Print ISSN: 2461-1409
Journal Sensi Online ISSN: 2655-5298

Research Article Open Access (CC–BY-SA)

Pembelajaran Mesin Untuk Prediksi Harga Rumah Dengan Metode Regresi Linier Berganda

Junaidi*[1], Novi Cholisoh[2], Deden Roni Nurjaman[3]
[1,2,3] Universitas Raharja; Jl Jend. Sudirman No 40, +6221 5529692
[1,3] Program Studi Teknik Informatika, Fakultas Sains Dan Teknologi, Tangerang
[3] Program Manajemen Retail, Fakultas Ekonomi Dan Bisnis, Tangerang

e-mail: *[1] junaidi@raharja.info, [2] novi@raharja.info, [3] deden.roni@raharja.info

Abstrak

Penelitian ini bertujuan untuk memprediksi harga rumah sesuai fitur yang diinginkan. Fokus penelitian membuat sebuah model pembelajaran mesin berdasarkan dataset yang tersedia, dengan memanfaatkan beberapa fitur yang dibutuhkan untuk memprediksi. Fitur pada dataset yang dijadikan variabel penelitian ini adalah lokasi rumah, jangkauan akses tol, jenis rumah serta jumlah lantai, kamar mandi, kamar tidur sebagai variabel x, dan harga jual sebagai variabel y. Metode yang digunakan adalah regresi linier berganda, dengan jumlah dataset sebanyak sepuluh ribu untuk di observasi. Setelah melalui tahapan pengolahan, kemudian data dibagi menjadi dua bagian: data untuk pelatihan dan pengujian. Penelitian ini menghasilkan Multiple R sebesar 93%, R Square sebesar 86%, Adjusted R Square sebesar 86% dengan standar error sebesar 2%, dengan intercept -0.14, menghasilkan persamaan y = 9E-05x - 0.142 dengan R^2 = 0.0004. Persamaan ini kemudian dilakukan uji coba terhadap data uji yang berbeda dari data latih, dan hasilnya bisa digunakan untuk memprediksi harga rumah, sehingga bisa membantu dengan cepat bagi yang berkepentingan.

Kata kunci— Pembelajaran Mesin, Regresi Linier Berganda, Prediksi

Abstract

This research aims to predict house prices according to the desired features. The research focus is creating a machine learning model based on the available dataset, by utilizing several features needed to predict. The features in the dataset that are used as variables for this research are the location of the house, toll road access range, type of house and number of floors, bathrooms, bedrooms as the x variable, and selling price as the y variable. The method used is multiple linear regression, with a dataset of ten thousand for observation. After going through the processing stages, the data is divided into two parts: data for training and testing. This research produced a Multiple R of 93%, R Square of 86%, Adjusted R Square of 86% with a standard error of 2%, with an intercept of -0.14, producing the equation y = 9E-05x - 0.142 with R^2 = 0.0004. This equation is then tested on test data that is different from the training data, and the results can be used to predict house prices, so that it can help those who are interested.

Keywords— *Machine Learning, Multiple Linear Regression, Prediction*

1. PENDAHULUAN

Tempat tinggal atau rumah merupakan salah satu kebutuhan primer bagi manusia [1], selain itu juga kebutuhan akan rumah sering dijadikan sebagai sumber investasi bagi pengusaha properti yang tentunya menghasilkan investasi yang tidak dapat diremehkan [2], karena Harga setiap tahu rumah semakin meningkat dan mahal [3], itu sebabnya pertimbangan dalam pembelian rumah diperlukan sebuah metode yang membantu untuk menentukan harga yang sesuai dengan keinginan pembeli [4], terkait hal tersebut, masyarakat membutuhkan informasi terkait prediksi harga rumah untuk keperluan perencanaan dimasa yang akan datang [5].

Pada penelitian ini peneliti mengimplementasikan metode multiple linear regression untuk memprediksi harga rumah dengan spesifikasi yang diinginkan [6], dengan mempertimbangkan variabel-variabel yang mempengaruhi [7], sehingga bisa digunakan untuk menganalisa fitur-fitur menggunakan algoritma machine learning yang akurat, dan nantinya digunakan untuk membuat model prediksi harga rumah [8]. Mungkin salah satu algoritma statistik dan pembelajaran mesin yang paling umum dan komprehensif adalah regresi linier, regresi linier digunakan untuk mencari hubungan linier antara satu atau lebih prediktor, regresi linier memiliki dua jenis, regresi sederhana dan regresi berganda [9]. Model yang diusulkan dikembangkan menggunakan data historis [10] dalam 12 (dua belas) tahun terakhir.

Selanjutnya makalah ini disusun dengan metode penelitian pada bagian 2, dilanjutkan hasil dan pembahasan pada bagian 3 dan dirumuskan dengan kesimpulan pada bagian 4, penelitian ini juga menyertakan saran untuk penelitian lanjutan pada bagian 5 dan ditutup dengan ucapan terima kasih dan studi pustaka.

2. METODE PENELITIAN

2.1 Regresi Linier Berganda

Penggunaan statistika dalam mengolah data penelitian akan mempengaruhi tingkat analisis dari suatu hasil penelitian dan akan memperoleh data yang hampir mendekati benar jika memperhatikan analisis regresi yang dipergunakan. Untuk memprediksi dan mengukur besarnya pengaruh suatu variabel bebas (independent/predictor) terhadap variabel tak bebas (dependent/ response) dapat digunakan uji regresi. Analisis regresi merupakan kajian terhadap hubungan satu variabel sebagai variabel yang diterangkan (the explained variable) dengan satu atau dua variabel yang menerangkan (the explanatory). Model regresi linier berganda merupakan suatu persamaan yang menggambarkan hubungan antara dua atau lebih variabel bebas/ predictor (X1, X2,…Xn) dan satu variabel tak bebas/ response (Y). Tujuan dari analisis regresi linier berganda adalah untuk memprediksi nilai variabel tak bebas/ response (Y) jika nilai variabel-variabel bebas/ predictor (X1, X2, ..., Xn) diketahui. Persamaan regresi linier berganda secara matematik diekspresikan oleh :

$$Y = a + b_1X_1 + b_2X_2 + \ldots + b_nX_n$$

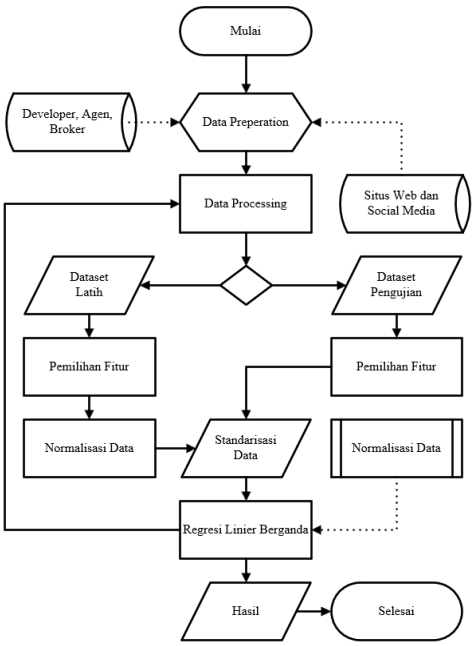
Dimana Y = variabel tak bebas (nilai yang akan diprediksi, a = konstanta, b = koefisien korelasi dan X = variabel bebas [11]. Penelitian ini memiliki 7 (tujuh) variabel bebas dan 1 (satu variabel terikat), sehingga penelitian ini memiliki bentuk persamaan sebagai berikut:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + b_7X_7$$

Dimana Y = Harga Jual, X_1 = lokasi rumah, X_2 = jarak rumah ke tol terdekat, X_3 = jenis rumah, X_4 = jumlah lantai, X_5 = terdapat garasi atau tidak, X_6 = jumlah kamar mandi dan X_7 = jumlah kamar tidur.

2.1 Dataset

Penelitian ini menggunakan teknik pengolahan kuantitatif. Penelitian ini mengolah data harga penjualan rumah yang dikumpulkan secara mandiri dari berbagai sumber, agen properti, penjual serta berbagai situs dan media sosial. Awalnya data terkumpul sebanyak 16.553 data, lalu dilakukan proses data standarisasi data, sehingga data yang berhasil terkumpul sebanyak 10.000 data. Proses pemilihan data dilakukan dengan memperhatikan ketersediaan 7 variabel yang dibutuhkan, yaitu lokasi rumah, jarak menuju akses tol terdekat, jenis lokasi rumah, jumlah lantai, ketersediaan garasi, jumlah kamar tidur dan jumlah kamar mandi. Setelah dataset berhasil disiapkan, maka data ini dibagi menjadi 2, masing-masing sebanyak 5.000 data, untuk digunakan sebagai data latih dan data uji. Berikut adalah alur pengolahan dataset:



Gambar 1 Alur Pengolahan Dataset

Gambar 1 menjelaskan bagaimana sebuah dataset terbentuk, mulai dari tahapan pengumpulan data yang didapat dari berbagai sumber diantaranya developer, agen, broker, situs web dan social media, kemudian data di disatukan sebagai langkah persiapan (data preparation). Pada tahap ini terjadi eliminasi data, data data yang tidak sesuai dengan fitur yang dibutuhkan akan dihapus, termasuk data yang tidak memiliki harga jual juga akan dihapus. Selanjutnya masuk ke dalam tahap pengolahan data (data processing), pada tahap ini data akan dibagi 2 (dua), untuk digunakan sebagai data latih guna pembelajaran mesin, data untuk digunakan sebagai data uji, guna pengujian persamaan yang berhasil dibentuk dengan metode regresi linier berganda.

Setiap dataset baik dataset latih dan dataset uji akan melalui langkah pemilihan fitur, dengan pengambil fitur yang berkaitan dengan penelitian. Sebanyak 8 (delapan) fitur akan digunakan, diantar 8 (delapan) fitur tersebut satu fitur yaitu harga rumah akan dijadikan variabel Y, dan 7 (tujuh) fitur lainnya dijadikan variabel X, mulai dari variabel X_1 sampai dengan variabel X_7. Demikian juga dengan dataset uji, dataset ini hanya memiliki 7 (tujuh) fitur sebagai variabel X, sementara fitur harga jual tidak digunakan dalam data uji ini, karena harga jual merupakan hasil yang akan diprediksi menggunakan persamaan yang berhasil dibentuk. Selanjutnya dataset dimasukkan ke dalam algoritma regresi linier untuk dilakukan pembelajaran mesin, sehingga menghasilkan sebuah persamaan regresi.

3. HASIL DAN PEMBAHASAN

Setelah dataset berhasil disiapkan dalam 2 (jenis) yaitu dataset yang digunakan untuk pembelajaran mesin dan dataset yang digunakan untuk pengujian, selanjutkan dilakukan tahapan pembelajaran mesin menggunakan algoritma regresi linier berganda dan pengujian pembelajaran mesin menggunakan persamaan yang berhasil dibentuk.

3.1 Pembahasan

Sebanyak 5.000 dataset digunakan untuk pembelajaran mesin, dengan memilih fitur yang diperlukan seperti terlihat pada tabel 1 berikut:

Tabel 1 Pengelompokkan dan Pembobotan Fitur Lokasi

FITUR	BOBOT	KETERANGAN
Pusat Kota	3	Berada tepat di dalam pusat kota, kriteria ini ditentukan berdasarkan kedekatan jarak wilayah administrasi pusat pemerintahan kabupaten kota, dan pusat bisnis lainnya.
Tengah Kota	2	Berada tepat di luar pusat kota dan tidak berada dalam area pinggir kota, kriteria ini ditentukan berdasarkan ketersediaan pusat bisnis dan kemudahan akses transportasi dari tengah kota ke pusat kota.
Pinggir Kota	1	Tidak berada dalam area pusat kota dan tengah kota, kriteria ini ditentukan berdasarkan ketersediaan pusat bisnis yang tidak sebanyak area tengah kota, dan kondisi akses dari pusat kota, tengah kota ke pinggir kota.

Tabel 1 merupakan pengelompokkan dan pemberian bobot terhadap suatu lokasi wilayah dimana rumah berada, terdapat 3 (tiga) pengelompokkan (kolom fitur) yaitu pusat kota dengan bobot 3 (tiga), tengah kota dengan bobot 2 (dua) dan pinggir kota dengan bobot 1 (satu), penjelasan dari pengelompokkan ini terdapat pada kolom keterangan.

Tabel 2 Pengelompokkan dan Pembobotan Fitur Akses Tol

FITUR	BOBOT	KETERANGAN
Dekat	3	Berada dengan lokasi yang tidak jauh dengan pintu akses tol masuk maupun pintu akses tol keluar, pengelompokkan ini didasarkan pada jarak dengan radius 5 km sd 10 km.
Sedang	2	Berada dengan lokasi yang tidak jauh dengan pintu akses tol masuk maupun pintu akses tol keluar, pengelompokkan ini didasarkan pada jarak dengan radius 10 km sd 15 km.
Jauh	1	Berada dengan lokasi yang tidak jauh dengan pintu akses tol masuk maupun pintu akses tol keluar, pengelompokkan ini didasarkan pada jarak dengan radius melebihi 15 km.

Tabel 2 merupakan pengelompokkan dan pemberian bobot terhadap suatu lokasi dari rumah ke akses pintu tol terdekat baik masuk maupun keluar, terdapat 3 (tiga) pengelompokkan (kolom fitur) yaitu dekat dengan bobot 3 (tiga), sedang dengan bobot 2 (dua) dan jauh dengan bobot 1 (satu), penjelasan dari pengelompokkan ini terdapat pada kolom keterangan.

Tabel 3 Pengelompokkan dan Pembobotan Fitur Jenis Rumah

FITUR	BOBOT	KETERANGAN
Elit	3	Pengelompokkan rumah yang berada di kawasan khusus atau elit yang memiliki konsep kota dalam kota.
Perumahan	2	Pengelompokkan rumah yang berada di kawasan perumahan yang dikembangkan oleh developer baik rumah subsidi maupun rumah non subsidi.
Perkampungan	1	Pengelompokkan rumah yang berada di luar kelompok elit dan perumahan.

Tabel 3 merupakan pengelompokkan dan pemberian bobot terhadap suatu jenis rumah berada pada suatu kawasan tertentu, terdapat 3 (tiga) pengelompokkan (kolom fitur) yaitu elit dengan bobot 3 (tiga), perumahan dengan bobot 2 (dua) dan perkampungan dengan bobot 1 (satu), penjelasan dari pengelompokkan ini terdapat pada kolom keterangan.

Tabel 4 Pengelompokkan dan Pembobotan Fitur Jumlah Lantai

FITUR	BOBOT	KETERANGAN
Tiga Lantai	3	Pengelompokkan rumah bertingkat dengan minimal jumlah lantai adalah 3, rumah yang memiliki jumlah lantai lebih dari tiga masuk dalam kelompok ini.
Dua Lantai	2	Pengelompokkan rumah bertingkat dengan minimal jumlah lantai adalah 2, rumah yang memiliki lantai ke 3 sebagai dak termasuk dalam kelompok ini.
Satu Lantai	1	Pengelompokkan rumah tidak bertingkat, rumah yang memiliki lantai ke 2 sebagai dak termasuk dalam kelompok ini.

Tabel 4 merupakan pengelompokkan dan pemberian bobot terhadap suatu jenis rumah bertingkat atau tidak, terdapat 3 (tiga) pengelompokkan (kolom fitur) yaitu tiga lantai dengan bobot 3 (tiga), dua lantai dengan bobot 2 (dua) dan tidak bertingkat dengan bobot 1 (satu), penjelasan dari pengelompokkan ini terdapat pada kolom keterangan.

Tabel 5 Pengelompokkan dan Pembobotan Fitur Ketersediaan Garasi

FITUR	BOBOT	KETERANGAN
Garasi Tertutup	3	Pengelompokkan rumah dengan ketersediaan garasi tertutup yang memiliki pintu sebagai garasi khusus.

Garasi Terbuka	2	Pengelompokkan rumah dengan ketersediaan garasi terbuka yang tidak memiliki pintu namun menyatu dengan atap rumah.
Garasi Lahan	1	Pengelompokkan rumah dengan garasi yang tidak disediakan khusus, yang bisa digunakan masuk dalam kategori ini

Tabel 5 merupakan pengelompokkan dan pemberian bobot terhadap ketersediaan garasi, terdapat 3 (tiga) pengelompokkan (kolom fitur) yaitu garasi tertutup dengan bobot 3 (tiga), garasi terbuka dengan bobot 2 (dua) dan garasi lahan dengan bobot 1 (satu), penjelasan dari pengelompokkan ini terdapat pada kolom keterangan.

Tabel 6 Pengelompokkan dan Pembobotan Fitur Kamar Mandi

FITUR	BOBOT	KETERANGAN
Jenis Mewah	3	Pengelompokkan rumah dengan ketersediaan jumlah dan jenis kamar mandi, kategori ini termasuk fasilitas di dalam nya seperti cermin rias, bathup dan shower dan closet duduk yang memiliki ruang khusus didalam kamar mandi.
Jenis Sedang	2	Pengelompokkan rumah dengan ketersediaan jumlah dan jenis kamar mandi, kategori ini termasuk fasilitas di dalam nya seperti cermin rias, shower dan closet duduk
Jenis Biasa	1	Pengelompokkan rumah dengan ketersediaan jumlah dan jenis kamar mandi, kategori ini memiliki batasan fitur hanya pada shower dan closet duduk atau jongkok.

Tabel 6 merupakan pengelompokkan dan pemberian bobot terhadap jumlah dan jenis kamar mandi, terdapat 3 (tiga) pengelompokkan (kolom fitur) yaitu jenis mewah dengan bobot 3 (tiga), jenis sedang dengan bobot 2 (dua) dan jenis biasa dengan bobot 1 (satu), penjelasan dari pengelompokkan ini terdapat pada kolom keterangan.

Tabel 7 Pengelompokkan dan Pembobotan Fitur Kamar Tidur

FITUR	BOBOT	KETERANGAN
Jenis 6	6	Pengelompokkan rumah dengan ketersediaan 6 kamar tidur, jenis ini memungkinkan untuk rumah dengan tipe bertingkat, atau rumah tanpa tingkat namun memiliki luas tanah yang memadai.
Jenis5	5	Pengelompokkan rumah dengan ketersediaan 5 kamar tidur, jenis ini memungkinkan untuk rumah dengan tipe bertingkat, atau rumah tanpa tingkat namun memiliki luas tanah yang memadai.

Jenis 4	4	Pengelompokkan rumah dengan ketersediaan 4 kamar tidur, jenis ini memungkinkan untuk rumah dengan tipe bertingkat, atau rumah tanpa tingkat namun memiliki luas tanah yang memadai.
Jenis 3	3	Pengelompokkan rumah dengan ketersediaan 3 kamar tidur, jenis ini memungkinkan untuk rumah tanpa tingkat, atau memiliki luas tanah yang memadai.
Jenis 2	2	Pengelompokkan rumah dengan ketersediaan 2 kamar tidur, jenis ini memungkinkan untuk rumah tanpa tingkat, atau memiliki luas tanah yang memadai.
Jenis 1	1	Pengelompokkan rumah dengan ketersediaan 1 kamar tidur, jenis ini memungkinkan untuk rumah tanpa tingkat, atau memiliki luas tanah yang kurang memadai.

Tabel 7 merupakan pengelompokkan dan pemberian bobot terhadap jumlah kamar tidur, terdapat 6 (enam) pengelompokkan (kolom fitur) yaitu jenis 6 dengan bobot 6 (enam), jenis 5 dengan bobot 5 (lima), jenis 4 dengan bobot 4 (empat), jenis 3 dengan bobot 3 (tiga), jenis 2 dengan bobot 2 (dua) dan jenis 1 dengan bobot 1 (satu), penjelasan dari pengelompokkan ini terdapat pada kolom keterangan.

Seluruh fitur tersebut diterapkan ke dalam dataset untuk pemberian bobot yang disesuaikan agar bisa dimasukan ke dalam algoritma regresi linier berganda sebagai bentuk pembelajaran mesin. Tabel 8 adalah tabel dari perwakilan dataset yang digunakan, terlihat data yang ditampilkan adalah 10 (sepuluh) data pertama dan 10 (sepuluh), total data ini secara keseluruhan berjumlah 5.000 data. Terdapat kolom No yang mewakili nomor data, kemudian kolom X1 yang mewakili variabel X_1 untuk mewakili bobot pada Tabel 1 (fitur lokasi), kolom X2 yang mewakili variabel X_2 untuk mewakili bobot pada Tabel 2 (fitur akses tol), kolom X3 yang mewakili variabel X_3 untuk mewakili bobot pada Tabel 3 (fitur jenis rumah),kolom X4 yang mewakili variabel X_4 untuk mewakili bobot pada Tabel 4 (fitur jumlah lantai),kolom X5 yang mewakili variabel X_5 untuk mewakili bobot pada Tabel 5 (fitur ketersediaan garasi), kolom X6 yang mewakili variabel X_6 untuk mewakili bobot pada Tabel 6 (fitur kamar mandi), kolom X7 yang mewakili variabel X_7 untuk mewakili bobot pada Tabel 7 (fitur kamar tidur). Nilai pada tabel X1 sampai dengan X7 merupakan nilai bobot yang telah diperkecil dengan pembagian 10.000 sebagai bentuk normalisasi data dan memudahkan perhitungan.

Masih pada Tabel 8, kolom paling kanan merupakan variabel Y, dimana variabel ini mewakili harga rumah pada setiap baris data, nilai yang tampil adalah nilai yang telah dinormalisasi dengan pembagian 10.000 setelah dijadikan dalam satuan juta. Data ini kemudian diolah dengan regresi linier berganda. Kolom bantu dibuat untuk menggambarkan algoritma ini secara visual dengan melakukan perkalian X_1 terhadap Y, kemudian X_1 terhadap X_2 sampai dengan X_1 terhadap X_7, Kemudian mengalikan X_2 dengan Y, kemudian X_2 terhadap X_3 sampai dengan X_2 terhadap X_7, kemudian mengalikan X_3 dengan Y, kemudian X_3 terhadap X_4 sampai dengan X_3 terhadap X_7, kemudian mengalikan X_4 dengan Y, kemudian X_4 terhadap X_5 sampai dengan X_4 terhadap X_7, mengalikan kemudian X_5 terhadap Y, kemudian X_5 terhadap X_6 dan X_5 terhadap X_7, kemudian mengalikan X_6 dengan Y dan X_6 terhadap X_7. Dilanjutkan Mengkuadratkan masing setiap variabel X_1^2, X_2^2, X_3^2, X_4^2, X_5^2, X_6^2, X_7^2, dan Y_1^2, Setelah itu, dilakukan penjumlahan setiap kolom yang terbentuk.

Tabel 8 Dataset Latih Dengan Pembobotan

No	X1	X2	X3	X4	X5	X6	X7	Y
1	0.00030	0.00010	0.00030	0.00020	0.00020	0.00010	0.00050	0.23200
2	0.00020	0.00010	0.00030	0.00030	0.00020	0.00020	0.00060	0.22700
3	0.00030	0.00030	0.00030	0.00010	0.00030	0.00010	0.00020	0.21150
4	0.00020	0.00010	0.00020	0.00020	0.00010	0.00010	0.00030	0.09850
5	0.00010	0.00030	0.00020	0.00010	0.00020	0.00020	0.00020	0.10850
6	0.00010	0.00010	0.00030	0.00020	0.00010	0.00010	0.00030	0.12150
7	0.00010	0.00010	0.00010	0.00010	0.00010	0.00020	0.00030	0.04000
8	0.00030	0.00020	0.00020	0.00030	0.00030	0.00030	0.00060	0.21350
9	0.00030	0.00020	0.00010	0.00030	0.00020	0.00020	0.00040	0.15800
10	0.00030	0.00020	0.00010	0.00020	0.00010	0.00010	0.00040	0.10350
4991	0.00020	0.00030	0.00030	0.00020	0.00020	0.00020	0.00040	0.21900
4992	0.00030	0.00030	0.00010	0.00010	0.00020	0.00010	0.00020	0.13150
4993	0.00010	0.00020	0.00030	0.00030	0.00030	0.00020	0.00050	0.17850
4994	0.00020	0.00010	0.00020	0.00010	0.00030	0.00020	0.00020	0.10800
4995	0.00030	0.00010	0.00010	0.00030	0.00010	0.00020	0.00060	0.14000
4996	0.00030	0.00030	0.00010	0.00020	0.00010	0.00010	0.00030	0.10750
4997	0.00020	0.00010	0.00020	0.00010	0.00010	0.00020	0.00030	0.07850
4998	0.00030	0.00020	0.00030	0.00010	0.00010	0.00020	0.00020	0.15400
4999	0.00010	0.00030	0.00010	0.00030	0.00030	0.00020	0.00050	0.22550
5000	0.00010	0.00010	0.00010	0.00010	0.00010	0.00010	0.00030	0.03550

Formula untuk menggambarkan paparan ini, dapat dinotasikan dalam bentuk formula matematis sebagai berikut

$$a = \frac{(\Sigma Y) - (b_1 \times \Sigma x_1) - (b_2 \times \Sigma x_2)}{n}$$

$$b_1 = \frac{[(\Sigma x_2^2 \times \Sigma x_1 y) - (\Sigma x_2 y \times \Sigma x_1 x_2)]}{[(\Sigma x_1^2 \times \Sigma x_2^2) - (\Sigma x_1 \times x_2)^2]}$$

$$b_2 = \frac{[(\Sigma x_1^2 \times \Sigma x_2 y) - (\Sigma x_1 y \times \Sigma x_1 x_2)]}{[(\Sigma x_1^2 \times \Sigma x_2^2) - (\Sigma x_1 \times x_2)^2]}$$

Gambar 2 Konstanta a dan koefisien-koefisien regresi b1 dan b2 d

Metode alternatif, yaitu metode matriks (metode kuadrat terkecil) a, b1 dan b2 dapat ditentukan dengan menyusun suatu persamaan :

$$\begin{aligned} an &+ b_1\sum X_1 &+ b_2\sum X_2 &= \sum Y \\ a\sum X_1 &+ b_1\sum X_1^2 &+ b_2\sum X_1X_2 &= \sum X_1Y \\ a\sum X_2 &+ b_1\sum X_2X_1 &+ b_2\sum X_2^2 &= \sum X_2Y \end{aligned}$$

Gambar 3 metode matriks (metode kuadrat terkecil) a, b1 dan b2

3.2 Hasil

Tabel 8 merupakan dataset yang sudah selesai diolah, dan setelah diimplementasikan kedalam algoritma regresi linear berganda, data terdiri dari 9 (sembilan) kolom, kolom 1 (satu) sebagai nomor id, kolom 2 (dua) sampai dengan kolom kolom 8 (delapan) merupakan data variabel bebas X_1 sampai dengan X_7, dan kolom terakhir sebagai variabel Y, hasilnya sebagai berikut:

SUMMARY OUTPUT PREDIKSI HARGA RUMAH

Regression Statistics	
Multiple R	0.925384022
R Square	0.856335588
Adjusted R Square	0.856134136
Standard Error	0.022930164
Observations	5000

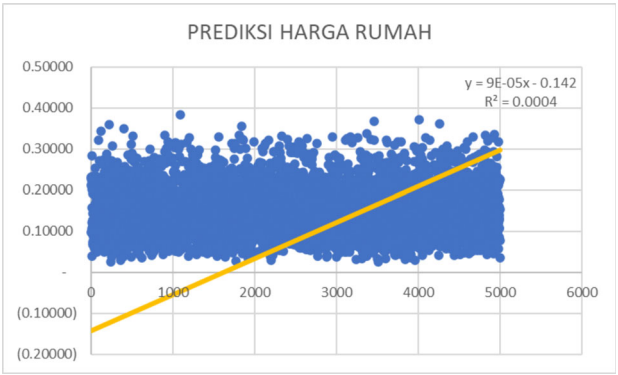
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	7	15.64529216	2.235041738	4250.806452	0
Residual	4992	2.624755674	0.000525792		
Total	4999	18.27004784			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-0.142023707	0.001966357	-72.22680997	0	-0.145878631	-0.138168783	-0.145878631	-0.138168783
lokasi	177.8361528	3.963088046	44.87312691	0	170.0667592	185.6055464	170.0667592	185.6055464
tol	176.8663983	3.971580844	44.53299712	0	169.080355	184.6524415	169.080355	184.6524415
jenis	416.491543	3.933068907	105.8947994	0	408.7810001	424.2020859	408.7810001	424.2020859
lantai	276.8634545	7.530165871	36.76724514	4.0929E-262	262.1010213	291.6258877	262.1010213	291.6258877
garasi	173.5110633	3.948443724	43.94416519	0	165.770379	181.2517476	165.770379	181.2517476
kamar mandi	186.2400992	5.849417308	31.83908574	1.0771E-202	174.7726715	197.7075268	174.7726715	197.7075268
kamar tidur	43.90336017	4.465063006	9.832640685	1.30548E-22	35.14987512	52.65684522	35.14987512	52.65684522

Gambar 4 Hasil Analisis Regresi Linier Berganda

Dapat dilihat pada gambar 4 Hasil penelitian ini menghasilkan Multiple R sebesar 93%, R Square sebesar 86%, Adjusted R Square sebesar 86% dengan standar error sebesar 2%, dengan intercept -0.14. Informasi yang didapat bahwa nilai korelasi X dan Y sebesar 93% dan ini masuk ke dalam kategori sangat kuat, hal ini mampu menjelaskan bawah harga jual rumah dipengaruhi oleh variabel bebas seperti lokasi, akses tol, jenis rumah, jumlah lantai, ketersediaan garasi, jumlah kamar tidur, jumlah kamar mandi dan sisanya dipengaruhi oleh faktor variabel lain. Berikut grafik (gambar 5) yang berhasil terbentuk dari penggunaan metode regresi linier berganda yang terlihat pada gambar 5:



Gambar 5 Hasil Analisis Regresi Linier Berganda

Sehingga model persamaan regresi linier sederhananya adalah: Y = 9E - 0.5x - 0.142 dan R^2 = 2. Terlihat pada gambar 5 persamaan diatas menghasilkan garis regresi linier yang menaik, hal ini dapat menggambarkan pengaruh variabel bebas terhadap harga jual rumah.

4. KESIMPULAN

Berdasarkan dataset yang kemudian diolah untuk memberikan bobot, kemudian direpresentasikan dengan regresi linier berganda sebagai salah satu metode algoritma pembelajaran mesin, dapat ambil kesimpulan sebagai berikut:

1. Nilai Multiple R sebesar 93%, R Square sebesar 86%, Adjusted R Square sebesar 86% dengan standar error sebesar 2%, dengan intercept -0.14. Informasi yang didapat bahwa nilai korelasi X dan Y sebesar 93% dan ini masuk ke dalam kategori sangat kuat.
2. Model persamaan regresi linier sederhananya adalah: Y = 9E - 0.5x - 0.142 dan R^2 = 2.
3. Hasil yang didapat dapat dikatakan bahwa hubungan variabel bebas / prediktor X dengan variabel terikat / response Y adalah sangat kuat.
4. Hasil ini juga mampu menerangkan bahwa, 93% variabel bebas / prediktor X dapat menerangkan / menjelaskan variabel tak bebas / response Y, dan 7% dijelaskan oleh variabel lainnya.
5. Hubungan antara beberapa fitur yang dibahas dengan harga jual memiliki keterkaitan, dimana semakin tinggi bobot setiap fitur, berdampak pada semakin tinggi harga jual yang dihasilkan, ini menggambarkan bahwa harga jual sebuah rumah bisa diprediksi berdasarkan fitur yang ada.

5. SARAN

Penelitian ini dapat dijadikan dasar pembelajaran mesin untuk memprediksi harga jual rumah berdasarkan kombinasi beberapa fitur yang diinginkan. Pemanfaatan data penjualan rumah dengan beberapa fitur yang memenuhi kebutuhan untuk memprediksi harga jual rumah, bukanlah satu satunya data yang bisa dijadikan ukuran, data berikutnya yang bisa dipertimbangkan untuk bisa dimasukkan kedalam persamaan yang dihasilkan dalam penelitian ini dapat disesuaikan dengan kebutuhan.

DAFTAR PUSTAKA

[1] Saiful, Andi., "Prediksi harga rumah menggunakan web scraping dan machine learning dengan algoritma linear regression", JATISI (Jurnal Teknik Informatika dan Sistem Informasi), Vol. 8, No. 1, pp.41-50, 2021. DOI: https://doi.org/10.35957/jatisi.v8i1.701

[2] Rahayuningtyas, Evi Febrion, Feny Novia Rahayu, and Yufis Azhar. "Prediksi harga rumah menggunakan general regression neural network", Jurnal Informatika, vol. 8, No. 1, pp. 59-66, 2021.

[3] Mu'tashim, Muhammad Labib, et al., "Analisis prediksi harga rumah sesuai spesifikasi menggunakan multiple linear regression", Informatik: Jurnal Ilmu Komputer, Vo. 17, No. 3, pp. 238-245, 2021. DOI: https://doi.org/10.52958/iftk.v17i3.3635

[4] Ridho, Ihda Innar, et al., "Metode Neural Network Untuk Penentuan Akurasi Prediksi Harga Rumah", Technologia: Jurnal Ilmiah, vol. 13, No. 1, pp.56-58, 2022. DOI: http://dx.doi.org/10.31602/tji.v13i1.6252

[5] Fitri, Evita., "Analisis Perbandingan Metode Regresi Linier, Random Forest Regression Dan Gradient Boosted Trees Regression Method Untuk Prediksi Harga Rumah", Journal of Applied Computer Science and Technology, vol. 4, no. 1, pp. 58-64, 2023. DOI: https://doi.org/10.52158/jacost.v4i1.491

[6] Maula, Inayatul, Linda Uswatun Hasanah, and Abu Tholib, "Analisis Prediksi Harga Rumah Di JABODETABEK Menggunakan Multiple Linear Regression", Jurnal

Informatika Kaputama (JIK), vol. 7, no. 2, pp. 216-224, 2023. DOI: https://doi.org/10.59697/jik.v7i2.135

[7] Vermaysha, Ashary, and Nurmalitasari Nurmalitasari, "Prediksi Harga Rumah di Kabupaten Karanganyar Menggunakan Metode Regresi Linear", Prosiding Seminar Nasional Teknologi Informasi dan Bisnis, 2023.

[8] Mustakim, Muhamad Imam, "Analisis Prediksi Harga Rumah Dengan Menggunakan Algoritma Machine Learning", DISS:Universitas Mercu Buana Jakarta, 2021.

[9] Hope, Thomas MH, "Linear regression", Machine Learning. Academic Press, p. 67-81, 2020. DOI: https://doi.org/10.38094/jastt1457

[10] Jumin, Ellysia, et al., "Machine learning versus linear regression modeling approach for accurate ozone concentrations prediction", Engineering Applications of Computational Fluid Mechanics, vol. 14, no. 1, p. 713-725,2020. DOI: https://doi.org/10.1080/19942060.2020.1758792

[11] I Made Yuliara, "Modul Regresi Linier Berganda", Fakultas Matematika dan Ilmu Pengetahuan Alam. Universitas Udayana, p. 1-12, 2016.