# Prediksi Harga Rumah Menggunakan Web Scrapping Dan Machine Learning Dengan Algoritma Linear Regression

Andi Saiful \*1, Septi Andryana 2, Aris Gunaryati 3

<sup>1,2,3</sup> Universitas Nasional; Jl. Sawo Manila No.61, Pejaten Barat, Pasar Minggu, Jakarta Selatan, DKI Jakarta 12520

<sup>1,2,3</sup> Jurusan Informatika, Universitas Nasional, DKI Jakarta e-mail: \*\frac{1}{2}andi.saiful3135@gmail.com, \frac{2}{2}septi.andryana@civitas.unas.ac.id, \frac{3}{2} gunaryati@civitas.unas.ac.id

### Abstrak

Tempat tinggal atau rumah merupakan salah satu dari banyaknya kebutuhan primer bagi manusia. Maka dari itu sangat penting untuk membuat perencanaan agar tiap keluarga dapat memiliki tempat tinggal pribadi. Dalam perencanaan tersebut dibutuhkan prediksi atau perkiraan harga di masa mendatang. Maka, tujuan penelitian ini bertujuan untuk membuat model prediksi harga rumah menggunakan metode machine learning adapun algoritmanya adalah linear regression. Dengan melakukan web scraping untuk mengumpulkan data, melalui beberapa website yang bergelut dibidang jual beli rumah. Adapun menurut developer rumah yang berhasil ditanyakan dilapangan terkait variabel yang mempengaruhi harga rumah diantaranya luas lahan, luas bangunan berdiri, banyaknya kamar tidur, banyaknya kamar mandi, dan ketersediaan tempat parkir mobil. Untuk mendapatkan nilai prediksi yang tinggi penelitian dilakukan secara berulang kali namun nilai prediksi terbesar dengan menggunakan 80% dataset untuk training dan 20% dataset digunakan untuk testing menghasilkan nilai output dengan tingkat akurasi dalam memprediksi sebesar 88%.

Kata kunci — Prediksi, Prediksi Harga Rumah, Web Scraping, Linear Regression

## Abstract

A place to live or a house is one of the many primary needs for humans. Therefore it is very important to make a plan so that each family can have a private residence. This planning requires a prediction or forecast of future prices. So, the aim of this study is to create a house price prediction model using machine learning methods while the algorithm is linear regression. By doing web scraping to collect data, through several websites that are involved in the sale and purchase of houses. Meanwhile, according to home developers who were successfully asked in the field related to variables that affect house prices, including land area, standing building area, number of bedrooms, number of bathrooms, and the availability of a car park. To get a high predictive value, research is carried out repeatedly but the largest predictive value is using 80% of the dataset for training and 20% of the dataset is used for testing to produce an output value with an accuracy level of predicting 88%.

**Keywords**— Prediction, House Price Prediction, Web Scrapping, Linear Regression

#### 1. PENDAHULUAN

Rumah adalah salah satu kebutuhan dari masyarakat yang tidak dapat untuk dihindari dikarenakan rumah merupakan kebutuhan primer, tempat untuk berlindung, tempat untuk beristirahat dari penatnya aktifitas harian. Tidak kalah dengan emas, rumahpun juga bisa dijadikan alat untuk berinvestasi dimasa yang akan datang dikarenakan pergerakan harganya yang berubah sewaktu-waktu dan semakin banyak orang yang membutuhkan rumah, terlebih lagi dekat dengan lapangan kerja, pusat perkantoran, pusat perbelanjaan, sarana transportasi, dan lain sebagainya pasti akan mempengaruhi harga rumah tersebut dengan cepat.

Seiring dengan perkembangan, prediksi digunakan sebagai alat bantu bahkan prediksi juga telah dijadikan suatu bahan pertimbangan dalam pengambilan keputusan pembelian [1]. Dari penelitian yang dilakukan oleh Sefto P. (2016) dengan penelitian yang berjudul "Prediksi Harga Tanah menggunakan Algoritma *Linear Regression*" mengatakan bahwa *linear regression* dapat memberikan hasil pengukuran tingkat akurasi yang tinggi dan RMSE (*Root Mean Square Error*) terendah dengan data yang digunakan untuk training 70% dan sisanya data yang digunakan untuk testing 30% [2].

Sindhu P. M. (2017) pada penelitiannya menghasilkan tingkat akurasi yang sangat baik dibandingkan dengan analisis deret waktu [3]. Adapun penelitian lainnya yang menggunakan algoritma serupa adalah penelitian prediksi curah hujan di Kabupaten Majalengka [4], penelitian prediksi penjualan laptop [5], penelitian prediksi jumlah peminat mata kuliah [6], prediksi hasil panen [7]. Berdasarkan penelitian tersebut dapat diambil kesimpulan bahwa melakukan prediksi menggunakan algoritma *Linear Regression* menghasilkan tingkat akurasi yang cukup baik jika digunakan untuk memprediksi.

Penelitian Sefto P. (2016) yang menerapkan Algoritma *Linear Regression* untuk memprediksi harga tanah di Kota Banjarmasin tidak menggunakan variabel luas bangunan, banyaknya kamar tidur, banyaknya kamar mandi, hingga ketersediaan tempat parkir mobil, sedangkan pada penelitian ini menggunakan variabel-variabel tersebut. Maka, untuk membuat model prediksi harga rumah menggunakan metode *machine learning* dan adapun algoritmanya adalah *linear regression* adalah tujuan dari penelitian ini. Adapun manfaat penelitiannya adalah untuk mempelajari dan mengetahui tentang penerapan *machine learning* menggunakan algoritma *linear regression* yang digunakan untuk memprediksi harga rumah.

## 2. METODE PENELITIAN

Untuk memprediksi harga rumah adapun pendekatan penelitiannya menggunakan metode kuantitatif, lalu pada tahap metode penelitian yang akan dilakukan adalah dengan pengumpulan data dengan variable terkait, pengolahan data awal (*pre-processing data*), lalu melakukan penerapan dengan algoritma *linear regression* pada data, dan hasil penelitian seperti gambar 1 dibawah ini.



Sumber: S. Pratama, 2016

Gambar 1. Tahap Metode Penelitian

E- ISSN 2503-2933

# 2.1 Pengumpulan Data

Pada pelaksanaannya penelitian ini menggunakan pengumpulan data untuk prediksi harga rumah yang dilakukan dengan studi dokumenter pada 2 website jual beli rumah terpercaya dikarenakan tidak semua website dapat memenuhi varibel dalam penelitian ini. Berdasarkan survei kepara para pengembang (developer) rumah, terdapat 5 variable yang mempengaruhi harga suatu rumah yaitu luas lahan, luas bangunan tersebut berdiri, banyaknya kamar tidur, banyaknya kamar mandi, hingga ketersediaan tempat parkir mobil.

Dengan melakukan *web scrapping* memungkinkan untuk mengumpulkan data dengan jumlah besar [8] dan mendapatkan sumber data terkait yang semua variable tersebut dikumpulkan secara berkala disetiap harinya mulai dari 1 Oktober 2020 hingga 31 Oktober 2020 selama penelitian ini dilakukan, jumlah data terkumpul dan digunakan sebanyak 7355 data.

# 2.2 Pengolahan Data Awal

Pre-processing adalah tahap untuk membantu metode yang digunakan agar dapat menghasilkan nilai output yang lebih baik, ini merupakan tahap yang tidak dapat dilewatkan dalam pengolahan data. Pre-processing data yang dilakukan adalah melakukan cleaning pada data dengan maksud untuk memperbaiki atau menghapus data yang rusak maupun data yang tidak relevan.

# 2.3 Penerapan Linear Regression

Linear regression merupakan salah satu metode statistik yang memberikan hasil output prediksi dengan melakukan pengembangan hubungan matematis antar variabel [9]. Variabel independen merupakan variabel yang mempengaruhi sedangkan variabel dependen merupakan variabel yang dipengaruhi [10]. Adapun perhitungan multiple linear regression yang digunakan maka dinyatakan pada (1).

$$Y = \alpha + b_1 X_1 + b_2 X_2 + \dots + b X$$
 (1)

dengan

Y = variabel dependen (variabel tak terikat)  $X_1, X_2 \cdots X$  = Variabel independen (variabel terikat)

 $\alpha$  = Konstanta

 $b_1, b_2 \cdots b = \text{Koefisien regresi}$ 

Maka pernyataan diatas dapat dijadikan acuan untuk menghitung prediksi harga rumah dengan algoritma *linear regression* pada penelitian ini dan dapat dinyatakan dalam pernyataan (2).

$$Y = \alpha + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + b_5 X_5 \tag{2}$$

dengan

Y = harga rumah $\alpha = konstanta$ 

 $b_1$  = koefisien luas lahan

 $X_1$  = luas lahan

 $b_2$  = koefisien luas bangunan

 $X_2$  = luas bangunan

 $b_3$  = koefisien kamar tidur

 $X_3$  = kamar tidur

 $b_4$  = koefisien kamar mandi

 $X_4$  = kamar mandi

 $b_5$  = koefisien parkir mobil

X<sub>5</sub> = parkir mobil

Langkah selanjutnya adalah melakukan pembentukan model dengan mencari nilai  $\alpha, b_1, b_2, \cdots b$  menggunakan metode kuadrat terkecil dengan persamaan umum yang dinyatakan pada (3).

$$a + b_{1}\Sigma X_{1} + b_{2}\Sigma X_{2} + b_{3}\Sigma X_{3} + b_{4}\Sigma X_{4} + b_{5}\Sigma X_{5} = \Sigma Y$$

$$a\Sigma X_{1} + b_{1}\Sigma(X_{1})^{2} + b_{2}\Sigma(X_{1}X_{2}) + b_{3}\Sigma(X_{1}X_{3}) + b_{4}\Sigma(X_{1}X_{4}) + b_{5}\Sigma(X_{1}X_{5})$$

$$= \Sigma X_{1}Y$$

$$a\Sigma X_{2} + b_{1}\Sigma(X_{1}X_{2}) + b_{2}\Sigma(X_{2})^{2} + b_{3}\Sigma(X_{2}X_{3}) + b_{4}\Sigma(X_{2}X_{4}) + b_{5}\Sigma(X_{2}X_{5})$$

$$= \Sigma X_{2}Y$$

$$a\Sigma X_{3} + b_{1}\Sigma(X_{1}X_{3}) + b_{2}\Sigma(X_{2}X_{3}) + b_{3}\Sigma(X_{3})^{2} + b_{4}\Sigma(X_{3}X_{4}) + b_{5}\Sigma(X_{3}X_{5})$$

$$= \Sigma X_{3}Y$$

$$a\Sigma X_{4} + b_{1}\Sigma(X_{1}X_{4}) + b_{2}\Sigma(X_{2}X_{4}) + b_{3}\Sigma(X_{3}X_{4}) + b_{4}\Sigma(X_{4})^{2} + b_{5}\Sigma(X_{4}X_{5})$$

$$= \Sigma X_{4}Y$$

$$a\Sigma X_{5} + b_{1}\Sigma(X_{1}X_{5}) + b_{2}\Sigma(X_{2}X_{5}) + b_{3}\Sigma(X_{3}X_{5}) + b_{4}\Sigma(X_{4}X_{5}) + b_{5}\Sigma(X_{5})^{2}$$

$$= \Sigma X_{5}Y$$

Setelah hasil intervensi diketahui, langkah selanjutnya adalah melakukan perkalian matriks determinan dengan  $\Sigma Y$ ,  $\Sigma X_1 Y$ ,  $\Sigma X_2 Y$ ,  $\Sigma X_3 Y$ ,  $\Sigma X_4 Y$ ,  $\Sigma X_5 Y$  untuk menghitung determinan matriks A,  $A_0$ ,  $A_1$ ,  $A_2$ ,  $A_3$ ,  $A_4$ , dan  $A_5$  menggunakan perhitungan yang dinyatakan pada (4).

$$\begin{bmatrix} N & \Sigma X_{1} & \Sigma X_{2} & \Sigma X_{3} & \Sigma X_{4} & \Sigma X_{5} \\ \Sigma X_{1} & \Sigma X_{1}X_{1} & \Sigma X_{1}X_{2} & \Sigma X_{1}X_{3} & \Sigma X_{1}X_{4} & \Sigma X_{1}X_{5} \\ \Sigma X_{2} & \Sigma X_{2}X_{1} & \Sigma X_{2}X_{2} & \Sigma X_{2}X_{3} & \Sigma X_{2}X_{4} & \Sigma X_{2}X_{5} \\ \Sigma X_{3} & \Sigma X_{3}X_{1} & \Sigma X_{3}X_{2} & \Sigma X_{3}X_{3} & \Sigma X_{3}X_{4} & \Sigma X_{3}X_{5} \\ \Sigma X_{4} & \Sigma X_{4}X_{1} & \Sigma X_{4}X_{2} & \Sigma X_{4}X_{3} & \Sigma X_{4}X_{4} & \Sigma X_{4}X_{5} \\ \Sigma X_{5} & \Sigma X_{5}X_{1} & \Sigma X_{5}X_{2} & \Sigma X_{5}X_{3} & \Sigma X_{5}X_{4} & \Sigma X_{5}X_{5} \end{bmatrix} x \begin{bmatrix} a \\ b_{1} \\ b_{2} \\ b_{3} \\ b_{4} \\ b_{5} \end{bmatrix} = \begin{bmatrix} \Sigma Y \\ \Sigma X_{1}Y \\ \Sigma X_{2}Y \\ \Sigma X_{3}Y \\ \Sigma X_{4}Y \\ \Sigma X_{5}Y \end{bmatrix}$$

$$(4)$$

Selanjutnya, hasil yang diperoleh dari perhitungan determinan matriks dapat digunakan untuk menemukan nilai dari  $a_1 b_1 b_2 b_3 b_4 b_5$  dengan mengacu pada pernyataan (5).

$$a = \frac{Det(A_0)}{Det(A)}$$

$$b_1 = \frac{Det(A_1)}{Det(A)}$$

$$b_2 = \frac{Det(A_2)}{Det(A)}$$

$$b_3 = \frac{Det(A_3)}{Det(A)}$$

$$b_4 = \frac{Det(A_4)}{Det(A)}$$

$$b_5 = \frac{Det(A_5)}{Det(A)}$$

$$(5)$$

Tahap selanjutnya untuk mengetahui tingkat keterkaitan kedua variabel perlu dilakukannya uji korelasi. Adapun nilai yang dicari adalah rX<sub>1</sub>Y, rX<sub>2</sub>Y, rX<sub>3</sub>Y, rX<sub>4</sub>Y, rX<sub>5</sub>Y, rX<sub>1</sub>X<sub>2</sub>,  $rX_1X_3$ ,  $rX_1X_4$ ,  $rX_1X_5$  dengan perhitungan nilai korelasi yang dilakukan seperti pada pernyataan (6).

$$\Sigma X_{1}Y = \Sigma X_{1}Y - \frac{(\Sigma X_{1}) \cdot (\Sigma Y)}{n}$$

$$\Sigma X_{2}Y = \Sigma X_{2}Y - \frac{(\Sigma X_{1}) \cdot (\Sigma Y)}{n}$$

$$\Sigma X_{3}Y = \Sigma X_{3}Y - \frac{(\Sigma X_{1}) \cdot (\Sigma Y)}{n}$$

$$\Sigma X_{4}Y = \Sigma X_{4}Y - \frac{(\Sigma X_{1}) \cdot (\Sigma Y)}{n}$$

$$\Sigma X_{5}Y = \Sigma X_{5}Y - \frac{(\Sigma X_{1}) \cdot (\Sigma Y)}{n}$$

$$\Sigma X_{1}X_{2} = \Sigma X_{1}X_{2} - \frac{(\Sigma X_{1}) \cdot (\Sigma X_{2})}{n}$$

$$\Sigma X_{1}X_{3} = \Sigma X_{1}X_{3} - \frac{(\Sigma X_{1}) \cdot (\Sigma X_{3})}{n}$$

$$\Sigma X_{1}X_{4} = \Sigma X_{1}X_{4} - \frac{(\Sigma X_{1}) \cdot (\Sigma X_{4})}{n}$$

$$\Sigma X_{2}X_{3} = \Sigma X_{2}X_{3} - \frac{(\Sigma X_{2}) \cdot (\Sigma X_{3})}{n}$$

$$\Sigma X_{2}X_{3} = \Sigma X_{2}X_{3} - \frac{(\Sigma X_{2}) \cdot (\Sigma X_{3})}{n}$$

$$\Sigma X_{2}X_{4} = \Sigma X_{2}X_{4} - \frac{(\Sigma X_{2}) \cdot (\Sigma X_{4})}{n}$$

$$\Sigma X_{2}X_{5} = \Sigma X_{2}X_{5} - \frac{(\Sigma X_{2}) \cdot (\Sigma X_{5})}{n}$$

$$\Sigma X_{3}X_{4} = \Sigma X_{3}X_{4} - \frac{(\Sigma X_{3}) \cdot (\Sigma X_{5})}{n}$$

$$\Sigma X_{4}X_{5} = \Sigma X_{2}X_{5} - \frac{(\Sigma X_{3}) \cdot (\Sigma X_{5})}{n}$$

$$\Sigma X_{4}X_{5} = \Sigma X_{2}X_{5} - \frac{(\Sigma X_{4}) \cdot (\Sigma X_{5})}{n}$$

Lalu selanjutnya, setelah mendapatkan nilai korelasi yaitu melakukan uji koefisiensi determinasi untuk mengetahui besarnya hubungan variabel independen dengan variabel dependen, maka dapat diketahui keterkaitan pada model tersebut. Koefisien determinasi hanya memiliki nilai berkisar antara 0 sampai 1 jika  $R^2 = 0$ , maka dapat diartikan bahwa tidak ada keterkaitan pada kedua variabel (independen dan dependen), dan sebaliknya jika R<sup>2</sup> semakin menjauhi angka 0, maka akan semakin kuat hubungan kedua variabel (independen dan dependen). Untuk mencari R<sup>2</sup> dapat dilakukan seperti pada pernyataan (7).

$$R^{2} = 1 - \frac{SS \ error}{SS \ total} = 1 - \frac{\Sigma (y_{i} - \hat{y}_{i})^{2}}{\Sigma (y_{i} - \bar{y})^{2}}$$
 (7)

dengan

observasi respons ke-i  $y_{i}$ = ramalan respons ke-i

ŷi

ÿ = rata-rata Dan mencari koefisien determinasi dapat dilakukan seperti pada pernyataan (8).

$$Kd = R^2 x 100\%$$
 (8)

# 2.4 Pengujian Hasil

Tahap pengujian hasil akurasi dengan menggunakan nilai RMSE yang mengacu pada pernyataan (9) dimana RMSE (Root Mean Squared Error) sering digunakan untuk mengevaluasi kinerja prediksi yang digunakan [11] dengan mengukur tingkat akurasi dari hasil prediksi. Jika nilai RMSE rendah menunjukan bahwa bentuk perubahan nilai yang dihasilkan oleh suatu model prediksi mendekati bentuk nilai aslinya. Dan sebaliknya, jika besar nilai RMSE maka keakuratan yang dihasilkan jauh dari bentuk nilai aslinya.

$$RMSE = \frac{\Sigma_{i}^{n} = \sqrt{(y_{i} - \hat{y}_{i})^{2}}}{n}$$
(9)

dengan

 $y_i$  = data awal  $\hat{y}_i$  = data akhir n = jumlah data

#### 3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan Google Colab berbasis web base yang akan memudahkan dalam melakukan penelitian prediksi harga rumah. Google Colab adalah sebuah web aplikasi open-source adapun kegunaan lainnya dari Google Colab di penelitian ini adalah untuk data cleaning, transformasi data, visualisasi data, machine learning, alat simulasi, dan lain sebagainya.

	luas_lahan	luas_bangunan	kmr_tidur	kmr_mandi	tmp_parkir	harga
0	60	36	2	1	0	335000
1	110	45	2	1	0	595000
2	73	45	2	1	0	425000
3	63	36	2	1	0	346500
4	66	36	2	1	0	418800
7437	72	44	2	1	1	462000
7438	132	80	3	1	0	500000
7439	82	130	3	2	0	850000
7440	71	44	2	1	1	458000
7441	90	90	2	1	0	699900

Keterangan: Harga x1000

Gambar 2. Data Set Sebanyak 7442 Data.

Terkumpul sebanyak 7442 data yang diperoleh melalui *web scrapping* dari 2 website jual-beli rumah terpercaya yang dikumpulkan mulai dari tanggal 1 Oktober 2020 sampai 31 Oktober 2020. Pada gambar 2 menunjukan data rumah yang berhasil dikumpulkan berisi kolom luas lahan, luas bangunan, kamar tidur, kamar mandi, tempat parkir, dan harga rumah.

luas_lahan	luas_bangunan	kmr_tidur	kmr_mandi	tmp_parkir	harga
60	36	2	1	0	335000
110	45	2	1	0	595000
73	45	2	1	0	425000
63	36	2	1	0	346500
66	36	2	1	0	418800
***	***			***	
75	45	2	1	0	360000
80	36	2	1	0	80000
75	45	2	1	0	360000
78	36	2	1	0	55000
60	30	2	1	0	55000

794 rows × 6 columns

Keterangan: Harga x1000

Gambar 3. Data Set Sebanyak 794 Data.

Dilakukan *data cleaning* pada tahap *pre-processing*. Kegunaannya adalah untuk memperbaiki atau menghapus data yang rusak maupun data yang tidak relevan. Selama masa pengumpulannya data yang terkumpul adalah 7442 data namun ditemukannya *data duplicate*, maka akan dilakukan *data cleaning* agar tidak memiliki data sama (*data duplicate*). Gambar 3 menunjukan data setelah dihapus data yang *duplicate*.

count	7.94000e+02	count	7.820000e+02
mean	4.350656e+05	mean	4.417418e+05
std	5.727996e+05	std	5.746196e+05
min	0.000000e+00	min	5.000000e+04
25%	2.80000e+05	25%	2.879000e+05
50%	3.64000e+05	50%	3.650000e+05
75%	4.500000e+05	75%	4.500000e+05
max	9.500000e+06	max	9.500000e+06
	(a)		(b)

Gambar 4. (a) Data set dengan nilai yang tidak diinginkan. (b) Data set dengan nilai sudah diperbaiki

Pada gambar 4(a) menunjukan terdapat data yang tidak diinginkan karena variabel harga tidak memiliki nilai atau 0. Untuk hasil prediksi yang akurat baiknya semua variabel harga memiliki nilai dan tidak 0 (nol). Dan setelah melalui tahap perbaikan data maka dapat dilihat pada gambar 4(b) yang menunjukan tidak terdapat pada variabel harga tidak memiliki nilai atau 0.

Hasil penerapan dengan algoritma *linear regression* menghasilkan hubungan antara luas lahan, luas bangunan, banyaknya kamar tidur, banyaknya kamar mandi, dan ketersediaan tempat parkir dengan mengacu pada pernyataan (10).

$$Y = a + b_1(luas\ lahan) + b_2(luas\ bangunan) + b_3(kamar\ tidur) + b_4(kamar\ mandi) + b_5(tempat\ parkir)$$
(10)

# 3.1 Uji Korelasi

Untuk mengetahui keterkaitan pada setiap variabel independen terhadap variabel dependen perlu adanya uji korelasi. Untuk menghitung nilai korelasi perlu mengetahui perhitungan nilai  $rX_1Y$ ,  $rX_2Y$ ,  $rX_3Y$ ,  $rX_4Y$ ,  $rX_5Y$ ,  $rX_1X_2$ ,  $rX_1X_3$ ,  $rX_1X_4$ ,  $rX_1X_5$ .

Pada gambar 5 dapat dilihat tabel korelasi antar variabel satu dengan variabel lainnya menunjukan yang paling mempengaruhi harga adalah luas bangunan berdiri dan dari tabel data tersebut dapat dilihat korelasi yang tidak mempengaruhi harga di data set adalah *car parking* atau tempat parkir mobil.

	luas_lahan	luas_bangunan	kmr_tidur	kmr_mandi	tmp_parkir	harga
luas_lahan	1.00	0.62	0.36	0.43	-0.02	0.74
luas_bangunan	0.62	1.00	0.64	0.71	-0.03	0.82
kmr_tidur	0.36	0.64	1.00	0.72	-0.02	0.53
kmr_mandi	0.43	0.71	0.72	1.00	-0.01	0.60
tmp_parkir	-0.02	-0.03	-0.02	-0.01	1.00	0.01
harga	0.74	0.82	0.53	0.60	0.01	1.00

Gambar 5. Data Korelasi dari Setiap Variabel

### 3.2 Uji Koefisiensi Determinasi

Tahap untuk mengetahui besar pengaruh antara kedua variabel adalah dengan melakukan uji koefisien determinasi, sehingga dapat diketahui akurasi dari penerapan algoritma linear regression. Pada koefisien ini dengan nilai berkisar antara 0 sampai 1. Jika nilai  $R^2=0$  dapat diartikan bahwa tidak adanya pengaruh atau hubungan antara kedua variabel (independen dan dependen) dan jika  $R^2=1$  atau mendekati 1, maka memiliki hubungan yang semakin kuat antara variabel independen dengan variabel dependen.

Tabel 1. Hasil Uji Koefisiensi Determinasi

Two of TV Trusti Of Tro of the Trustian E over the trustian E			
R	$R^2$		
0.7008425785021375	0.8859830993050699		

Dari hasil yang ditunjukan pada tabel 1 menunjukan bahwa penelitian prediksi harga rumah dengan *web scrapping* dan *machine learning* menggunakan algoritma *linear regression*, dengan menggunakan data training 80% dan data testing 20% menghasilkan tingkat akurasi prediksi sebesar 0.8859830993050699 atau 88%.

Tabel 2. RMSE (Root Mean Square Error)

Tabel 2: RVISE (Root Weath Square Error)
RMSE
259171.91

Selanjutnya adalah menguji nilai RMSE (*Root Mean Square Error*) untuk mengevaluasi hasil prediksi yang digunakan lalu mengukur nilai akurasi dari hasil penerapan prediksi pada

model. Pada tabel 2 menunjukan *Root Mean Square Error* prediksi sebesar 259171.91 atau Rp. 259.171.91.

#### 4. KESIMPULAN

Dari hasil penerapan prediksi harga rumah dapat disimpulkan bahwa pengolahan awal data yang dilakukan pada data set 7442 data menjadi 794 data sangat mempengaruhi dalam tingkat akurasi prediksi harga rumah. Dengan algoritma *linear regression* untuk memprediksi harga rumah dapat memberikan hasil keakuratan prediksi harga rumah dengan baik. Adapun hasil akurasi terbaik dengan menggunakan data set untuk training sebesar 80% dan menggunakan 20% untuk data testing menghasilkan tingkat akurasi prediksi sebesar 0.8859830993050699 atau 88%. Namun, memiliki nilai galat cukup tinggi sebesar 259171.91 atau Rp. 259.171,91.

### 5. SARAN

Analisis tentang prediksi harga rumah menggunakan *web* scrapping dan *machine learning* dengan algoritma *linear regression* dilihat dari nilai RMSE, dimana nilai yang dihasilkan pada penelitian ini cukup tinggi. Maka, untuk penelitian dan pengujian yang akan datang agar hasil yang didapatkan lebih baik maka perlu diperhatikan:

- 1. Data yang digunakan sebagai data set ditambahkan variabel tak pasti.
- 2. Perlu ditambahkannya algoritma lain untuk mendapatkan hasil prediksi lebih akurat.

# **UCAPAN TERIMA KASIH**

Penulis mengucapkan terima kasih kepada dosen pembimbing di jurusan Informatika Universitas Nasional yang telah memberi bimbingan terhadap penulisan penelitian ini.

#### DAFTAR PUSTAKA

- [1] G. N. Ayuni dan D. Fitrianah. 2019, *Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ, J. Telemat., Vol. 14, No. 2, hal. 79–86*, [Daring]. Tersedia pada: https://journal.ithb.ac.id/telematika/article/view/321.
- [2] S. Pratama. 2016, Prediksi Harga Tanah Menggunakan Algoritma Linear Regression, Technologia, Vol. 7, No. 2, hal. 122–130,.
- [3] S. P. Menon, R. Bharadwaj, P. Shetty, P. Sanu, dan S. Nagendra. 2018, Prediction of Temperature Using Linear Regression," *Int. Conf. Electr. Electron. Commun. Comput. Technol. Optim. Tech. ICEECCOT* 2017, hal. 670–671, doi: 10.1109/ICEECCOT.2017.8284588.
- [4] E. Dewi, S. Mulyani, I. Septianingrum, N. Nurjanah, dan R. Rahmawati, 2019, *Prediksi Curah Hujan Di Kabupaten Majalengka Dengan Menggunakan Algoritma Regresi Rainfall Prediction in Majalengka District Using Regression Algorithm, Vol. 8, No. 1.*

- [5] C. Siregar, A. S. Sembiring, dan H. K. Siburian. 2018, *Perancangan Aplikasi Prediksi Penjualan Laptop Dengan Menerapkan Metode Regresi Linier*," *J. Pelita Inform.*, *Vol. 17*, *No. 4*, *hal. 416–421*, [Daring]. Tersedia pada: https://ejurnal.stmikbudidarma.ac.id/index.php/pelita/article/view/1079/932.
- [6] N. K. Afkarina, A. W. Widodo, dan M. T. Furqon. 2019, *Implementasi Regresi Linier Berganda Untuk Prediksi Jumlah Peminat Mata Kuliah Pilihan*, *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, Vol. 3, No. 11, hal. 10462–10467.
- [7] V. Sellam dan E. Poovammal. 2016, Prediction of Crop Yield Using Regression Analysis," Indian J. Sci. Technol., Vol. 9, No. 38, , doi: 10.17485/ijst/2016/v9i38/91714.
- [8] A. A. Maulana, A. Susanto, dan D. P. Kusumaningrum, 2019, *Rancang Bangun Web Scraping Pada Marketplace di Indonesia*," *JOINS (Journal Inf. Syst.*, Vol. 4, No. 1, hal. 41–53, doi: 10.33633/joins.v4i1.2544.
- [9] A. Muqtadir, A. A. Suryanto. 2018, Penerapan Regresi Linear Untuk Memprediksi Kebutuhan Produksi Padi, Vol. 3, hal. 331–332.
- [10] N. Kusumawati, F. Marisa, dan I. D. Wijaya, 2017, *Prediksi Kurs Rupiah Terhadap Dollar Amerika Dengan Menggunakan Metode Regresi Linear*, *J I M P J. Inform. Merdeka Pasuruan*, Vol. 2, No. 3, hal. 45–56, doi: 10.37438/jimp.v2i3.79.
- [11] A. Chakraborty dan D. Goswami, 2017, Prediction of Slope Stability Using Multiple Linear Regression (MLR) and Artificial Neural Network (ANN), Arab. J. Geosci., Vol. 10, No. 17, hal. 1–11, doi: 10.1007/s12517-017-3167-x.