

## Analisis Prediksi Harga Rumah Sesuai Spesifikasi Menggunakan *Multiple Linear Regression*

Muhammad Labib Mu'tashim<sup>1</sup>, Sekar Ayu Damayanti<sup>2</sup>, Hanan Nadia Zaki<sup>3</sup>, Toni Muhayat<sup>4</sup>, Rio Wirawan<sup>5</sup>

S1 Informatika / Fakultas Ilmu Komputer

Universitas Pembangunan Nasional Veteran Jakarta

Jl. RS. Fatmawati Raya, Pd. Labu, Kec. Cilandak, Kota Depok, Daerah Khusus Ibukota Jakarta

<sup>1</sup> muhammadlm@upnvj.ac.id, <sup>2</sup> sekarayu@upnvj.ac.id, <sup>3</sup> hanannz@upnvj.ac.id, <sup>4</sup> tonim@upnvj.ac.id,

<sup>5</sup> rio.wirawan@upnvj.ac.id

**Abstrak.** Harga rumah khususnya di kawasan ibukota DKI Jakarta merupakan suatu wadah sebagai investasi bagi para pencari rumah jaman sekarang. Berbagai rumah yang terdapat di wilayah DKI Jakarta, terutama di Jakarta Selatan memiliki banyak perbedaan spesifikasi. Spesifikasi itu yang berbeda baik dari harga, luas, dan jumlah ruangan yang terdapat di rumah tersebut. Sehingga, terjadinya persaingan harga diantara agen-agen properti. Selain itu harga rumah setiap tahun akan semakin mahal, sehingga dibutuhkan analisis serta pemilihan yang tepat yang harus dilakukan untuk membeli sebuah rumah. Dengan menggunakan algoritma regresi dengan menggunakan faktor-faktor dari harga rumah, luas bangunan, luas tanah, ruang kamar tidur, garasi, dan kota. Sehingga, hasil penelitian yang kami dapatkan dengan menggunakan algoritma regresi dengan nilai akurasi sebesar 66%. Demikian dari hasil penelitian kami ditemukan bahwa harga rumah mengalami peningkatan. Diharapkan semoga penelitian ini berguna bagi para pembeli ataupun agen properti dalam melakukan prediksi penjualan rumah.

**Kata Kunci :** Rumah, Spesifikasi, Harga, Regresi

### 1. Pendahuluan

Rumah merupakan bangunan perihal primer yang diperlukan oleh makhluk hidup buat menetap serta bertempat tinggal. Terlepas dari itu, rumah tinggal juga wajib mempunyai nilai serta mempunyai ukuran, keadaan, kondisi, dan paling utama adalah aspek ekonomi yang mempunyai investasi masa panjang yakni dimensi buat mengenali bagaimana kehidupan seorang. Kehidupan di wilayah ibu kota pasti mempunyai harga investasi yang cukup tinggi dan pastinya menjanjikan semua kalangan, dikarenakan harga tanah, rumah, maupun fasilitas umum sangat memadai mampu menjadikan harga setiap tahun semakin meningkat.

Harga setiap tahu rumah semakin meningkat dan mahal, peningkatan harga ini dapat diukur melalui beberapa aspek ataupun faktor pendukung perihal harga yang ditawarkan. Harga tidak selalu pasti dan terprediksi dengan akurat membuat pembeli rumah butuh sistem yang bisa memperkirakan harga rumah berdasarkan faktor pendukungnya [1]. Dikarenakan harga rumah tiap tahun berbeda sehingga kami melakukan penelitian menggunakan algoritma regresi dalam memperhitungkan spesifikasi harga rumah dengan menggunakan *Multiple linear regression* dengan menganalisis data untuk membangun sebuah model *Machine Learning* dengan tujuan memprediksi harga rumah.

### 2. Landasan Teori

#### 2.1 Rumah

Umumnya rumah bisa diartikan sebagai tempat berlindung ataupun tempat bernaung suatu keluarga dari segala macam kondisi diluar sana, seperti panasnya matahari, badai, hujan, hingga dinginnya malam, Serta sebagai pencerminan diri pribadi dalam upaya peningkatan taraf hidup, serta pembentukan watak, karakter dan kepribadian bangsa [2].

Kalau dilihat berdasarkan fisiknya, rumah merupakan suatu bangunan atau tempat bagi keluarga untuk berkumpul, bekerja, sebagai juga tempat tidur dan istirahat untuk memulihkan Kembali kondisi fisik dan mental setelah melaksanakan tugasnya sehari-hari masing-masing, bagi para penghuni rumah tersebut. Hal ini juga

diperkuat oleh Tuner [3] yang menyatakan bahwa rumah memiliki fungsi utama, yakni penunjang rasa aman, kesempatan, dan identitas keluarga. Namun bila dilihat dari aspek psikologinya, rumah adalah sebuah tempat yang membuat keluarga tentram, aman, serta menyenangkan bagi seluruh penghuninya.

Penelitian [4] menggunakan UCI Machine Learning. Karakteristik data terdiri dari 7 variabel. 6 variabel yang dijadikan variabel X atau disebut variabel dependen dan 1 variabel yang dijadikan variabel Y atau variabel independen. Implementasinya berupa sistem prediksi harga rumah diharapkan mampu memberikan informasi harga rumah yang sesuai dengan keadaan yang diharapkan.

## 2.2 Harga Rumah

Penentuan harga rumah yang harus diperhitungkan baik kepada penjual maupun pembeli. Penentuan harga ini berdasarkan luas bangunan dan luas tanah yang menjadi pokok dalam penjual rumah, selain dari itu harga pajak dari rumah tersebut atau bisa disebut Nilai Jual Objek Pajak (NJOP) berbeda-beda tiap daerah sehingga harga pajak tersebut menjadi salah satu hal penting dalam pembelian rumah, serta apabila harga NJOP besar dapat menjadi investasi untuk para pemilik rumah apabila ingin menjual rumahnya.

Selain cek harga tanah, ada beberapa faktor lain yang menentukan harga jual rumah [5]. Pembeli harus memeriksa pada spesifikasi bangunan yang berupa luas bangunan, jumlah kamar tidur, jumlah kamar mandi, serta kondisi rumah yang akan pembeli huni. Kemudian, melakukan *cross check* terhadap kelebihan dan kekurangan terhadap properti rumah tersebut. Seperti area plafon ataupun rangka ruma. Lalu, cek apakah rumah yang akan di bel mengalami kerusakan dan perlu perbaikan. Setelahnya melakukan perbandingan harga rumah dengan harga pasaran dan memperhatikan dimana lokasi hunian yang akan ditinggali apabila lokasi rumah tersebut semakin strategis dan dekat dengan berbagai area serta memiliki akses yang mudah, sehingga semakin bagus kualitas rumah semakin tinggi pula harga rumah [6].

## 2.3 Multiple Linear Regression

*Multiple Linear Regression*, yakni model regresi berganda yang variabelnya lebih dari satu yang dimana pemodelan regresi linear berganda melakukan prediksi dengan menggunakan data skala interval atau rasio serta terdapat lebih dari satu prediksi. Selain itu metode ini melakukan analisis dari adanya variable bebas dan variable terikat [7].

Model regresi linear berganda bisa digambarkan dengan persamaan sebagai berikut:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_n X_n + e$$

Keterangan:

Y = Variabel terikat (Dependen)

X = Variabel bebas (Independen)

$\alpha$  = Konstanta (*Intercept*)

$\beta$  = Slope atau Koefisien estimate.

e = Error

## 3. Metode Penelitian

### 3.1 Pengumpulan Data

Penelitian pada jurnal ini mengambil dataset dari *kaggle.com* [8], yang banyak berisikan atribut-atribut yang diperlukan, yakni data harga rumah yang berada di Jakarta Selatan. Kemudian proses selanjutnya adalah mempelajari dan mengumpulkan literatur yang berkaitan erat dengan *Multiple Linear Regression*, yang sumbernya bisa dari internet maupun buku-buku tambahan.

### 3.2 Pengolahan Data

Data harga rumah yang telah diambil dari *kaggle.com* yaitu berisi atribut Harga, Luas Tanah, Luas Bangunan, jumlah Kamar Tidur, jumlah Kamar Mandi, ada/ tidaknya Garasi, serta nama Kota. Data yang sudah diambil selanjutnya masuk kedalam pre-processing data, atau dimana data yang belum lengkap, berisikan NaN, ataupun data yang tidak stabil diseleksi dan dihapus. Lalu pada relevation data, data yang tidak dibutuhkan akan dihapus juga dan hanya diambil data yang relevan dengan penelitian ini. Dalam penelitian ini tidak ada atribut yang dihapus karena semua atribut yang ada relevan dan dapat digunakan.



**Gambar. 1.** Bagan Alur

Pada Gambar 1, penelitian akan dimulai dengan studi Literatur terlebih dulu, yakni kami mencari referensi dan metode terdekat yang sesuai dengan penelitian kami, yakni terfokus pada prediksi harga rumah. Kemudian pada pengumpulan data, kami mencari data dari website *Kaggle.com*, dan lalu melakukan pra-processing data sehingga data sudah bersih. Data tersebut akan di lanjutkan ke proses perhitungan menggunakan Multiple Linear Progression, disini awalnya data akan dibagi 2, yakni data testing atau data uji dan data training atau data latih. Setelah hal tersebut dilakukan, baru bisa mengetahui prediksi Harga Rumah menggunakan akurasi dan spesifikasi yang tersedia [9].

## 4. Pembahasan dan Hasil

### 4.1 Pra-proses Data

Terdapat total 1001 baris data dan 7 kolom. Data tersebut sudah dibersihkan dan bisa dilanjutkan ke tahap berikutnya.

**Tabel 2.** Data yang sudah di Clean

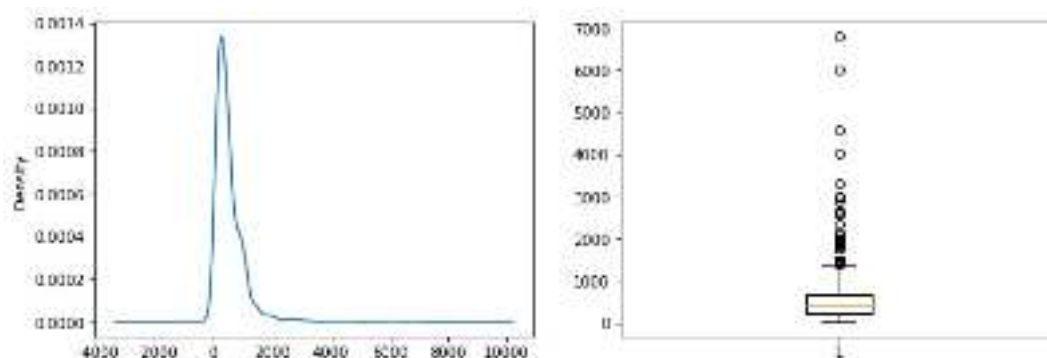
Index	HARGA	LT	LB	JKT	JKM	GRS	KOTA
0	28000000000	1100	700	5	6	1	1
1	19000000000	824	800	4	4	1	1
2	4700000000	500	400	4	3	1	1
3	4900000000	251	300	5	4	1	1
4	28000000000	1340	575	4	5	1	1
5	10000000000	460	300	4	4	1	1

Penjelasan setiap kolom :

1. Harga : Harga setiap rumah
2. LT : Luas Tanah
3. LB : Luas Bangunan
4. JKT : Jumlah Kamar Tidur
5. JKM : Jumlah Kamar Mandi
6. GRS : Garasi
7. KOTA : Kota (Jakarta Selatan)

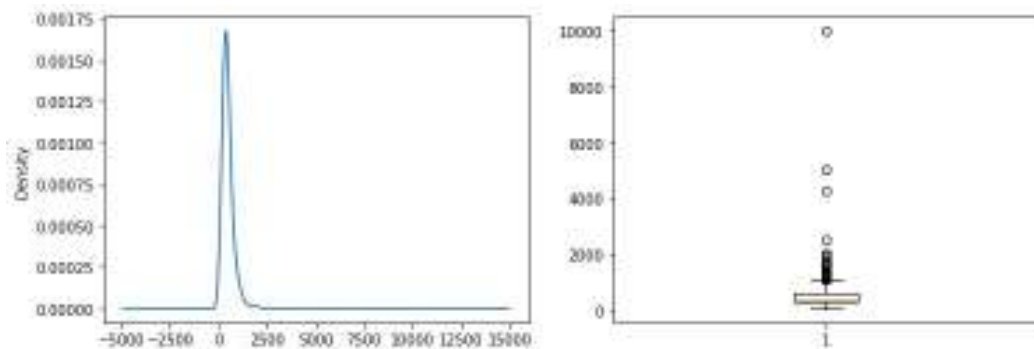
Data yang ditampilkan sudah bersih dari missing value, dan semua variable serta atribut bisa dipakai untuk diproses selanjutnya.

## 4.2 Analisa Data



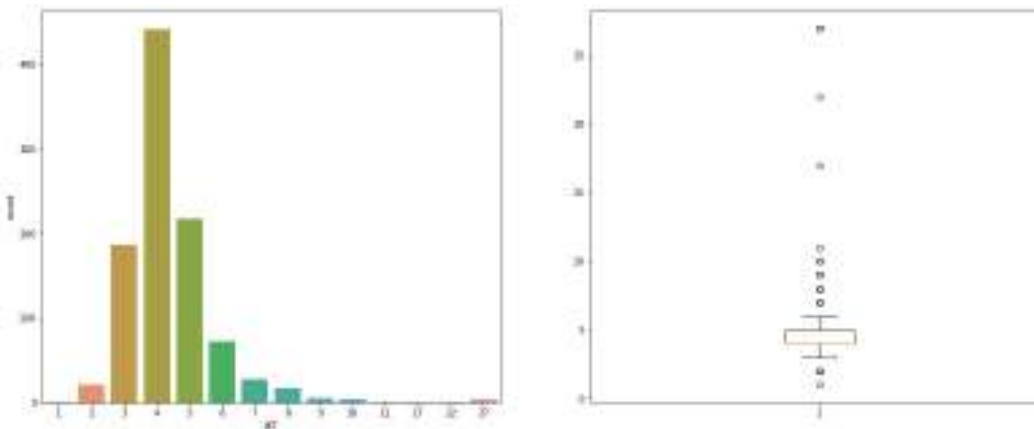
**Gambar. 2.** Chart Luas Tanah (LT)

Pada Chart diatas, luas tanah berada di sekitar angka 1 hingga 1000 m2, menandakan kebanyakan rumah memiliki luas tanah dengan rentang tersebut, dan terdapat banyak *outliers* (pencilan).



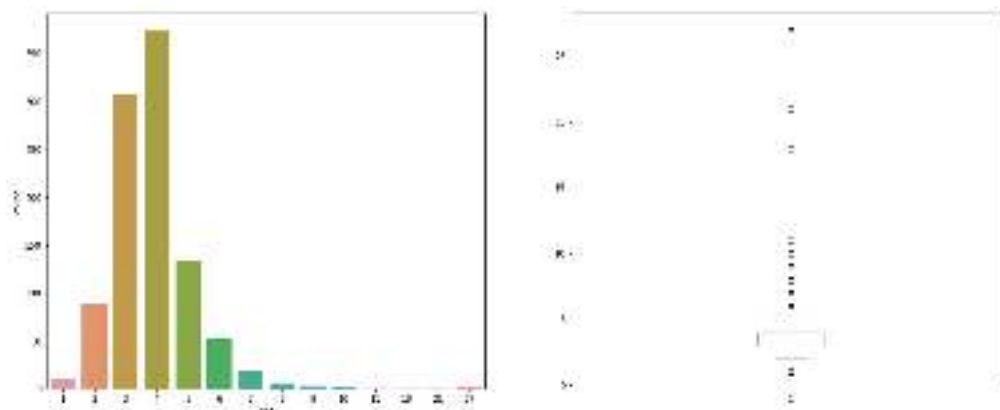
Gambar. 3. Chart Luas Bangunan (LB)

Begitu juga dengan Luas Bangunan, Luas Bangunan banyak dimiliki setiap rumah pada rentang 1 hingga 1000 m<sup>2</sup>, dan terdapat *outliers* (pencilan) jauh dari rata-rata Luas Bangunan lainnya, yakni 10.000m<sup>2</sup>.



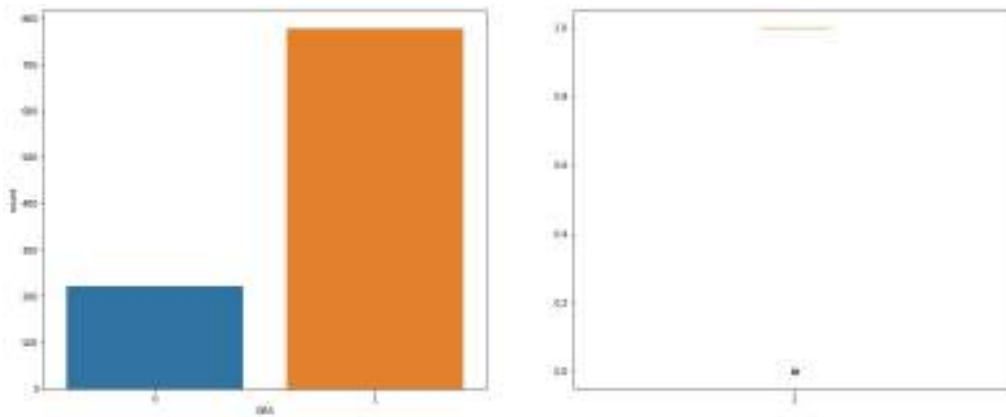
Gambar. 4. Chart Jumlah Kamar Tidur (JKT)

Jumlah Kamar Tidur pada setiap rumah memiliki rentang 3 sampai 5 kamar tidur, terdapat juga beberapa *outliers* yakni lebih dari 25 kamar tidur.



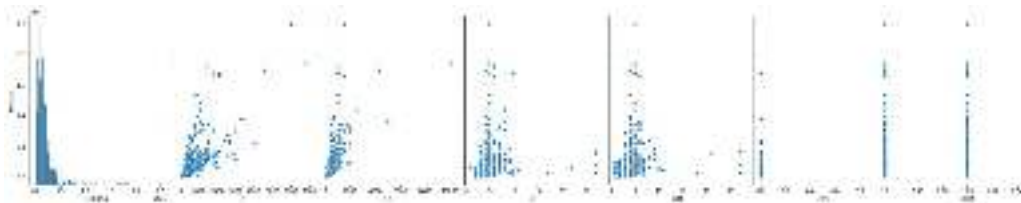
Gambar. 5. Chart Jumlah Kamar Mandi (JKM)

Pada Chart Jumlah Kamar Mandi, setiap rumah rata-rata setidaknya memiliki 3 – 4 kamar mandi. Sama seperti Jumlah Kamar Tidur, Chart ini memiliki *outliers* yang lebih dari 25 kamar mandi, itu berarti setiap kamar tidur memiliki 1 kamar mandi.



**Gambar. 6.** Chart Garasi (GRS)

Setiap rumah setidaknya memiliki Garasi yang di wakili dengan nilai 1, dari total 1001 data, terdapat kira-kira 800 data yang memiliki garasi.



**Gambar. 7.** Analisis antara variable Independen dengan Dependen

### 4.3 Pembagian Data

Data diatas sudah bersih, tahap selanjutnya yakni pembagian data menjadi 2, yakni data dibagi menjadi data uji dan data latih. Total data yang sudah bersih ada 1001 baris. Pada proses ini data dibagi meliputi 80% data latih dan 20% data uji. Pembagian ini melihat dari data yang cukup banyak, maka efektif untuk memakai ukuran pembagi 80 : 20 dibanding 70 : 30. Dari 1001 data tersebut, terpisahkan menjadi 800 data latih dan 201 data uji.

### 4.4 Hasil

Pemodelan Multiple Linear Regression diawali dengan memisahkan data menjadi data latih dan data uji, yakni 800 data latih dan 201 data uji. Digunakan 6 variabel Dependen yaitu Luas Tanah (LT), Luas Bangunan (LB), Jumlah Kamar Tidur (JKT), Jumlah Kamar Mandi (JKM), Garasi (GRS). Untuk target variable atau Independen yaitu harga pada setiap rumah. Data traning yang berjumlah 800 data dipakai untuk membuat model Multiple Linear Regression, Sedangkan data testing digunakan untuk mengetahui akurasi seberapa baik pengujian Algoritma ini digunakan [9].

Model yang sudah dihitung dari data training menggunakan bantuan Aplikasi Python Ananconda menghasilkan persamaan dibawah ini.

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_n X_n + e$$

Keterangan:

Y = Variabel terikat (Dependen)

X = Variabel bebas (Independen)

$\alpha$  = Konstanta (*Intercept*)

$\beta$  = Slope atau Koefisien estimate.

e = Error

#### Model Persamaan Multiple Linear Regression

$$Y = -2821411329.521103 + 16760457.7 X^1 + 14365458.8 X^2 + 572614432 X^3 + 51722216.8 X^4 + 1931322280 X^5 + 0 X^6$$

Setelah mendapatkan persamaan Model *Multiple Linear Regression* diatas, kemudian melatih model yang sudah didapat dengan melakukan perhitungan *Accuracy Score* di Python Anaconda.

```
In [21]: akurasi = lin_reg.score(x_test, y_test)
...: print("akurasi = ", akurasi)
akurasi = 0.6619662062746469
```

Dari pengujian diatas terlihat bahwa Akurasi yang didapat sebesar 0.66 atau 66%, cukup baik untuk memprediksikan harga rumah berdasarkan variable independent yang ada. Kemudian kami mengetes hasil akurasi tadi dengan memprediksikan harga dari suatu rumah dengan spesifikasi berikut.

1. Luas Tanah = 500 m<sup>2</sup>
2. Luas Bangunan = 625 m<sup>2</sup>
3. Jumlah Kamar Tidur = 5
4. Jumlah Kamar Mandi = 3
5. Terdapat Garasi = 1 (diwakili nilai 1, berarti terdapat garasi)
6. Rumah di Jakarta Selatan = 1 (diwakili nilai 1, berarti di Jakarta Selatan)

```
In [22]: prediksi = lin_reg.predict([[500, 625, 5, 3, 1, 1]])
...: print("prediksi = ", prediksi)
prediksi = [1.94867964e+10]
```

Harga Prediksi dari spesifikasi diatas adalah Rp 19.486.790.400 (19 Milyar Rupiah).

HARGA	LT	LB	LKT	LKM	LGS	KOTA
4790000000	500	480	4	3	1	1
2100000000	500	525	4	5	1	1
1650000000	500	780	4	4	1	1
1371000000	500	480	4	3	0	1
650000000	500	480	17	15	1	1
1500000000	500	765	4	0	1	1
2250000000	500	515	5	5	1	1
1050000000	500	700	4	0	1	1
1500000000	500	1000	5	5	1	1
1050000000	500	700	4	0	1	1

**Gambar. 8.** Luas Tanah dengan rentang 500 m<sup>2</sup>.

Melihat dari daftar harga rumah yang data Luas Tanah 500 m<sup>2</sup>, kisaran harga mencapai 15 – 22 Milyar Rupiah, dan bila dibandingkan dengan Harga prediksi, yakni 19 Milyar Rupiah, dapat dinyatakan bahwa prediksi harga rumah cukup akurat. Hal ini menandakan bahwa Model Multiple Linear Regression diatas memiliki akurasi yang Akurat.

## 5. Penutup

### 5.1 Kesimpulan

Faktor dari suatu harga rumah yaituberupa faktor luas tanah, luas bangunan, jumlah kamar tidur ataupun kamar mandi, ada/ tidaknya garasi, dan lokasi. Untuk mencari prediksi harga ini, maka perlu suatu metode yang mendukung hal tersebut, dan itu bisa didapat dari metode *Multiple Linear Regression*. Prediksi bisa didapat dari berbagai macam factor pada variable. Factor tersebut haruslah dipilih sesuai kriteria agar akurat dalam melakukan prediksi. Dalam uji sample data menggunakan 1001 baris data dan 7 kolom yang berisikan data harga rumah yang ada di Jakarta Selatan. Setelah data di bersihkan, kemudian data dipisah menjadi data training dan data testing, lalu dari data tersebut dicari akurasi Model *Multiple Linear Regression*, yang menghasilkan akurasi 66%, cukup baik untuk memprediksikan harga rumah sesuai spesifikasi yang diperlukan.

### 5.2 Saran

Dalam penelitian ini, penulis masih memiliki banyak kekurangan, sehingga penulis akan menyempurnakan penelitian ini dengan beberapa panduan. Selain sadar akan kekurangan dan perlu disempurnakan untuk meningkatkan akurasi dalam melakukan prediksi harga rumah mengingat metode yang digunakan dalam penelitian ini, masih terdapat banyak metode lainnya dan variable yang digunakan bisa ditambahkan agar akurasi bisa lebih akurat.

## 6. Daftar Pustaka

- [1] Rahayuningtyas, Evi Febrion, dkk. 2021. Prediksi Harga Rumah Menggunakan General Regression Neural Network. *Jurnal Informatika*, Vol.8 No.1 April 2021 : Halaman 59-66
- [2] Sanusi, R. M., Ansori, A. S. R., & Wijaya, R. (2020). Prediksi Harga Rumah Di Kota Bandung Bagian Timur Dengan Menggunakan Metode Regresi. *eProceedings of Engineering*, 7(3).
- [3] Adhyaksa. 2021. Definisi Rumah Sebagai Sebuah Bangunan untuk Tempat Tinggal. <https://www.adhyaksapersada.co.id/pengertian-rumah/> (diakses 15 Juni 2021)
- [4] Ardiyanto, Wahyu. 2017. Jangan Salah Kasih Harga! Ini 5 Cara Menetapkan Harga Jual Rumah. <https://www.rumah.com/berita-properti/2017/9/159522/jangan-salah-kasih-harga-ini-5-cara-menetapkan-harga-jual-rumah> (diakses 15 Juni 2021)
- [5] Azizah, E. N., Cholissodin, I., & Mahmudy, W. F. (2015). Optimasi fungsi keanggotaan fuzzy tsukamoto menggunakan algoritma genetika untuk penentuan harga jual rumah. *Journal of Environmental Engineering and Sustainable Technology*, 2(2), 79-82.
- [6] Kurniawan, Y. D. (2021). Analisis Faktor-Faktor Yang Mempengaruhi Pembelian Rumah di Kota Madiun. *J-MACC: Journal of Management and Accounting*, 4(1), 72-85.
- [7] <https://skillplus.web.id/multiple-linear-regression-pendahuluan/> (diakses 02 Juli 2021)
- [8] <https://www.kaggle.com/wisnuanggara/daftar-harga-rumah> (diakses 10 Juni 2021)
- [9] Ayuni, G. N., & Fitrianah, D. (2019). Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ. *Jurnal Telematika*, 14(2), 79-86.