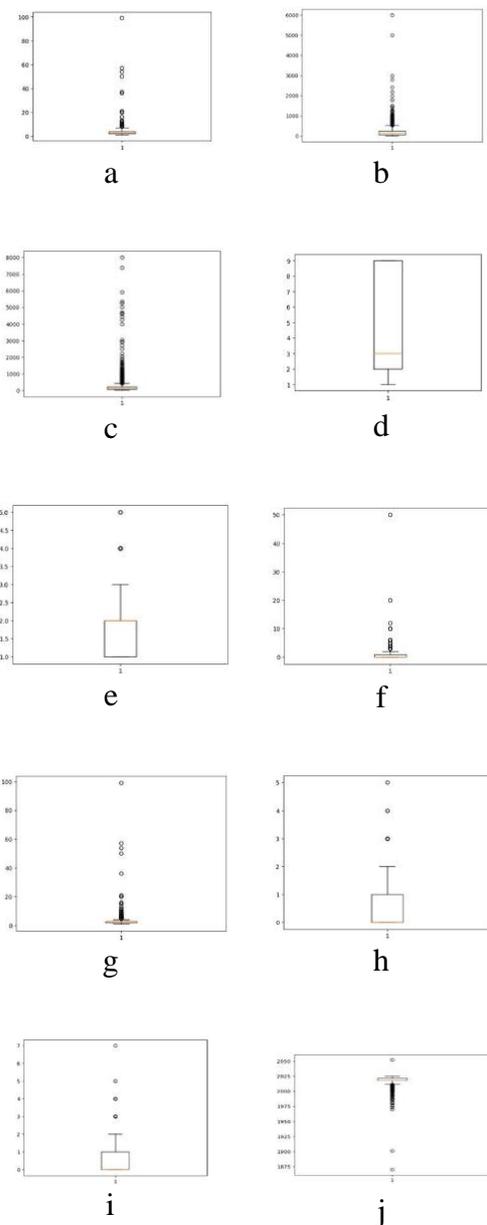
Tabel 3.1. Informasi Tipe Data

#	Column	Non-Null Count	Dtype
0	Harga	3553 non-null	Float64
1	Kota	3553 non-null	Int64
2	Kamar tidur	3553 non-null	Int32
3	Kama mandi	3553 non-null	Int32
4	Luas lahan	3553 non-null	Float64
5	Luas bangunan	3553 non-null	Float64
6	Kamar pembantu	3553 non-null	Int64
7	Kamar Mandi pembantu	3553 non-null	Int64
8	Lantai	3553 non-null	Int32
9	Tahun	3553 non-null	Int32
10	Garasi	3553 non-null	Int32

Tahap selanjutnya pengecekan data duplikat, jika terdapat baris data yang memiliki nilai yang sama, maka baris data tersebut diidentifikasi sebagai data duplikat. Dari hasil pengecekan terdapat sebanyak

711 rows data duplikat pada penelitian ini. Data duplikat perlu dihapus dari dataset untuk menjaga kebersihan dan kualitas data harga rumah.

Proses selanjutnya dalam *cleaning data* ialah identifikasi *oulier*. *Outlier* adalah nilai yang berbeda secara signifikan dari pola umumnya. Persebaran outlier dapat diidentifikasi dengan menggunakan boxplot, scatter plot atau histogram.



Gambar 3.3 Distribusi a)Kamar tidur b)Luas bangunan c)Luas lahan d)Kota e)Lantai f)Garasi g)Kamar mandi h)Kamar mandi pembantu i)Kamar pembantu j)Tahun

Berdasarkan visualisasi penyebaran suatu variabel diketahui bahwa variabel Kota berdistribusi secara normal sedangkan variabel luas bangunan, luas lahan, kamar tidur, kamar mandi, kamar pembantu, kamar mandi pembantu, lantai, garasi, dan tahun mengalami *outlier*. Penentuan *outlier* dapat dilakukan dengan metode Z-Score yaitu batas *outlier* berdasarkan jumlah standar deviasi antara nilai observasi dan rata-rata. Setelah batas *outlier* ditentukan maka nilai yang melebihi batas tersebut akan dihapus secara permanen dari dataset.

Setelah dilakukan data *cleaning* selanjutnya analisis deskriptif terhadap dataset harga rumah Jobodetabek. Analisis deskriptif dapat dilihat pada tabel 2 berikut.

**Tabel 3.2. Deskripsi dataset**

	Harga	Kota	Kamar tidur	Kamar mandi	Luas lahan
count	965	965	965	965	965
mean	1.522.008.000	4.26	2.86	2.18	87.67
Std	1.476.453.000	3.06	0.83	0.98	38.37
min	70.000.000	1.0	1.0	1.0	15.0
25%	600.000.000	2.0	2.0	1.0	61.0
50%	970.000.000	3.0	3.0	2.0	75.0
75%	1.790.000.000	8.0	3.0	3.0	105.0
max	8.820.000.000	9.0	6.0	6.0	224.0

Dalam tabel 3.2 diatas dapat

	Luas Bangunan	Kamar Pembantu	Kamar mandi Pembantu	Lantai	Gara si	Thn
count	965	965	965	965	965	965
mean	94.98	0.24	0.24	1.73	0.61	2021
Std	60.01	0.44	0.42	0.58	0.82	0.84
min	19.0	0.0	0.0	1.0	0.0	2019
25%	50.0	0.0	0.0	1.0	0.0	2021
50%	75.0	0.0	0.0	2.0	0.0	2022
75%	126.0	0.0	0.0	2.0	1.0	2022
max	340.0	2.0	2.0	3.0	4.0	2023

disimpulkan bahwa dataset tersebut tampak lengkap karena tidak terdapat *missing values* pada setiap kolom variabel. Selain itu, tidak terdapat *outliers* yang terdeteksi. Ini mengindikasikan bahwa tidak ada nilai yang ekstrem yang dapat mempengaruhi hasil analisis statistik. Dengan demikian, analisis

lebih lanjut dapat dilakukan untuk melihat hubungan antara variabel satu dengan variabel lainnya.

### 3.3 Matrik Korelasi

Matriks korelasi adalah sebuah tabel yang menggambarkan hubungan antara dua atau lebih variabel dalam suatu data. Setiap sel dalam matriks korelasi berisi koefisien korelasi antara dua variabel yang diukur dalam rentang -1 hingga 1. Nilai positif menunjukkan hubungan positif, sedangkan nilai negatif menunjukkan hubungan negatif. Semakin dekat nilai dengan 1 atau -1, semakin kuat hubungan antara variabel-variabel tersebut. Nilai 0 menunjukkan tidak adanya hubungan linier antara variabel-variabel tersebut.

	harga	kota	kamar_tidur	kamar_mandi	luas_lahan	luas_bangunan	kamar_pembantu	kamar_mandi_pembantu	lantai	tahun	garasi
harga	1.00	0.49	0.63	0.67	0.76	0.86	0.59	0.66	0.49	-0.13	0.15
kota	0.49	1.00	0.34	0.39	0.23	0.38	0.28	0.32	0.37	-0.09	0.09
kamar_tidur	0.63	0.34	1.00	0.78	0.57	0.74	0.44	0.47	0.56	-0.06	0.25
kamar_mandi	0.67	0.39	0.78	1.00	0.54	0.78	0.53	0.54	0.67	-0.08	0.21
luas_lahan	0.76	0.23	0.57	0.54	1.00	0.77	0.56	0.58	0.25	-0.26	0.14
luas_bangunan	0.86	0.38	0.74	0.78	0.77	1.00	0.65	0.67	0.59	-0.16	0.16
kamar_pembantu	0.59	0.28	0.44	0.53	0.56	0.65	1.00	0.88	0.36	-0.13	0.11
kamar_mandi_pembantu	0.66	0.32	0.47	0.54	0.58	0.67	0.88	1.00	0.37	-0.11	0.12
lantai	0.49	0.37	0.56	0.67	0.25	0.59	0.36	0.37	1.00	0.00	0.14
tahun	-0.13	-0.09	-0.06	-0.08	-0.26	-0.16	-0.13	-0.11	0.00	1.00	0.05
garasi	0.15	0.09	0.25	0.21	0.14	0.16	0.11	0.12	0.14	0.05	1.00

Gambar 3.4. Matrik korelasi

Pada gambar 3.4 dapat dilihat tabel korelasi yang menunjukkan hubungan antara variabel satu dengan variabel lainnya. Dari tabel tersebut, terlihat bahwa variabel yang paling berpengaruh terhadap variabel harga adalah luas bangunan. Hal ini menunjukkan bahwa semakin besar luas bangunan rumah, maka harga rumah juga semakin meningkat. Selain itu, dari data tabel tersebut juga terlihat bahwa variabel tahun tidak memiliki korelasi yang berpengaruh terhadap harga dalam dataset tersebut.

### 3.4 Multiple Linear Regression

Setelah melalui proses data *cleaning*, terdapat total 965 data yang siap untuk dianalisis. Dataset tersebut akan dibagi menjadi data *training* dan data *testing*, dengan 80% data sebagai data *training* dan 20% data sebagai data *testing*. Dalam pemodelan ini, menggunakan 10 variabel dependen (x) yaitu kota, kamar tidur, kamar mandi, luas tanah, luas bangunan, kamar pembantu, kamar mandi pembantu, lantai, tahun dan garasi. Variabel independent (y) adalah harga pada setiap rumah.

Berdasarkan persamaan (1) Maka, Model Persamaan Multiple Linear Regression pada penelitian ini sebagai berikut.

$$y = -187908165997.4 + 100528235.1688x_1 - 96430701.7285x_2 + 6237491.2411x_3 + 11571599.0847x_4 + 12906292.8436x_5 - 378712189.3596x_6 + 629716003.4461x_7 + 57694109.536x_8 + 92432882.1724x_9 - 5356150.7447x_{10}$$

Dengan:

y = Harga sebagai variabel dependen

x<sub>1</sub> = Kota

x<sub>2</sub> = Kamar Tidur

x<sub>3</sub> = Kamar Mandi

x<sub>4</sub> = Luas Lahan

x<sub>5</sub> = Luas Bangunan

x<sub>6</sub> = Kamar Pembantu

x<sub>7</sub> = Kamar Mandi Pembantu

x<sub>8</sub> = Lantai

x<sub>9</sub> = Tahun

x<sub>10</sub> = Garasi

dengan evaluasi akurasi prediksi :

Tabel 3. Akurasi

CC	MAE	RMSE
0.85235	375428909.51	515738165.50

Berdasarkan Tabel 3 diatas diperoleh nilai *Correlation Coefficient* (CC)

sebesar 0.85235. Hal ini menunjukkan adanya hubungan linier positif yang kuat antara variabel-variabel yang digunakan dalam model regresi. Selanjutnya, *Mean Absolute Error* (MAE) memiliki nilai sebesar 375428909.51. Dalam hal ini, nilai MAE yang cukup besar menunjukkan bahwa rata-rata kesalahan absolut antara nilai prediksi dan nilai sebenarnya cukup tinggi. Sedangkan nilai *Root Mean Squared Error* (RMSE) sebesar 515738165.50. menunjukkan bahwa terdapat variasi yang signifikan antara nilai prediksi dan nilai sebenarnya.

Selanjutnya kami mencoba memprediksi harga dari suatu rumah berdasarkan spesifikasi berikut.

1. Kota = 1 (Bekasi)
2. Kamar tidur = 4
3. Kamar mandi = 2
4. Luas lahan = 200 m<sup>2</sup>
5. Luas bangunan = 150 m<sup>2</sup>
6. Kamar pembantu = 0
7. Kamar mandi pembantu = 0
8. Lantai = 1
9. Tahun = 2018
10. Garasi = 1

```
In [68]: reg.predict([[1,4,2,200,150,0,0,1,2018,1]])
Out[68]: array([2.65127234e+09])
```

Gambar 6. Hasil prediksi

Harga prediksi dari spesifikasi diatas adalah 2.651.272.340 (2.6 Milyar Rupiah ). Dilihat dari daftar harga rumah diwilayah Bekasi dengan luas tanah 200 m<sup>2</sup> dan luas bangunan 150m<sup>2</sup> rentang harganya mencapai 2 - 3 Miliar Rupiah. Jika dibandingkan dengan harga prediksi sebesar 2.6 Milyar Rupiah, dapat disimpulkan bahwa prediksi harga rumah ini cukup akurat. Hal ini menunjukkan bahwa model *Multiple Linear Regression* yang digunakan memiliki tingkat akurasi yang tinggi.

#### 4. Kesimpulan

Berdasarkan dari hasil analisis, dapat disimpulkan bahwa metode *Multiple Linear Regression* dapat digunakan untuk memprediksi harga rumah di wilayah Jabodetabek dengan tingkat akurasi yang cukup baik yakni sebesar 85%. Variabel independen seperti jumlah kamar, jumlah kamar mandi, luas tanah, luas bangunan, garasi dan kota menjadi faktor yang memiliki pengaruh signifikan terhadap harga rumah di wilayah Jabodetabek. Namun, terdapat tingkat kesalahan prediksi yang signifikan dengan MAE dan RMSE yang cukup tinggi. Hal ini menandakan bahwa model tersebut belum mampu menghasilkan prediksi yang sangat akurat.

#### 5. SARAN

Penulis menyadari bahwa masih banyak kekurangan dalam penelitian ini, untuk itu penulis akan melengkapi penelitian ini dengan beberapa saran. Untuk meningkatkan akurasi dalam melakukan prediksi harga rumah diwilayah jabodetabek, penulis menyarankan agar dapat menggunakan metode lainnya di penelitian berikutnya. Selanjutnya penulis menyarankan untuk referensi dataset harga rumah kedepannya dapat ambil dari web crawling.

#### DAFTAR PUSTAKA

- [1] A. D. Prastiwi, 'URBAN HEAT ISLAND DI KOTA TANGERANG SELATAN', *Jurnal Geosaintek*, vol. 8, no. 2, p. 182, Aug. 2022, doi: 10.12962/j25023659.v8i2.11721.
- [2] M. D. Rachmansyah, T. Ekowati, and W. Dyah Prastiwi, 'Tekanan Penduduk dan Daya Dukung Lahan Padi Gogo di Kecamatan Mustikajaya Kota Bekasi', *Forum Agribisnis*, vol. 12, no. 2, pp. 88-97, Sep. 2022, doi: 10.29244/fagb.12.2.88-97.
- [3] A. Aljufri, A. A. Meidhani, H. Adriel, and R. W. Limas, 'Faktor yang Mempengaruhi Niat Membeli rumah', 2021.

- [4] Z. Faradilla Daldiri -, M. Rafly -, and I. Veritawati -, 'Clustering Daftar Harga Rumah di Jakarta Dengan Algoritma K-Means', *Journal of Informatics and Advanced Computing (JIAC)*, vol. 3, no. 2, 2022, [Online]. Available: <https://www.kaggle.com/datasets/wisnuanggara/daf>
- [5] M. Labib Mu'tashim, S. A. Damayanti, H. N. Zaki, T. Muhayat, and R. Wirawan, 'Analisis Prediksi Harga Rumah Sesuai Spesifikasi Menggunakan Multiple Linear Regression', vol. 3, p. 2021.
- [6] K. Puteri and A. Silvanie, 'MACHINE LEARNING UNTUK MODEL PREDIKSI HARGA SEMBAKO DENGAN METODE REGRESI LINIER BERGANDA', 2020. [Online]. Available: [www.data.jakarta.go.id](http://www.data.jakarta.go.id).
- [7] M. Reza Fahlepi, A. Widjaja, and J. Surya Sumantri No, 'Penerapan Metode Multiple Linear Regression Untuk Prediksi Harga Sewa Kamar Kost', 2019.
- [8] H. T. Siregar, N. H. Harani, C. Prianto, and A. Bachelor, 'IMPLEMENTATION OF MULTIPLE LINEAR REGRESSION METHODS AS PREDICTION OF VILLAGE SPENDING ON VILLAGE FINANCIAL MANAGEMENT SYSTEM', *Jurnal Ilmiah KURSOR*, vol. 10, no. 2, pp. 53–64, 2019.
- [9] P. Kurnia Putri and I. Mahendra, 'IMPLEMENTASI METODE ANALYTICAL HIERARCHY PROCESS (AHP) DALAM SISTEM PENDUKUNG KEPUTUSAN PEMBELIAN RUMAH DI KOTA TANGERANG', 2019.
- [10] Nafis Barizki, 'Daftar Harga Rumah Jabodetabek', *kaggle.com*, 2022. <https://www.kaggle.com/datasets/nafisbarizki/daftar-harga-rumah-jabodetabek> (accessed May 08, 2023).
- [11] mie.binus.ac.id, 'Teknik pre-processing dan classification dalam data science', *mie.binus.ac.id*, Aug. 26, 2022. <https://mie.binus.ac.id/2022/08/26/teknik-pre-processing-dan-classification-dalam-data-science/> (accessed May 11, 2023).
- [12] A. Novebrian Maharadja, I. Maulana, and B. Arif Dermawan, 'Penerapan Metode Regresi Linear Berganda untuk Prediksi Kerugian Negara Berdasarkan Kasus Tindak Pidana Korupsi', 2021. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [13] C. Haryanto, N. Rahaningsih, and F. Muhammad Basysyar, 'KOMPARASI ALGORITMA MACHINE LEARNING DALAM MEMREDIKSI HARGA RUMAH', 2023.
- [14] R. Paninggalih, B. Nugroho, and M. I. Takaendengan, 'Prediksi Saldo Produksi Hasil Ternak Kabupaten Blitar Menggunakan Regresi Linier Berganda', 2023. [Online]. Available: [www.data.go.id](http://www.data.go.id).
- [15] S. Bramasto, D. Khairiani, J. Raya, P. Serpong, and T. Selatan, 'Prediksi Daya Output Sistem Pembangkit Listrik Tenaga Surya (PLTS) Menggunakan Regresi Linear Berganda', vol. 15, no. 3, pp. 1979–276, 2022, doi: 10.30998/faktorexacta.v15i2.13254.