

Prediksi Harga Rumah menggunakan Machine Learning Algoritma Regresi Linier

Rosalia Roja Hallan^a, Ika Nur Fajri^b

^aProgram Studi Sistem Informasi, Ilmu Komputer, Universitas Amikom Yogyakarta, rosaliarajahallan@students.amikom.ac.id

^bfajri@amikom.ac.id

Submitted: 12-12-2024, Reviewed: 24-12-2024, Accepted 05-1-2025

<https://doi.org/10.47233/jteksis.v7i1.1732>

Abstract

The property sector plays a vital role in the global economy, especially regarding property price prediction, which is a complex challenge influenced by factors such as building size, number of rooms, location, and property condition. This study aims to build a property price prediction model using the Linear Regression algorithm. The data used in this research was obtained from Kaggle, consisting of 1460 data points on house prices in Ames, USA. The preprocessing phase includes handling missing data, outlier management, and feature standardization using StandardScaler to ensure data consistency. The linear regression model was trained and evaluated using R-squared (R^2) and Root Mean Squared Error (RMSE) metrics. The evaluation results show an R^2 of 0.81, indicating the model explains 81% of the variation in house prices. Additionally, the RMSE value of 35,830.40 shows the model's relatively low and consistent error when tested with different data. Features such as overall house quality (OverallQual) and living area size (GrLivArea) significantly correlate with house prices. These findings demonstrate that linear regression is an effective tool for predicting property prices.

Keywords: Property price prediction, Linear Regression, StandardScaler, R-squared, Root Mean Squared Error, Machine learning

Abstrak

Sektor properti memiliki peran penting dalam perekonomian global, terutama dalam prediksi harga properti yang merupakan tantangan kompleks yang dipengaruhi oleh berbagai faktor seperti ukuran bangunan, jumlah kamar, lokasi, dan kondisi properti. Penelitian ini bertujuan untuk membangun model prediksi harga properti menggunakan algoritma Regresi Linier. Data yang digunakan dalam penelitian ini diperoleh dari Kaggle, yang mencakup 1460 data mengenai harga rumah di Ames, Amerika Serikat. Proses pra-pemrosesan mencakup pengelolaan data yang hilang, penanganan outlier, dan standarisasi fitur menggunakan StandardScaler untuk memastikan konsistensi data. Model regresi linier dilatih dan dievaluasi menggunakan metrik R-squared (R^2) dan Root Mean Squared Error (RMSE). Hasil evaluasi menunjukkan R^2 sebesar 0.81, yang menunjukkan model dapat menjelaskan 81% variasi harga rumah. Selain itu, nilai RMSE sebesar 35,830.40 menunjukkan bahwa model memiliki kesalahan yang relatif rendah dan konsisten saat diuji dengan data yang berbeda. Fitur-fitur seperti kualitas rumah secara keseluruhan (OverallQual) dan luas ruang hidup (GrLivArea) memiliki korelasi signifikan dengan harga rumah. Temuan ini menunjukkan bahwa regresi linier adalah alat yang efektif untuk memprediksi harga properti.

Keywords: Prediksi harga properti, Regresi Linier, StandardScaler, R-squared, Root Mean Squared Error, Machine learning

This work is licensed under Creative Commons Attribution License 4.0 CC-BY International license



PENDAHULUAN

Properti merupakan salah satu sektor ekonomi yang memiliki peran penting dalam kehidupan manusia [1]. Pertumbuhan kebutuhan akan hunian yang layak seiring dengan meningkatnya populasi menjadikan sektor ini sebagai salah satu investasi yang strategis [2]. Penentuan harga properti menjadi tantangan yang kompleks karena dipengaruhi oleh berbagai faktor seperti luas bangunan, jumlah kamar, lokasi, serta kondisi properti itu sendiri [3] [4]. Oleh karena itu, dibutuhkan metode yang efektif untuk membantu memprediksi harga properti secara akurat guna mendukung pengambilan keputusan baik oleh pembeli, penjual, maupun investor.

Dalam beberapa dekade terakhir, perkembangan teknologi dan ketersediaan data besar (big data) telah mendorong penggunaan metode berbasis machine learning untuk

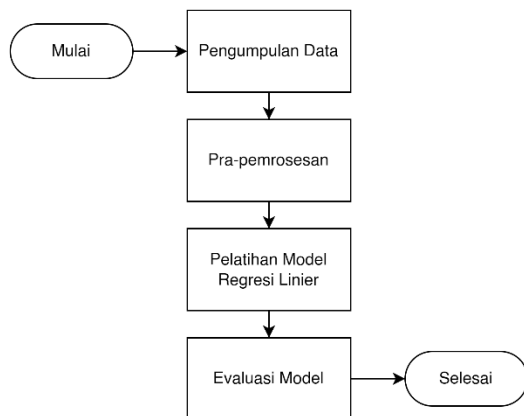
menyelesaikan berbagai permasalahan [5] termasuk prediksi harga properti [6]. Machine learning memungkinkan komputer untuk belajar dari data historis dan membuat prediksi tanpa harus diprogram secara eksplisit [7]. Salah satu algoritma yang sering digunakan dalam masalah prediksi adalah Linear Regression, yang terkenal karena kesederhanaannya dan kemampuannya dalam memodelkan hubungan antara variabel independen dan dependen [8] [9] yang akan dievaluasi menggunakan metrik evaluasi untuk hasil yang lebih optimal dan akurat [10].

Penelitian ini bertujuan untuk membangun model prediksi harga properti menggunakan algoritma Linear Regression. Model ini diharapkan mampu menganalisis pengaruh berbagai faktor, seperti luas bangunan, jumlah kamar, dan kondisi properti, terhadap harga. Selain itu, penelitian ini

juga mengevaluasi performa model yang dihasilkan menggunakan metrik evaluasi seperti Root Mean Squared Error (RMSE) dan R-squared (R^2). Dengan adanya model ini, diharapkan dapat memberikan kontribusi dalam penyelesaian permasalahan prediksi harga properti yang lebih akurat dan efisien.

Penelitian ini memberikan fokus pada analisis fitur-fitur yang paling berpengaruh terhadap harga properti serta bagaimana algoritma Linear Regression dapat digunakan sebagai pendekatan yang efektif dalam prediksi harga. Selain itu, penelitian ini mengesampingkan faktor lokasi geografis karena keterbatasan data yang spesifik untuk wilayah Amerika Serikat, sehingga model ini dirancang untuk menggeneralisasi pola-pola yang memengaruhi harga properti secara umum.

METODE PENELITIAN



Gambar 1. Metode Penelitian

Untuk memprediksi harga rumah, penelitian ini menggunakan pendekatan kuantitatif. Berdasarkan Gambar 1, proses penelitian terdiri dari beberapa tahapan, yaitu pengumpulan data yang mencakup variabel-variabel terkait, pengolahan dan pembersihan data, eksplorasi data, penerapan algoritma regresi linear terhadap data yang telah diproses, dan akhirnya, analisis terhadap hasil yang diperoleh. Berikut adalah rincian dari setiap langkah dalam penelitian ini.

2.1. Pengumpulan Data

Pada tahap ini, data yang digunakan diperoleh dari situs Kaggle. Data tersebut mencakup informasi mengenai harga dan spesifikasi rumah yang terletak di kota Ames, Amerika Serikat, yang telah terjual. Terdapat total 1460 data yang digunakan dalam penelitian ini.

2.2. Pra-pemrosesan

Pra-pemrosesan data adalah langkah awal yang penting dalam membangun model machine learning,

termasuk untuk prediksi harga rumah. Tahapan ini mencakup penghapusan data yang hilang, penanganan outlier, standarisasi fitur, dan konversi variabel kategori menjadi format numerik [11]. Dalam hal ini, StandardScaler digunakan untuk menstandarisasi fitur-fitur numerik [12], seperti luas bangunan, jumlah kamar, atau usia rumah. Proses standarisasi dilakukan dengan persamaan:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

Persamaan (1) di mana x adalah nilai asli fitur, μ adalah rata-rata, dan σ adalah standar deviasi fitur tersebut.

Standarisasi penting dalam prediksi harga rumah karena fitur-fitur yang memiliki skala berbeda, seperti harga dan luas tanah, dapat mempengaruhi model secara tidak proporsional. Dengan standarisasi, model machine learning seperti regresi linier atau algoritma berbasis gradient descent dapat bekerja lebih efisien, mempercepat konvergensi, dan menghasilkan prediksi yang lebih akurat [13].

2.3. Pelatihan Model Regresi Linier

Regresi linear adalah salah satu algoritma machine learning yang digunakan untuk memodelkan hubungan antara satu atau lebih variabel independen (X) dan variabel dependen (Y) [14]. Model regresi linear dapat direpresentasikan dalam bentuk persamaan berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon \quad (2)$$

Persamaan (2) dimana Y adalah variabel dependen (harga properti dalam penelitian ini). β_0 adalah intercept (konstanta). $\beta_1, \beta_2, \dots, \beta_n$ adalah koefisien regresi yang menunjukkan pengaruh setiap variabel independen terhadap variabel dependen. X_1, X_2, \dots, X_n adalah variabel independen (seperti luas bangunan, jumlah kamar, dll.). ϵ adalah error term, yang menunjukkan deviasi antara nilai aktual dan nilai prediksi.

Data dibagi menjadi dua bagian, yaitu data pelatihan dan data uji. Model kemudian dilatih menggunakan data pelatihan untuk mengenali pola dan hubungan antara fitur-fitur yang ada dengan harga properti.

Regresi linear sering digunakan karena kesederhanaannya dan interpretabilitasnya. Algoritma ini bekerja dengan meminimalkan Root Mean Squared Error (RMSE) untuk menemukan garis terbaik yang dapat menjelaskan hubungan antara variabel [15] [16].

2.4. Evaluasi Model

Model yang telah dilatih kemudian dievaluasi menggunakan beberapa metrik evaluasi, seperti R-squared (R^2) dan Root Mean Squared Error (RMSE).

R-squared (R^2) adalah metrik lain yang digunakan untuk mengevaluasi model regresi. R^2 mengukur proporsi variasi dalam variabel dependen (Y) yang dapat dijelaskan oleh variabel independen (X) dalam model [17]. R^2 dinyatakan dalam persamaan berikut:

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (3)$$

Persamaan (3) dimana \bar{Y} adalah rata-rata dari Y. Y_i adalah nilai aktual. \hat{Y}_i adalah nilai prediksi. Nilai R^2 berada dalam rentang 0 hingga 1. Nilai yang mendekati 1 menunjukkan bahwa model dapat menjelaskan variasi target dengan baik.

Root Mean Squared Error (RMSE) adalah metrik yang digunakan untuk mengevaluasi akurasi model regresi dengan mengukur rata-rata kesalahan prediksi antara nilai aktual dan nilai prediksi yang dihasilkan oleh model [18]. RMSE menghitung akar kuadrat dari rata-rata kuadrat selisih antara nilai aktual dan nilai prediksi [19]. RMSE memberikan informasi tentang seberapa besar kesalahan model dalam satuan yang sama dengan variabel dependen. RMSE dihitung dengan persamaan berikut:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (4)$$

Persamaan (4) dimana Y_i adalah nilai aktual dari data. \hat{Y}_i adalah nilai prediksi yang dihasilkan oleh model. n adalah jumlah data sampel. Nilai RMSE yang lebih kecil menunjukkan bahwa model memiliki prediksi yang lebih akurat, karena semakin kecil nilai kesalahan rata-rata, semakin baik model dalam mengestimasi nilai target untuk RMSE memberikan gambaran tentang seberapa besar kesalahan yang dihasilkan oleh model [20] dalam satuan yang sama dengan variabel target (prediksi harga rumah dalam USD), sehingga memudahkan interpretasi hasil evaluasi model.

Tujuannya adalah untuk mengukur sejauh mana model dapat memprediksi harga rumah dengan akurat.

HASIL DAN PEMBAHASAN

Pada tahap ini, data yang digunakan diperoleh dari situs web Kaggle. Data tersebut mencakup informasi mengenai harga rumah yang

terletak di kota Ames, Amerika Serikat, yang telah terjual.

(1460, 81)

PoolArea	PoolQC	Fence	MiscFeature	MiscVal	MoSold	YrSold	SaleType	SaleCondition	SalePrice
0	NaN	NaN	NaN	0	2	2008	WD	Normal	208500
0	NaN	NaN	NaN	0	5	2007	WD	Normal	181500
0	NaN	NaN	NaN	0	9	2008	WD	Normal	223500
0	NaN	NaN	NaN	0	2	2006	WD	Abnorml	140000
0	NaN	NaN	NaN	0	12	2008	WD	Normal	250000

Gambar 2. Data Harga Rumah pada kota Ames

Berdasarkan Gambar 2, data terdiri dari beberapa variabel yang dijadikan sebagai variabel X (dependen), serta satu variabel sebagai variabel Y (independen). Terdapat total 1460 data dan 81 kolom yang digunakan, dengan variabel SalePrice yang berisi nilai harga dalam satuan USD.

Berikut adalah data yang telah ditampilkan bersama statistiknya, mulai dari nilai rata-rata (mean), nilai minimum, hingga nilai maksimum harga rumah dalam USD, beserta data dependen yang akan digunakan nantinya.

	SalePrice	YrSold	FullBath	GarageArea	PoolArea	YearBuilt	OverallCond	OverallQual
count	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000
mean	180921.195890	2007.815753	1.565068	472.980137	2.758904	1971.267808	5.575342	6.099315
std	79442.502883	1.328095	0.550916	213.804841	40.177307	30.202904	1.112799	1.382997
min	34900.000000	2006.000000	0.000000	0.000000	0.000000	1872.000000	1.000000	1.000000
25%	129975.000000	2007.000000	1.000000	334.500000	0.000000	1954.000000	5.000000	5.000000
50%	163000.000000	2008.000000	2.000000	480.000000	0.000000	1973.000000	5.000000	6.000000
75%	214000.000000	2009.000000	2.000000	576.000000	0.000000	2000.000000	6.000000	7.000000
max	755000.000000	2010.000000	3.000000	1418.000000	738.000000	2010.000000	9.000000	10.000000

Gambar 3. Statistik Data Harga Rumah pada kota Ames

Berdasarkan Gambar 3 terlihat bahwa harga rumah terendah adalah USD 34.900, sementara harga rumah tertinggi mencapai USD 755.000. Selanjutnya, hanya data numerik dengan korelasi lebih dari 0,5 dan data kategorikal yang bertipe objek yang akan digunakan yang hasilnya ditunjukkan pada Gambar 4 di bawah ini.

OverallQual (0.7910), YearBuilt (0.5229), YearRemodAdd (0.5071), TotRmsBsSF (0.6136),
1stFlrSF (0.6659), GrLivArea (0.7086), FullBath (0.5667), TotRmsAbvGrd (0.5337),
GarageCars (0.6404), GarageArea (0.6234), SalePrice (1.0000),

Kategorikal: ['MSZoning', 'Utilities', 'BldgType', 'KitchenQual', 'SaleCondition', 'Landslope']

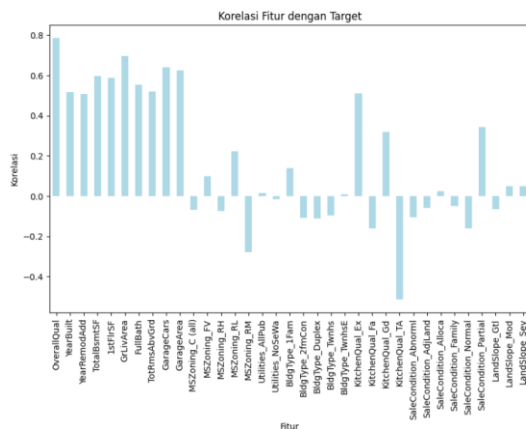
Gambar 4. Variabel yang akan Digunakan

Selanjutnya, kita mengubah nilai-nilai dalam dataset menjadi format yang lebih sesuai untuk diproses oleh model regresi linear. Variabel-variabel yang sebelumnya bertipe objek diubah menjadi bilangan integer. Berikut adalah hasil dataset yang telah ditransformasi, yang selanjutnya akan dilakukan standarisasi menggunakan Standard Scaler.

	OverallQual	YearBuilt	YearRemodAdd	TotRmsBsSF	1stFlrSF	GrLivArea	FullBath	TotRmsAbvGrd	GarageCars	GarageArea
0	0.651479	1.050994	0.878668	-0.459303	-0.793434	0.370333	0.789741	0.912210	0.311725	0.351000
1	-0.071836	0.156734	-0.429577	0.466465	0.257140	-0.482512	0.789741	-0.318683	0.311725	-0.060731
2	0.651479	0.984752	0.830215	-0.313369	-0.627826	0.515013	0.789741	-0.318683	0.311725	0.631726
3	0.651479	-1.863632	-0.720298	-0.687324	-0.521734	0.383659	-1.026041	0.296763	1.650307	0.790804
4	1.374795	0.951632	0.733308	0.199680	-0.045611	1.299326	0.789741	1.527656	1.650307	1.698485

Gambar 5. Data setelah Standarisasi

Hasil dari standarisasi yang ditunjukkan pada Gambar 5 di atas menunjukkan bahwa setelah data distandarisasi, nilai-nilai variabel tidak lagi memiliki rentang yang sangat besar. Beberapa nilai variabel menunjukkan angka negatif, yang berarti berada di bawah rata-rata, sementara nilai yang lebih besar dari 1 menunjukkan bahwa data tersebut berada lebih jauh dari rata-rata, namun tetap dalam skala yang seragam dengan deviasi standar 1. Hal ini membantu model dalam menangani variabel dengan skala yang berbeda-beda secara lebih efisien. Kemudian data training dan data testing dibagi dengan rasio 80:20 dengan rincian data sebanyak 1168 data training dan sebanyak 292 data pada data testing.



dukungan yang tak ternilai selama proses penelitian ini.

2. Universitas Amikom Yogyakarta, yang telah menyediakan fasilitas dan sumber daya yang mendukung kelancaran penelitian ini.
3. Teman sejawat, yang telah membantu dalam pengumpulan data dan analisis yang sangat berharga bagi penyelesaian penelitian ini.
4. Semua pihak yang telah memberikan kontribusi baik secara langsung maupun tidak langsung, baik berupa saran, bantuan teknis, atau motivasi selama penelitian ini berlangsung.

Penulis juga ingin menyampaikan terima kasih kepada kontributor data pada Kaggle yang telah menyediakan data yang digunakan dalam penelitian ini, yang sangat mendukung keberhasilan penelitian ini.

Semoga hasil penelitian ini dapat memberikan kontribusi yang bermanfaat bagi perkembangan ilmu pengetahuan, khususnya dalam bidang prediksi harga rumah menggunakan metode machine learning.

DAFTAR PUSTAKA

- [1] R. Khoirudin and M. L. A. Kurniawan, 'A time-varying of property residential price in Indonesia: a VAR approach', *Jurnal Ekonomi & Studi Pembangunan*, vol. 24, no. 1, pp. 69–80, May 2023, doi: 10.18196/jesp.v24i1.17750.
- [2] Suharti, Yusrizal, and L. Eprianti, 'Evaluasi Pengaruh Faktor Ekonomi Makro Terhadap Harga Saham Sektor Properti Yang Terdaftar Di BEI Periode 2017-2020', *Management Studies and Entrepreneurship Journal*, vol. 4, no. 1, pp. 834–851, 2023.
- [3] B. Wisnuadhi, I. Setiawan, and P. Korespondensi, 'REKOMENDASI FITUR YANG MEMPENGARUHI HARGA SEWA MENGGUNAKAN PENDEKATAN MACHINE LEARNING', vol. 8, no. 4, pp. 673–682, 2021, doi: 10.25126/jtiik.202183305.
- [4] I. G. A. A. Putra, 'Penentuan Nilai Pasar Properti Tanah Kosong di Kelurahan Kerobokan Kelod, Kabupaten Badung', *Jurnal Teknik Sipil*, vol. 15, no. 1, pp. 48–53, 2023.
- [5] S. H. Hasanah and E. Julianti, 'Analysis of CART and Random Forest on Statistics Student Status at Universitas Terbuka', *INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi*, vol. 6, no. 1, pp. 56–65, Feb. 2022, doi: 10.29407/intensif.v6i1.16156.
- [6] M. Jason and D. Prayogo, 'Prediksi Financial Distress Pada Perusahaan Terbuka di Sektor Konstruksi dan Properti yang Terdaftar di Bursa Efek Indonesia dengan Metode Integrasi Differential Evolution dan Least Squares Support Vector Machine', *Dimensi Utama Teknik Sipil*, vol. 10, no. 1, pp. 77–85, Apr. 2023, doi: 10.9744/duts.10.1.77-85.
- [7] I. M. Faiza, Gunawan, and W. Andriani, 'Tinjauan Pustaka Sistematis: Penerapan Metode Machine Learning untuk Deteksi Bencana Banjir', *Jurnal Minfo Polgan*, vol. 11, no. 2, pp. 59–63, Dec. 2022, doi: 10.33299/jpkop.22.2.1752.
- [8] Anjar Setiawan, Ema Utami, and Dhani Ariatmanto, 'Cattle Weight Estimation Using Linear Regression and Random Forest Regressor', *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 1, pp. 72–79, Feb. 2024, doi: 10.29207/resti.v8i1.5494.
- [9] L. Qodariah, M. Nurjihad, E. Pembangunan, F. Ekonomi, D. Bisnis, and U. T. Sumbawa, 'Pengaruh Sektor-Sektor Ekonomi Prioritas dan Variabel Demografis Terhadap Konsumsi Energi Listrik di Provinsi Nusa Tenggara Barat', *Journal of Macroeconomics and Social Development*, no. 1, pp. 1–14, 2024, [Online]. Available: <https://economics.pubmedia.id/index.php/jmsd>
- [10] A. A. Kurniawan, M. Mustikasari, and P. Korespondensi, 'EVALUASI KINERJA MLLIB APACHE SPARK PADA KLASIFIKASI BERITA PALSU DALAM BAHASA INDONESIA', *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 9, no. 3, 2022, doi: 10.25126/jtiik.202293538.
- [11] R. Martha and D. E. Herwindiati, 'Prediksi Hujan Menggunakan Metode Artificial Neural Network, K-Nearest Neighbors, dan Naïve Bayes', *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 6, no. 4, pp. 859–865, Nov. 2024, doi: 10.47233/jteksis.v6i4.1650.
- [12] Hermansyah, A. Abdullah, and P. Y. Utami, 'Penerapan Metode Regresi Linier Berganda Untuk Memprediksi Panen Kelapa Sawit', *Progresif: Jurnal Ilmiah Komputer*, vol. 20, no. 1, pp. 540–554, 2024.
- [13] E. Wibowo and I. Pratama, 'Analisis Sentimen Terhadap Ulasan Hotel Melalui Platform Google Review Menggunakan Metode Stacking', *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 6, no. 4, pp. 774–784, Oct. 2024, doi: 10.47233/jteksis.v6i4.1475.
- [14] S. Karbala, I. A. Program, S. Komputerisasi, A. D3, and F. T. Industri, 'MEMPREDIKSI HARGA BERAS ECERAN MENGGUNAKAN ALGORITMA REGRESI LINIER', *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 3, 2023.
- [15] N. Nuris, 'Analisis Prediksi Harga Rumah Pada Machine Learning Metode Regresi Linear', *EXPLORE*, vol. 14, no. 2, pp. 108–112, 2024.
- [16] M. Adha, E. Utami, and Hanafi, 'Model Hibrid Algoritma Apriori dan Regresi Linear untuk Perkiraan Produksi Jagung (Studi Kasus : Kabupaten Dompu)', *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 8, no. 3, pp. 441–450, 2022.
- [17] D. Permata Sari, R. Bayu Putra, H. Fitri, A. Ramadhanu, and F. Cahyani Putri, 'PENGARUH PEMAHAMAN PAJAK, PELAYANAN APARAT PAJAK, SANKSI PERPAJAKAN DAN PREFERENSI RISIKO PERPAJAKAN TERHADAP KEPATUHAN WAJIB PAJAK(STUDI KASUS UMKM TOKO ELEKTRONIK DI KECAMATAN SITIUNG DHARMASRAYA)', *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 1, no. 2, pp. 18–22, Jul. 2019, doi: 10.47233/jteksis.v1i2.46.
- [18] A. R. M. Togatorop, A. I. L. Bahari, and A. Choiruddin, 'Neural Networks-Based Forecasting Platform for EV Battery Commodity Price Prediction', *INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi*, vol. 7, no. 2, pp. 243–261, Aug. 2023, doi: 10.29407/intensif.v7i2.19999.
- [19] T. Nurmansyah, R. Kurniawan, Y. A. Wijaya, P. Studi, T. Informatika, and I. Cirebon, 'Analisis Data Stok Alat Kesehatan menggunakan Metode Regresi Linier Berdasarkan Nilai RMSE', vol. 6, no. 1, pp. 177–182, 2024.
- [20] M. Sholeh, Suraya, and D. Andayati, 'Machine Linear untuk Analisis Regresi Linier Biaya Asuransi

Kesehatan dengan Menggunakan Python Jupyter Notebook', *JEPIN (Jurnal Edukasi dan Penelitian Informatika)* , vol. 8, no. 1, pp. 20–27, 2022, [Online]. Available: www.data.jakarta.go.id.