



SUBMIT

(Jurnal Ilmiah Teknologi Informasi dan Sains)

Vol.4 No.1 (2024)30-35

ISSN Media Elektronik: 2798-6861

PREDIKSI HARGA RUMAH SESUAI SPESIFIKASI MENGGUNAKAN METODE MULTIPLE LINEAR REGRESSION

Artika Widyastuti

Informatika, Universitas Islam Majapahit

Email: artikawidy007@gmail.com

(Naskah masuk: 20 Juni 2024, diterima untuk diterbitkan: 30 Juni 2024)

Abstrak

Tempat tinggal atau rumah adalah salah satu kebutuhan utama manusia yang berguna sebagai tempat tinggal, tempat bersantai dan berkumpul dengan keluarga. Sehingga pertimbangan dalam pembelian rumah memerlukan sebuah metode untuk membantu menentukan harga yang sesuai pembeli inginkan. Dengan menggunakan metode *Multiple Linear Regression* dari data rumah yang didapatkan dari website *kaggle.com* terdapat variabel-variabel potensial seperti harga rumah, luas tanah, luas bangunan, jumlah ruang kamar tidur, jumlah ruang kamar mandi dan garasi. Penelitian ini menggunakan dataset sebanyak 1010 data dengan 6 variabel. Data tersebut akan melalui proses *pre-processing* data sebelum dilakukannya training model. Selanjutnya pada tahap evaluasi menggunakan matrik pengujian, yaitu MAE (Mean Absolute Error), MSE (Mean Squared Error) digunakan untuk menilai kinerja model. Berdasarkan hasil analisis dari dataset, dimana data tersebut dibagi menjadi dua bagian yaitu, sebanyak 80% sebagai data training dan 20% sebagai data testing dari hasil analisis tersebut diperoleh tingkat akurasi sebesar 0,77 atau 77% dengan tingkat error MAE sebesar 1980.3555 dan tingkat error MSE sebesar 10675731. Sehingga proses perhitungan prediksi harga yang dilakukan untuk sebuah rumah dengan spesifikasi tertentu dapat menjadi lebih mudah, menghemat waktu dan hasil yang didapat lebih akurat.

Kata kunci: *prediksi harga rumah, multiple linear regression*

PREDICTION OF HOUSE PRICES ACCORDING TO SPECIFICATIONS USING MULTIPLE LINEAR REGRESSION METHOD

Abstract

Housing is one of the main human needs that is useful as a place to live, a place to relax and gather with family. So that consideration in buying a house requires a method to help determine the price that the buyer wants. By using the Multiple Linear Regression method from house data obtained from the kaggle.com website, there are potential variables such as house price, land area, building area, number of bedrooms, number of bathrooms and garages. This research uses a dataset of 1010 data with 6 variables. The data will go through a data pre-processing process before training the model. Furthermore, the evaluation stage uses a testing matrix, namely MAE (Mean Absolute Error), MSE (Mean Squared Error) used to assess model performance. Based on the results of the analysis of the dataset, where the data is divided into two parts, namely, as much as 80% as training data and 20% as testing data from the results of the analysis obtained an accuracy rate of 0.77 or 77% with an MAE error rate of 1980.3555 and an MSE error rate of 10675731. So that the price prediction calculation process carried out for a house with certain specifications can be easier, save time and the results obtained are more accurate.

Keywords: *house price prediction, multiple linear regression*

1. PENDAHULUAN

Tempat tinggal atau rumah adalah salah satu kebutuhan utama manusia yang berguna sebagai tempat tinggal, tempat bersantai, berkumpul dengan keluarga, tempat untuk berlindung, serta untuk beristirahat setelah melakukan aktifitas harian. Tempat tinggal juga memiliki peran sebagai representasi dari identitas individu dalam usaha meningkatkan kualitas hidup, serta membentuk sifat, karakter, dan kepribadian suatu bangsa (Sanusi, Ansori and Rifki, 2020). Dari segi psikologis, rumah menjadi lingkungan yang menciptakan suasana yang nyaman, aman, dan penuh kegembiraan bagi semua penghuninya.

Kebutuhan fisiologis manusia akan semakin bertambah, salah satunya adalah kebutuhan akan tempat tinggal atau rumah. Pengusaha properti akan berlomba-lomba untuk membangun properti seperti rumah sebagai sarana investasi. Hal ini akan membuat masyarakat dalam membeli rumah berfikir apakah rumah yang ia beli akan mempunyai nilai keuntungan yang baik atau tidak, dikarenakan harga rumah yang akan semakin naik setiap waktu. Semakin baik kualitas rumah dan semakin strategis lokasinya, maka harga rumah kemungkinan akan lebih tinggi (Kurniawan, 2021).

Seiring dengan perkembangan waktu, prediksi digunakan sebagai alat bantu bahkan prediksi juga telah dijadikan suatu bahan pertimbangan dalam pengambilan keputusan pembelian. Kemampuan melakukan prediksi dengan tingkat akurasi yang dapat dipertanggungjawabkan dapat membantu sebuah perusahaan atau instansi untuk mengambil keputusan yang tepat.

Sehubungan dengan hal tersebut maka diperlukan suatu sistem keputusan yang membantu calon pembeli dalam menentukan 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*. *Multiple Linear Regression* adalah salah satu metode statistik untuk memprediksi variabel dependen (variabel yang akan diprediksi) berdasarkan beberapa variabel independen (variabel yang dapat mempengaruhi variabel dependen). *Multiple Linear Regression* merupakan versi yang lebih kompleks dari regresi linier sederhana, sedangkan regresi linier sederhana hanya menggunakan satu variabel independen untuk memprediksi variabel dependen (Saputra Aldi Agusta, Munir and Daffa Rizki, 2023). 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

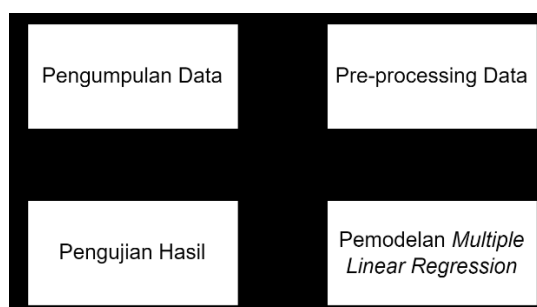
berdasarkan nilai-nilai variabel independen yang diberikan (Puteri and Silvanie, 2020).

Beberapa penelitian terkait metode *Multiple Linear Regression* sebagai sistem penunjang keputusan dilakukan oleh peneliti terdahulu. Penelitian (Mu'tashim *et al.*, 2021) dilakukan tentang prediksi harga rumah berdasarkan spesifikasi dengan menggunakan metode *Multiple Linear Regression*. Penelitian ini menggunakan beberapa variabel independen yang mempengaruhi proses prediksi harga rumah yaitu jumlah kamar, jumlah kamar mandi, luas tanah, luas bangunan, garasi dan kota. Hasil dari penelitian ini, ditemukan bahwa model memiliki tingkat akurasi sebesar 66%. (Sanusi, Ansori and Rifki, 2020) pada penelitiannya Prediksi Harga Rumah Di Kota Bandung Bagian Timur Dengan Menggunakan Metode Regresi menjelaskan bahwa memprediksi Kota Bandung Timur termasuk Kecamatan saja. Pengujian dengan sistem Regresi Linier mendapatkan nilai rata-rata semua data yang terdapat Kota Bandung Timur nilai akurasi sekitar 85-91% dan nilai error sekitar 9-15%.

Pada penelitian ini peneliti mengimplementasikan metode *Multiple Linear Regression* untuk membantu memprediksi kisaran harga dalam membeli rumah sesuai spesifikasi yang ada. Proses analisis pada penelitian ini menggunakan 5 variabel independen yaitu luas bangunan, luas tanah, jumlah kamar tidur, jumlah kamar mandi 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 dengan harga rumah yang sesuai dengan keadaan individu.

2. METODOLOGI

Pada tahap penelitian ini dilakukan beberapa tahapan dalam mengimplementasikan metode *Multiple Linear Regression* dalam proses prediksi harga rumah. Tahapan proses yang akan dilakukan pada Gambar 1, adalah dengan pengumpulan data, pre-processing data, lalu melakukan penerapan model dengan algoritma *multiple linear regression* pada data, dan melakukan pengujian hasil.



Gambar 1. Tahap Metode Penelitian

2.1. Pengumpulan Data

Pada penelitian ini dataset di dapat dari *kaggle*. Dataset tersebut terdiri dari 1010 row dan 6 column. Dataset yang digunakan dalam penelitian ini mengandung informasi mengenai harga rumah, luas bangunan, luas tanah, jumlah kamar tidur, jumlah kamar mandi dan garasi.

2.2. Pengolahan Data Awal

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 (Hasanah, Maula and Tholib, 2023).

2.3. Pemodelan Multiple Linear Regression

Multiple Linear Regression merupakan 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 (Siregar, Sitorus and Ananta, 2023).

Pada persamaan 1 merupakan perhitungan rumus persamaan model *multiple linear regression*.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_n X_n + \varepsilon \quad (1)$$

Keterangan:

- Y = Variabel terikat (Dependen)
- X₁, X₂, ..., X_n = Variabel bebas (Independen)
- β₀, β₁, ..., β_n = Slope atau Koefisien estimate
- ε = Error

2.4. Pengujian Hasil

Tahap pengujian hasil dilakukan pada proses perhitungan menggunakan *Multiple Linear Regression*, disini awalnya data akan dibagi 2, yakni data testing atau data uji dan data training atau data latih. Setelah hal ini dilakukan, akurasi dan spesifikasi yang tersedia dapat digunakan untuk menentukan estimasi rumah.

3. HASIL DAN PEMBAHASAN

3.1. Data Pre-processing

Pada tahap ini dilakukan *data cleaning* untuk memastikan kualitas data yang akan digunakan. Dataset ini terdapat total 1010 baris data dan 6 kolom. Data telah dibersihkan dan dapat dilanjutkan ke tahap berikutnya.

Tabel 1. Data Pre-processing

	HARGA	LB	LT	KT	KM	GRS
0	3800000000	220	220	3	3	0

1	4600000000	180	137	4	3	2
2	3000000000	267	250	4	4	4
3	4300000000	40	25	2	2	0
4	9000000000	400	355	6	5	3

Pada Tabel 1, adalah sebagian teratas dari dataset yang telah melewati proses pembersihan, di mana atribut-atribut yang tidak diperlukan, seperti nomor dan nama rumah telah dihilangkan. Data ini sekarang siap untuk diproses pada langkah berikutnya.

Penjelasan setiap kolom:

HARGA = Harga Rumah

LB = Luas Bangunan

LT = Luas Tanah

KT = Jumlah Kamar Tidur

KM = Jumlah Kamar Mandi

GRS = Jumlah Garasi

Selanjutnya proses tahap pemeriksaan *missing value*, dalam dataset tersebut diperiksa apakah terdapat data yang bernilai *null* (*missing value*).

```
HARGA    0
LB        0
LT        0
KT        0
KM        0
GRS       0
dtype: int64
```

Gambar 2. *missing value*

Berdasarkan Gambar 2, pada pengecekan *missing value*, tidak ada data yang kosong sehingga dapat dilanjutkan ke tahapan selanjutnya pemeriksaan type data.

Tabel 2. Informasi Type Data

#	Column	Non-Null Count	Dtype
0	HARGA	1010 non-null	int64
1	LB	1010 non-null	int64
2	LT	1010 non-null	int64
3	KT	1010 non-null	int64
4	KM	1010 non-null	int64
5	GRS	1010 non-null	int64

Pada Tabel 2, dapat dilihat informasi dari tipe data yang berjumlah 1010 record. Kolom HARGA, N, LT, KT, KM, GRS memiliki tipe data int64. Karena tipe data yang akan dilakukan proses sudah numerik, jadi tidak perlu melakukan perubahan tipe data

3.2. Exploratory Data

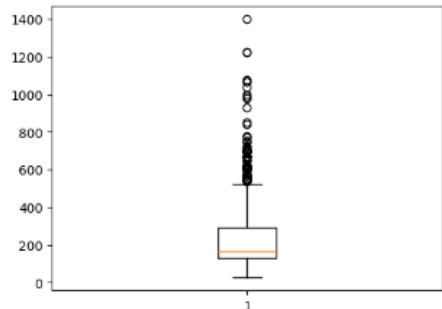
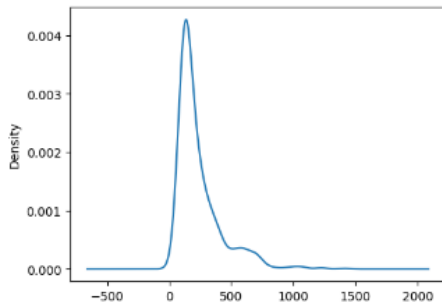
Setelah dilakukan *data cleaning*, selanjutnya dilakukan statistik deskriptif untuk merangkum dan mendeskripsikan dataset.

	HARGA	LB	LT	KT	KM	GRS
count	1010.000000	1010.000000	1010.000000	1010.000000	1010.000000	1010.000000
mean	7628.982178	276.539604	237.432673	4.668317	3.607921	1.920792
std	7340.945285	177.864557	179.957604	1.572776	1.420066	1.510998
min	430.000000	40.000000	25.000000	2.000000	1.000000	0.000000
25%	3262.500000	150.000000	130.000000	4.000000	3.000000	1.000000
50%	5000.000000	216.500000	165.000000	4.000000	3.000000	2.000000
75%	9000.000000	350.000000	290.000000	5.000000	4.000000	2.000000
max	65000.000000	1126.000000	1400.000000	10.000000	10.000000	10.000000

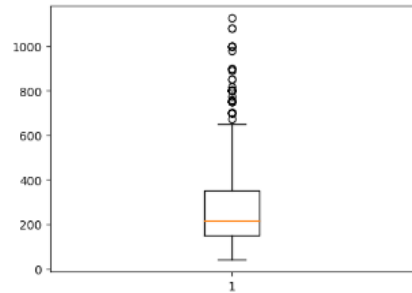
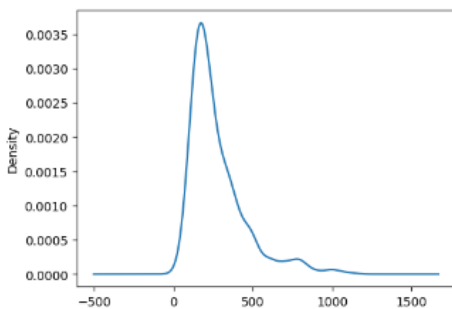
Gambar 3. Deskripsi Data

Pada Gambar 3, dapat dilihat nilai dari data, mulai dari total(count), standar deviasi, nilai rata-rata(mean), nilai minimal(min), serta nilai maksimal(max). Dapat disimpulkan bahwa dataset tersebut tampak lengkap karena tidak terdapat *missing values* pada setiap kolom variabel.

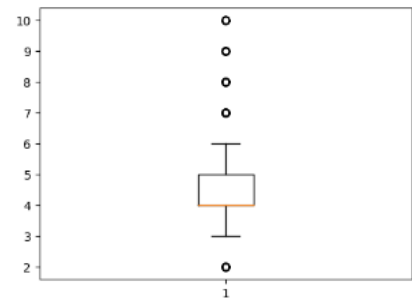
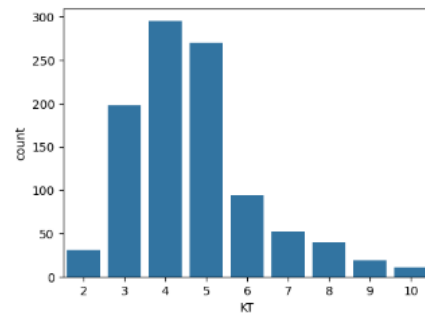
Proses selanjutnya adalah melibatkan penemuan dan menghapus outlier pada setiap atribut data rumah dengan merepresentasikan visual data untuk masing-masing atribut yang menunjukkan adanya outlier.



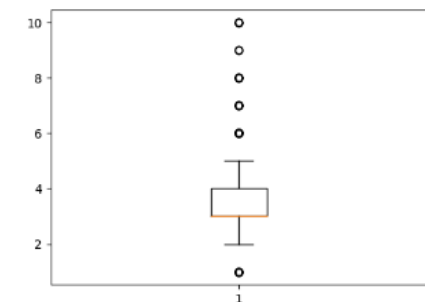
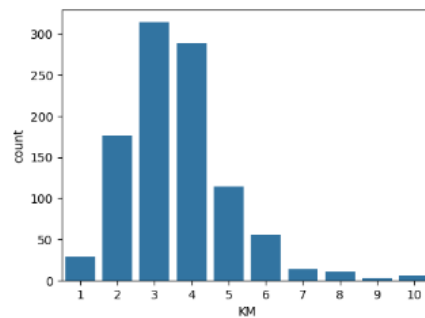
Gambar 4. Distribusi Data Luas Tanah



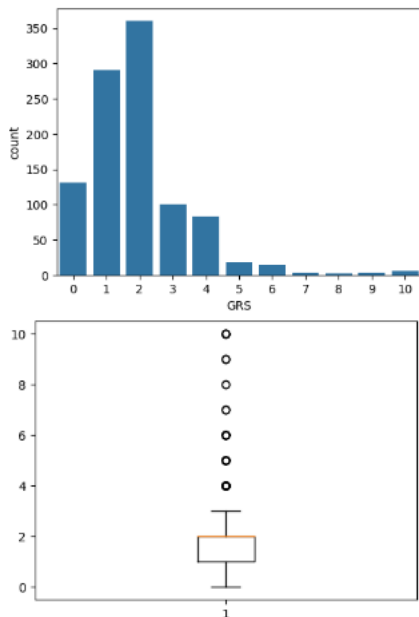
Gambar 5. Distribusi Data Luas Bangunan



Gambar 6. Distribusi Data Jumlah Kamar Tidur



Gambar 7. Distribusi Data Jumlah Kamar Mandi



Gambar 8. Distribusi Data Jumlah Garasi

3.3. Matriks Korelasi

	HARGA	LB	LT	KT	KM	GRS
HARGA	1.0	0.7	0.8	0.3	0.4	0.5
LB	0.7	1.0	0.7	0.4	0.5	0.5
LT	0.8	0.7	1.0	0.4	0.4	0.5
KT	0.3	0.4	0.4	1.0	0.7	0.3
KM	0.4	0.5	0.4	0.7	1.0	0.3
GRS	0.5	0.5	0.5	0.3	0.3	1.0

Gambar 10. Matriks Korelasi

Pada Gambar 10, dapat dilihat bahwa warna korelasi koefisien yang biru tua menunjukkan hubungan linear yang positif yang kuat. Sedangkan warna putih menunjukkan hubungan linear yang negatif yang berarti tidak ada korelasi antar variabel.

3.4. Pemodelan Multiple Linear Regression

Perhitungan *multiple linear regression* dilakukan dengan menggunakan library *sklearn*. Dataset berjumlah 1010 dibagi menjadi menjadi data *training* dan data *testing*, 80% data *training* (808 data) dan 20% data *testing* (202 data). Dalam pemodelan ini menggunakan 5 variabel independen (x) yaitu LB, LT, KT, KM, dan GRS variabel dependen (y) yaitu HARGA.

Berdasarkan persamaan (1) maka telah didapatkan model persamaan *Multiple Linear Regression* pada penelitian ini.

$$y = -775.9292971976356 + 12.28525398X1 + 23.2681398X2 - 638.75382306X3 + 556.61941339X4 + 247.98297331X5$$

3.5. Pengujian Hasil

Setelah mendapatkan persamaan Model *Multiple Linear Regression*. Selanjutnya yaitu Proses evaluasi akurasi dan memprediksi harga sesuai spesifikasi yang ada pada dataset.

Tabel 3. Hasil Akurasi

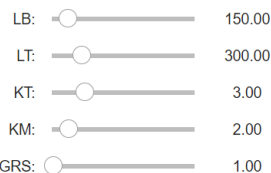
Mean Absolute Error (MAE)	Mean Squared Error (MSE)	Model Accuracy
1980.3555183722162	10675731.022166723	0.7713132526788451

Dari Tabel 3, dapat dilihat bahwa tingkat akurasi diperoleh nilai sebesar 0.77 atau 77%. Hal ini menunjukkan adanya hubungan linier positif yang kuat antara variabel-variabel yang digunakan dalam model *multiple linear regression*.

Pada data hasil Mean Absolute Error (MAE) memiliki nilai sebesar 1980.3555183722162. Dalam hal ini, nilai MAE yang cukup besar menunjukkan bahwa rata-rata kesalahan absolut antara nilai prediksi dan nilai sebenarnya cukup tinggi. Sedangkan nilai Mean Squared Error (MSE) sebesar 10675731.022166723. menunjukkan bahwa terdapat variasi yang signifikan antara nilai prediksi dan nilai sebenarnya.

Selanjutnya, tingkat akurasi dari hasil tersebut akan diuji melalui proses prediksi harga untuk sebuah rumah dengan spesifikasi tertentu.

1. Luas Bangunan = 150 m²
2. Luas Tanah = 300 m²
3. Kamar Tidur = 3
4. Kamar Mandi = 2
5. Garasi = 1



Harga rumah impian anda diperkirakan sekitar IDR 7,492.261 juta

Gambar 11. Prediksi Harga Sesuai Spesifikasi

Hasil perkiraan harga berdasarkan spesifikasi dari Gambar 11, adalah 7,492.261 (7,4 Juta Rupiah).

4. KESIMPULAN DAN SARAN

Faktor dari suatu harga rumah yaitu berupa faktor luas tanah, luas bangunan, jumlah kamar tidur ataupun kamar mandi, dan ada/ tidaknya garasi. Untuk mencari prediksi harga ini, diperlukan metode untuk mendukungnya, dan ini dapat diperoleh

dengan metode *Multiple Linear Regression*. Prediksi dapat ditentukan dengan menggunakan berbagai factor dalam *variable*.

Dalam pengujian menggunakan sampel data sebanyak 1010 baris dengan 6 kolom yang memuat informasi tentang harga rumah, setelah melalui tahap pembersihan data, langkah selanjutnya adalah memisahkan antara dataset pelatihan dan pengujian. Dari dataset tersebut, dilakukan perhitungan akurasi menggunakan Model *Multiple Linear Regression*, mencapai tingkat keakuratan sebesar 77%. Hasil ini dianggap memadai untuk melakukan prediksi harga rumah sesuai dengan spesifikasi yang diinginkan.

Bagi penulis hasil akurasi mungkin belum optimal, sehingga dapat disarankan bahwa diperlukan beberapa prosedur tambahan diantaranya memperbanyak jumlah dataset dan menggunakan metode lainnya di penelitian berikutnya. Sehingga dapat memperoleh hasil yang akurat dan optimal.

DAFTAR PUSTAKA

- Hasanah, L.U., Maula, I. and Tholib, A. (2023) 'Analisis Prediksi Harga Rumah Di Jabodetabek Menggunakan Multiple Linear Regression', *Jurnal Informatika Kaputama (JIK)*, 7(2), pp. 216–224. Available at: <https://doi.org/10.59697/jik.v7i2.135>.
- Kurniawan, Y.D. (2021) 'Analisis Faktor-Faktor Yang Mempengaruhi Pembelian Rumah Di Kota Madiun', *J-MACC: Journal of Management and Accounting*, 4(1), pp. 72–85. Available at: <https://doi.org/10.52166/j-macc.v4i1.2406>.
- Mu'tashim, M.L. *et al.* (2021) 'Analisis Prediksi Harga Rumah Sesuai Spesifikasi Menggunakan Multiple Linear Regression', *Informatik: Jurnal Ilmu Komputer*, 17(3), pp. 238–245.
- Puteri, K. and Silvanie, A. (2020) 'Machine Learning untuk Model Prediksi Harga Sembako', *Jurnal Nasional Informatika*, 1(2), pp. 82–94.
- Sanusi, R.M., Ansori, A.S.R. and Rifki, W. (2020) 'PREDIKSI HARGA RUMAH DI KOTA BANDUNG BAGIAN TIMUR DENGAN MENGGUNAKAN METODE REGRESI', *Electoral Governance Jurnal Tata Kelola Pemilu Indonesia*, 7(3), p. 9381. Available at: <https://talenta.usu.ac.id/politeia/article/view/3955>.
- Saputra Aldi Agusta, Munir, M. and Daffa Rizki, Z.A. (2023) 'Peramalan Pendapatan dari Penjualan Bawang Merah Menggunakan Metode Regresi Linier Berganda', *Stains (Seminar Nasional Teknologi & Sains)*, 2(1), pp. 383–389. Available at: <https://proceeding.unpkediri.ac.id/index.php/stains/article/view/2900>.
- Siregar, R.N.T., Sitorus, V. and Ananta, W.P. (2023) 'Analisis Prediksi Harga Rumah di Bandung

Menggunakan Regresi Linear Berganda', *Journal of Creative Student Research (JCSR)*, 1(6), pp. 395–404. Available at: <https://doi.org/10.55606/jcsrpolitama.v1i6.3038>.