



Prediksi Harga Rumah dengan Regresi Linier

Fathoni Dwi Atmoko

Program Studi Teknik Informatika, Universitas Nahdlatul Ulama, Indonesia

*Penulis Korespondensi: fathonidwiatmoko03@gmail.com

Abstract. Property price determination is a complex challenge influenced by various factors, thus requiring an effective method for accurate prediction to support investment decision-making. In the current digital era, conventional approaches are being replaced by data-driven and artificial intelligence methods, where Linear Regression remains a popular choice due to its simplicity and effectiveness in modeling linear relationships. This study aims to analyze the relationship between the physical characteristics of a house and its selling price, and to build an accurate predictive model using the Linear Regression algorithm. A quantitative method was used, focusing on Building Area, Number of Rooms, and Building Age against the House Selling Price. Correlation analysis results show that Building Area has the strongest correlation (0.81) with price, while Building Age shows a negative correlation (-0.52). The Linear Regression model demonstrated very strong and stable performance. The model achieved an R^2 Score of 0.9396 on the testing data, meaning 93.96% of house price variability can be explained by the model. Furthermore, the low MAE of only 11.31 million rupiah indicates a small prediction error, and the consistency of R^2 scores confirms that the model does not suffer from overfitting. This study concludes that the Linear Regression model provides excellent, stable, and reliable prediction performance for projecting house selling prices.

Keywords: House Price Prediction; Linear Regression; Building Area; Machine Learning.

Abstrak . Dalam memfasilitasi pengambilan keputusan investasi, pendekatan yang efisien untuk memprediksi harga secara andal diperlukan karena penetapan harga properti adalah tugas rumit yang dipengaruhi oleh beberapa faktor. Di era digital saat ini, pendekatan konvensional mulai tergantikan oleh metode berbasis data dan kecerdasan buatan, di mana Regresi Linier masih menjadi pilihan populer karena kesederhanaan dan efektivitasnya. Penelitian ini bertujuan untuk menganalisis hubungan antara karakteristik fisik rumah dan harga jual, serta membangun model prediktif yang akurat menggunakan algoritma Regresi Linier. Metode kuantitatif digunakan dengan fokus pada variabel Luas Bangunan, Jumlah Kamar, dan Usia Bangunan terhadap Harga Jual Rumah. Hasil analisis korelasi menunjukkan bahwa Luas Bangunan memiliki korelasi yang paling kuat (0.81) terhadap harga, sementara Usia Bangunan berkorelasi negatif (-0.52). Evaluasi kinerja model Regresi Linier menunjukkan performa yang sangat kuat dan stabil. Model mencapai nilai R^2 Score sebesar 0.9396 pada data pengujian, yang berarti 93,96% variabilitas harga rumah dapat dijelaskan oleh model. Selain itu, nilai MAE yang hanya 11,31 juta rupiah menunjukkan tingkat kesalahan prediksi yang kecil, dan konsistensi nilai R^2 antara data pelatihan dan pengujian mengonfirmasi bahwa model tidak mengalami overfitting. Dengan demikian, penelitian ini menyimpulkan bahwa model Regresi Linier memberikan performa prediksi yang sangat baik, stabil, dan dapat diandalkan untuk memproyeksikan harga jual rumah dalam dataset ini.

Kata Kunci: Prediksi Harga Rumah; Regresi Linier; Luas Bangunan; Machine Learning.

1. LATAR BELAKANG

Salah satu bidang ekonomi yang krusial bagi kehidupan manusia adalah properti (Khoirudin & Kurniawan, 2023). Industri ini merupakan investasi yang bijaksana mengingat populasi yang terus bertambah dan kebutuhan akan rumah yang layak huni semakin meningkat. (Yusrizal & Eprianti, 2023). Harga properti merupakan indikator ekonomi yang penting dan dipengaruhi oleh berbagai faktor. Pemahaman mengenai faktor-faktor penentu harga sangat krusial bagi pengembang, investor, dan calon pembeli. Karena bergantung pada banyak variabel, seperti luas bangunan, jumlah kamar, lokasi, dan kondisi properti itu sendiri,

penetapan harga properti menjadi masalah yang kompleks (Putra, 2023; Wisnuadhi & Setiawan, 2021). Kenaikan harga rumah niscaya akan sangat dipengaruhi oleh semua faktor ini (Saiful, 2021). Oleh karena itu, untuk membantu pembeli, penjual, dan investor dalam mengambil keputusan, diperlukan teknik prediksi harga properti yang akurat. Di era digital saat ini, teknik berbasis data dan kecerdasan buatan mulai menggantikan cara tradisional dalam menentukan harga rumah.

Pembelajaran mesin (*machine learning*) merupakan salah satu bidang kecerdasan buatan yang relevan di sini (Lindholm, Wahlström, Lindsten, & Schön, 2022). Tanpa pemrograman eksplisit, pembelajaran mesin memungkinkan komputer untuk belajar dari data masa lalu dan menghasilkan prediksi (Faiza, Andriani, & others, 2022). Karena dapat mensimulasikan hubungan linear antara variabel masukan (fitur) dan keluaran tujuan, yaitu harga rumah, teknik regresi linear terus menjadi pilihan populer dalam konteks prediksi harga rumah. Selain mudah digunakan, pendekatan ini menghasilkan temuan yang dapat dijelaskan secara matematis (Qodariah & Nurjihadi, 2024). Hasil ini akan dievaluasi menggunakan ukuran evaluasi untuk mendapatkan hasil yang lebih tepat dan ideal. (A. A. Kurniawan, Mustikasari, & Korespondensi, 2022).

Penelitian ini berfokus pada analisis kuantitatif hubungan antara karakteristik fisik rumah seperti Luas Bangunan, Jumlah Kamar, dan Usia Bangunan terhadap Harga Jual Rumah. Tujuan penelitian ini adalah untuk menganalisis hubungan antara berbagai karakteristik properti dan harga jual rumah, membangun model Regresi Linier yang mampu memprediksi harga secara akurat, serta mengevaluasi performa model tersebut melalui metrik statistik seperti MAE, MSE, dan R^2 Score. Melalui pendekatan ini, penelitian berupaya memberikan gambaran menyeluruh mengenai faktor-faktor yang paling berpengaruh terhadap nilai pasar sebuah rumah sekaligus menilai sejauh mana model prediktif dapat merepresentasikan variasi harga di lapangan.

2. KAJIAN TEORITIS

Faktor yang mempengaruhi nilai jual rumah

Harga jual rumah dipengaruhi oleh sejumlah faktor selain nilai tanah. Calon pembeli disarankan untuk menilai fitur-fitur bangunan, seperti ukuran, jumlah kamar tidur, kamar mandi, dan kondisi umum. Disarankan juga untuk memeriksa secara menyeluruh kelebihan dan kekurangan rumah, termasuk kondisi rangka dan langit-langit bangunan. Penting juga untuk menentukan apakah rumah yang Anda pertimbangkan untuk dibeli perlu diperbaiki atau rusak. Selanjutnya, bandingkan harga rumah dengan harga yang berlaku dan pertimbangkan

lokasinya, terutama jika dekat dengan sejumlah fasilitas. Harganya kemungkinan akan meningkat seiring dengan kualitas properti dan lokasinya yang strategi. (Kurniawan, 2018)

Machine Learning

Tujuan pembelajaran mesin, sebuah subbidang kecerdasan buatan, adalah menyediakan metode yang memungkinkan komputer belajar dari pengalaman dan data tanpa pemrograman eksplisit. Dengan sedikit atau tanpa keterlibatan manusia, pembelajaran mesin memungkinkan komputer mengenali pola dalam data, memperkirakan hasil, dan membuat penilaian.

Regresi Linear

Salah satu metode paling populer dalam analisis statistik untuk menentukan hubungan antara variabel dependen (y) dan satu atau lebih variabel independen (x) adalah regresi linier. (Padilah & Adam, 2019). Model regresi linear dapat dinyatakan sebagai:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon$$

Di mana:

- y adalah variabel dependen,
- x_1, x_2, \dots, x_n adalah variabel independen,
- $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ adalah koefisien regresi,
- ϵ adalah kesalahan acak.

Tujuan dari regresi linear adalah untuk menemukan koefisien β yang meminimalkan kesalahan kuadrat rata-rata antara nilai yang diamati dan nilai yang diprediksi oleh model.

Regularisasi adalah teknik yang digunakan untuk mengurangi overfitting dalam model regresi linear dengan menambahkan penalti pada besar koefisien. Dua metode regularisasi yang umum digunakan adalah regresi Ridge dan Lasso.

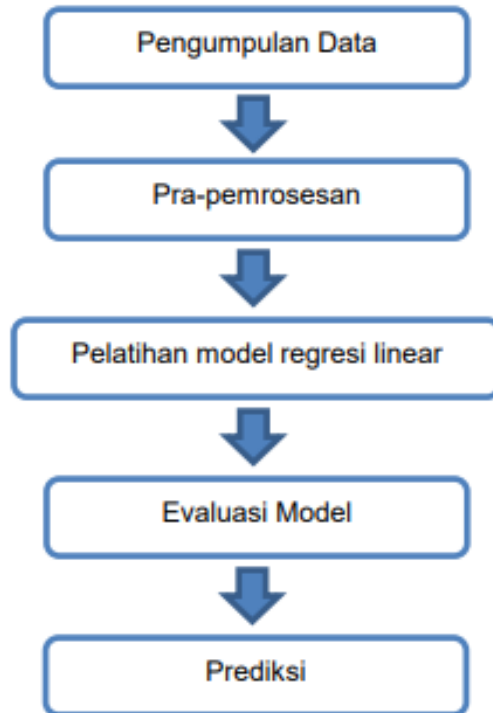
Regresi Ridge: Regresi Ridge menambahkan penalti berbentuk kuadrat dari magnitude koefisien ke fungsi objektif, yang dikenal sebagai norma L_2 . Tujuan dari penalti ini adalah untuk mencegah koefisien menjadi terlalu besar dan memperbaiki stabilitas model.

Penerapan regularisasi dalam regresi linear dapat membantu meningkatkan kinerja model dengan mengontrol kompleksitas dan mengurangi overfitting. Pemilihan antara regresi Ridge dan Lasso tergantung pada kebutuhan spesifik dari dataset dan tujuan analisis.

3. METODE PENELITIAN

Studi ini memperkirakan harga rumah menggunakan metode kuantitatif. Pengumpulan data, termasuk variabel terkait, pemrosesan dan pembersihan data, eksplorasi data, penerapan

algoritma regresi linier pada data yang telah diolah, dan interpretasi temuan merupakan beberapa langkah dalam proses penelitian. Gambar 1 mengilustrasikan hal ini.



Gambar 1. Flowchart Penelitian

Pengumpulan Data

Pada tahap ini, Dataset yang digunakan berasal dari file dataset_harga_rumah.csv yang berisi 1000 entri dan 4 variabel utama. Variabel yang dianalisis meliputi Luas Bangunan, Jumlah Kamar, Usia Bangunan, dan Harga.

Pra-pemrosesan Data:

Kemudian harus diproses untuk memastikan kebersihan dan kecocokan untuk model *Machine Learning*. Ini melibatkan penghapusan nilai yang hilang, penanganan outlier, dan konversi variabel kategori menjadi format yang bisa digunakan oleh algoritma.

Pelatihan Model Regresi Linear

Data dibagi menjadi data pelatihan dan data uji. Model dilatih menggunakan data pelatihan untuk belajar pola dan hubungan antara fitur dan harga properti.

Evaluasi Model

Model dievaluasi menggunakan metrik evaluasi yang sesuai seperti Mean Squared Error (MSE), R-squared (R^2), atau Mean Absolute Error (MAE) untuk memahami seberapa baik model dapat memprediksi harga rumah.

Prediksi

Setelah model dioptimalkan dan dievaluasi dengan baik, ia dapat digunakan untuk melakukan prediksi harga rumah untuk data baru atau properti yang belum dilihat sebelumnya.

4. HASIL DAN PEMBAHASAN

Statistik Deskriptif

Tabel 1. Statistik Deskriptif

Variabel	Count	Mean	Std. Dev	Min	Max
Luas Bangunan	1000	99.88	19.60	35.00	177.00
Jumlah Kamar	1000	4.05	1.42	1.00	6.00
Usia Bangunan	1000	25.52	14.14	2.00	49.00
Harga (Juta)	1000	300.45	60.88	90.00	511.00

Tabel 1 menyajikan ringkasan statistik deskriptif dari seluruh variabel yang digunakan dalam penelitian. Berdasarkan hasil tersebut, variabel Luas Bangunan memiliki nilai rata-rata 99,88 m² dengan standar deviasi 19,60, menunjukkan adanya variasi ukuran bangunan yang cukup lebar dalam dataset, mulai dari 35 m² hingga 177 m². Variabel Jumlah Kamar rata-rata sebesar 4,05 kamar dengan rentang antara 1 hingga 6 kamar, mengindikasikan bahwa sebagian besar rumah dalam dataset merupakan rumah berukuran menengah. Sementara itu, Usia Bangunan memiliki rata-rata 25,52 tahun dan standar deviasi 14,14, dengan usia termuda 2 tahun dan tertua 49 tahun, mencerminkan distribusi usia bangunan yang cukup beragam.

Untuk variabel Harga, nilai rata-rata tercatat sebesar 300,45 juta rupiah dengan variasi harga mulai dari 90 juta hingga 511 juta rupiah. Mayoritas rumah dalam kumpulan data berada dalam kisaran harga menengah, berdasarkan nilai rata-rata ini. Dengan mempertimbangkan semua hal, ringkasan deskriptif ini menawarkan gambaran awal tentang fitur-fitur distribusi data sebelum penyelidikan lebih lanjut.

Pra-pemrosesan

Saat mengembangkan model pembelajaran mesin, terutama untuk memprediksi harga rumah, Pra-pemrosesan data merupakan langkah awal yang penting. Pada langkah ini, data yang hilang dihilangkan, outlier dikelola, fitur-fitur distandarisasi, dan variabel kategorikal dikonversi ke format numerik, misalnya :

- a. Tidak ditemukan data duplikat (0 duplicated data) dan Tidak terdapat missing values.

```
Jumlah data duplicated: 0
0
LuasBangunan 0
JumlahKamar 0
UsiaBangunan 0
Harga 0

dtype: int64
```

Gambar 2. Bukti Data

- b. Konversi Nilai

Variabel Harga dikonversi ke satuan Juta Rupiah. Dan Variabel Luas Bangunan dibulatkan ke bawah dan dikonversi menjadi tipe integer. Agar mudah di baca

	LuasBangunan	JumlahKamar	UsiaBangunan	Harga
0	109	5	43	330
1	97	2	19	291
2	112	4	5	384
3	130	6	11	453
4	95	4	12	304

Gambar 3. Konversi Nilai

- c. Mengklasifikasikan tingkatan harga

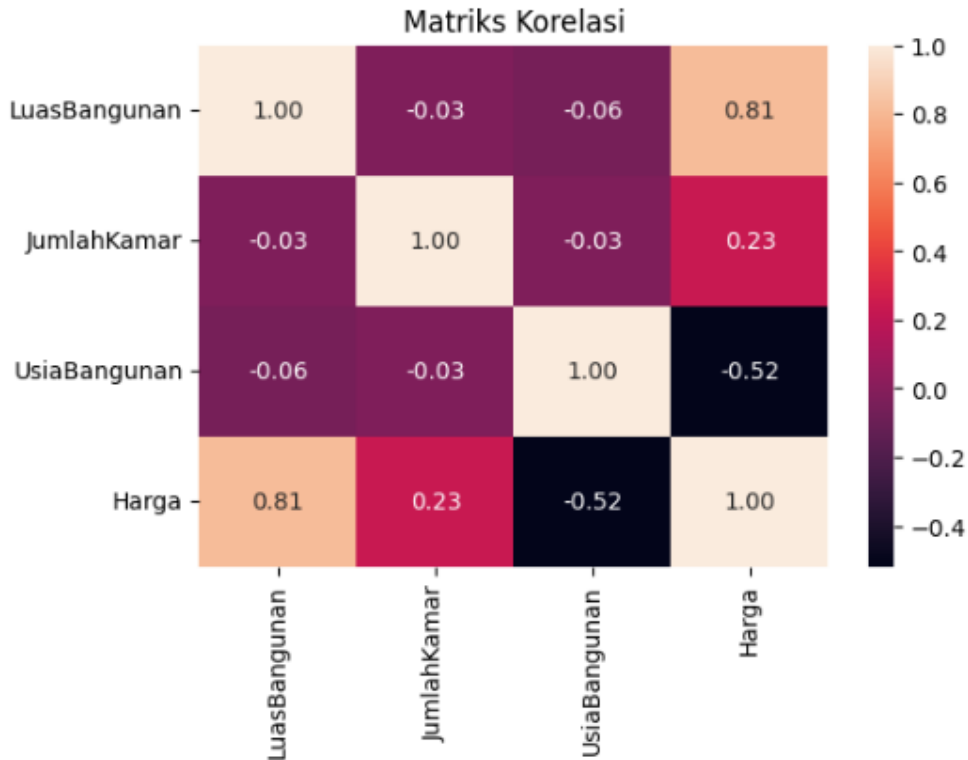
Gambar berikut menunjukkan bagaimana setiap rumah diklasifikasikan ke dalam kategori tingkat harga berdasarkan nilai harga jualnya. Variabel tingkat_harga membagi rumah menjadi beberapa kelompok, seperti Mahal, Menengah, atau kategori lainnya jika ada, untuk memudahkan analisis.

	LuasBangunan	JumlahKamar	UsiaBangunan	Harga	tingkat_harga
0	109	5	43	330	Mahal
1	97	2	19	291	Menengah
2	112	4	5	384	Mahal
3	130	6	11	453	Mahal
4	95	4	12	304	Mahal

Gambar 4. Klasifikasi tingkatan harga

Matriks Korelasi

Hasil korelasi antar variabel menunjukkan bahwa Luas Bangunan memiliki korelasi paling kuat (0.81), Jumlah Kamar berkorelasi lemah (0.23), dan Usia Bangunan berkorelasi negatif (-0.52).



Gambar 5. Matriks Korelasi

Evaluasi Model Regresi Linier

Tabel 2. Evaluasi Kinerja Model

Metrik	Data Pelatihan	Data Pengujian
MAE	12.23	11.31
MSE	233.68	215.65
R ² Score	0.9374	0.9396

Nilai $R^2 = 0.9396$ pada data pengujian menunjukkan bahwa model memiliki kemampuan prediktif yang sangat kuat dan konsisten dengan data pelatihan.

Tabel 2 menunjukkan hasil evaluasi kinerja model Regresi Linier berdasarkan tiga metrik utama, yaitu MAE, MSE, dan R² Score. Nilai MAE pada data pengujian sebesar 11,31 menandakan bahwa rata-rata kesalahan prediksi harga rumah hanya sekitar 11 juta rupiah dari nilai sebenarnya, yang merupakan tingkat kesalahan yang relatif kecil. Demikian pula, nilai MSE yang lebih rendah pada data pengujian dibandingkan data pelatihan menunjukkan bahwa model tidak hanya akurat, tetapi juga stabil dalam menghadapi variasi data baru.

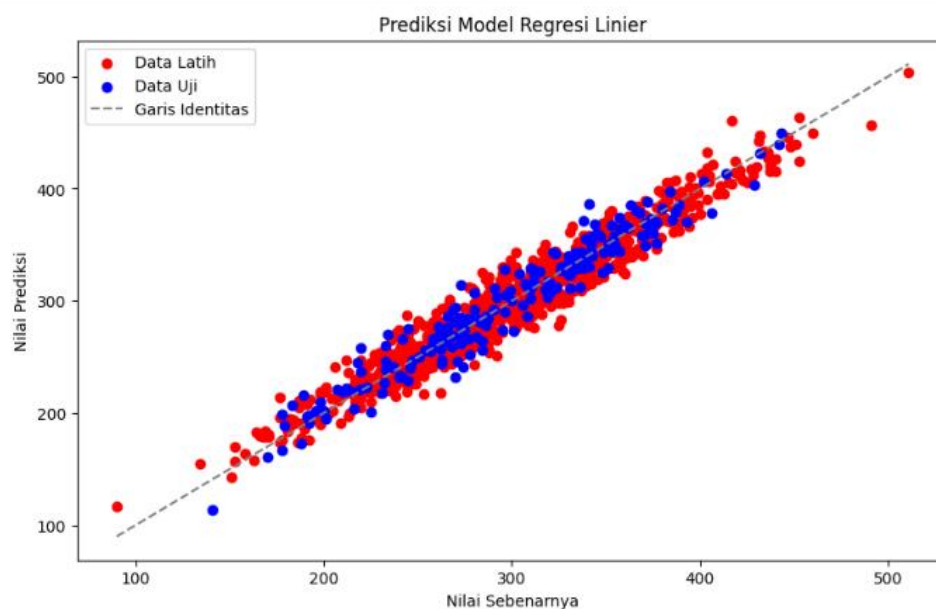
Sementara itu, nilai R² Score sebesar 0.9396 pada data pengujian memiliki makna yang sangat penting. Angka ini menunjukkan bahwa sekitar 93,96% variabilitas harga rumah dapat dijelaskan oleh model, yang berarti model memiliki kemampuan prediktif yang sangat kuat. Tidak hanya itu, konsistensi nilai R² antara data pelatihan (0.9374) dan pengujian (0.9396) juga

mengindikasikan bahwa model tidak mengalami overfitting. Dengan kata lain, model mampu mempelajari pola pada data pelatihan tanpa kehilangan kemampuan generalisasi ketika diterapkan pada data baru. Secara keseluruhan, hasil evaluasi dalam Tabel 2 menggambarkan bahwa model Regresi Linier yang dibangun memberikan performa prediksi yang sangat baik, stabil, dan dapat diandalkan untuk memproyeksikan harga jual rumah dalam dataset ini.

Prediksi

Bagian prediksi ini menunjukkan bahwa model Regresi Linier memiliki akurasi yang sangat baik dan konsisten dalam memperkirakan harga rumah. Hal ini terlihat dari titik-titik prediksi yang berkumpul dekat dengan garis identitas, menandakan bahwa nilai prediksi hampir sama dengan nilai sebenarnya. Baik data pelatihan maupun data pengujian mengikuti pola yang sama, sehingga model terbukti mampu melakukan generalisasi dengan baik dan tidak mengalami overfitting.

Distribusi titik yang rapi dan minim penyimpangan juga menggambarkan bahwa kesalahan prediksi sangat kecil, sejalan dengan hasil evaluasi seperti MAE, MSE, dan R^2 yang menunjukkan performa tinggi. Secara keseluruhan, visualisasi ini menegaskan bahwa model mampu menangkap hubungan antarvariabel secara efektif dan memberikan estimasi harga rumah yang sangat mendekati kondisi nyata.



Gambar 6. Prediksi

5. KESIMPULAN DAN SARAN

Penelitian ini menyimpulkan bahwa model Regresi Linier sangat efektif dan andal dalam memprediksi harga rumah, dengan Luas Bangunan (korelasi 0.81) menjadi faktor paling berpengaruh, dan model mencapai kemampuan prediktif yang sangat kuat dengan R² Score 0.9396 dan MAE rata-rata kesalahan hanya 11,31 juta rupiah, menunjukkan model stabil dan tidak mengalami overfitting. Untuk penelitian mendatang, disarankan untuk memperluas cakupan variabel dengan menyertakan faktor-faktor eksternal krusial seperti lokasi strategis dan kondisi makroekonomi, serta melakukan perbandingan kinerja model Regresi Linier ini dengan algoritma *Machine Learning* lain seperti *Regresi Ridge*, *Lasso*, atau *Random Forest Regressor* guna menentukan model yang paling optimal.

DAFTAR REFERENSI

- Faiza, I. M., Andriani, W., & others. (2022). Tinjauan Pustaka Sistematis: Penerapan Metode Machine Learning untuk Deteksi Bencana Banjir. *Jurnal Minfo Polgan*, 11(2), 59–63.
- Khoirudin, R., & Kurniawan, M. L. A. (2023). A time-varying of property residential price in Indonesia: a VAR approach. *Jurnal Ekonomi & Studi Pembangunan*, 24(1), 69–80.
- Kurniawan, A. A., Mustikasari, M., & Korespondensi, P. (2022). Evaluasi Kinerja MLLIB APACHE SPARK pada Klasifikasi Berita Palsu dalam Bahasa Indonesia. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 9(3).
- Kurniawan, Y. (2018). *Analisis Faktor-Faktor Yang Mempengaruhi Pembelian Rumah di Kota Madiun*. Universitas Muhammadiyah Surakarta.
- Lindholm, A., Wahlström, N., Lindsten, F., & Schön, T. B. (2022). *Machine learning: a first course for engineers and scientists*. Cambridge University Press.
- Padilah, T. N., & Adam, R. I. (2019). Analisis regresi linier berganda dalam estimasi produktivitas tanaman padi di Kabupaten Karawang. *FIBONACCI: Jurnal Pendidikan Matematika Dan Matematika*, 5(2), 117–128.
- Putra, I. G. A. A. (2023). Penentuan Nilai Pasar Properti Tanah Kosong di Kelurahan Kerobokan Kelod, Kabupaten Badung. *Portal: Jurnal Teknik Sipil*, 15(1), 48–53.
- Qodariah, L., & Nurjihadi, M. (2024). Pengaruh Sektor-Sektor Ekonomi Prioritas dan Variabel Demografis Terhadap Konsumsi Energi Listrik di Provinsi Nusa Tenggara Barat. *Journal of Macroeconomics and Social Development*, 1(3), 1–14.
- Saiful, A. (2021). Prediksi harga rumah menggunakan web scrapping dan machine learning dengan algoritma linear regression. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 8(1), 41–50.

- Wisnuadhi, B., & Setiawan, I. (2021). Rekomendasi fitur yang mempengaruhi harga sewa menggunakan pendekatan machine learning. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(4), 673–682.
- Yusrizal, S., & Eprianti, L. (2023). Evaluasi Pengaruh Faktor Ekonomi Makro Terhadap Harga Saham Sektor Properti Yang Terdaftar Di BEI Periode 2017-2020. *Management Studies and Entrepreneurship Journal*, 4(1), 834–851.