



Comparative Analysis of Machine Learning Algorithms for Predicting Commercial Property Rental Prices

Analisis Komparasi Algoritma Machine Learning untuk Prediksi Harga Sewa Properti Komersial

Dwi Juli Yantono^{1*}, Nabila Rahmawati Puspaningrum², Salsabila Amani Zein³, Adolfintje Hawila Amberam Winona Resky Marani⁴, Gregorius Januario Namat⁵

^{1,2,3,4,5}Program Studi Manajemen Aset Publik, Politeknik Keuangan Negara STAN, Indonesia

E-Mail: ¹4111220009_anton@pknstan.ac.id, ²4111220053_nabila@pknstan.ac.id, ³4111220052_sani@pknstan.ac.id, ⁴4111220090_adol@pknstan.ac.id, ⁵411122006_gerry@pknstan.ac.id

Received Nov 05th 2025; Revised Dec 16th 2025; Accepted Dec 27th 2025; Available Online Jan 16th 2026

Corresponding Author: Dwi Juli Yantono

Copyright © 2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

Determining rental prices for commercial properties (shophouses/ruko) in South Tangerang is inefficient due to subjective valuation methods and market heterogeneity. This study aims to develop an objective rental price prediction model by comparing three Machine Learning algorithms: Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR). The dataset consists of 275 listings collected via web scraping from the Lamudi platform in December 2025. The features used include building area and location (district). The evaluation results using the testing dataset show that Random Forest is the best model with an R-Squared (R^2) score of 0.6801 and a Mean Absolute Percentage Error (MAPE) of 27.40%. In contrast, XGBoost and SVR performed poorly with negative R^2 values, indicating an inability to effectively capture data patterns in small datasets. Feature importance analysis reveals that building area is the most dominant factor influencing rental prices, followed by location. This study proves that Random Forest is a robust method for property valuation in the context of public asset management.

Keywords: Commercial Property, Machine Learning, Random Forest, Rental Price, South Tangerang.

Abstrak

Penentuan harga sewa properti komersial (ruko) di Kota Tangerang Selatan seringkali menghadapi kendala inefisiensi akibat metode penilaian yang subjektif dan heterogenitas pasar. Penelitian ini bertujuan untuk mengembangkan model prediksi harga sewa yang objektif dengan membandingkan tiga algoritma *Machine Learning*: *Random Forest* (RF), *Extreme Gradient Boosting* (XGBoost), dan *Support Vector Regression* (SVR). Dataset terdiri dari 275 data listing yang dikumpulkan melalui teknik web scraping dari platform Lamudi pada Desember 2025. Fitur yang digunakan meliputi luas bangunan dan lokasi (kecamatan). Hasil evaluasi menggunakan data *testing* menunjukkan bahwa *Random Forest* adalah model terbaik dengan skor *R-Squared* (R^2) sebesar 0,6801 dan *Mean Absolute Percentage Error* (MAPE) sebesar 27,40%. Sebaliknya, XGBoost dan SVR menunjukkan performa buruk dengan nilai R^2 negatif, mengindikasikan ketidakmampuan menangkap pola data secara efektif pada dataset berskala kecil. Analisis fitur penting (*feature importance*) mengungkapkan bahwa luas bangunan menjadi faktor paling dominan yang memengaruhi harga sewa dibandingkan lokasi. Penelitian ini membuktikan bahwa *Random Forest* merupakan metode yang robust untuk valuasi properti dalam konteks manajemen aset publik.

Kata Kunci: Harga Sewa, *Machine Learning*, Properti Komersial, *Random Forest*, Tangerang Selatan.

1. PENDAHULUAN

Kota Tangerang Selatan telah bertransformasi menjadi pusat pertumbuhan ekonomi strategis, terutama setelah penetapan BSD City sebagai Kawasan Ekonomi Khusus (KEK). Data pasar terbaru kuartal IV 2024 menunjukkan lonjakan harga properti yang signifikan di kawasan ini, didorong oleh pembangunan infrastruktur masif seperti Tol Kataraja [1]. Sejalan dengan itu, Badan Pusat Statistik Kota Tangerang Selatan

mencatat adanya pertumbuhan ekonomi yang signifikan pada sektor *Real Estate* dan Perdagangan [2], menjadikan Ruko sebagai aset vital dalam ekosistem bisnis lokal.

Namun, tingginya volatilitas harga ini memunculkan masalah inefisiensi pasar, di mana pemilik ruko seringkali menentukan harga sewa secara spekulatif tanpa acuan data yang jelas. Metode penilaian konvensional (manual) yang saat ini umum digunakan dinilai tidak lagi memadai untuk merespons dinamika pasar yang cepat tersebut. Penelitian terbaru dari Lee (2024) menegaskan bahwa penilaian berbasis manusia memiliki kelemahan fatal berupa "*Surveyor Bias*" atau subjektivitas penilai yang mendistorsi harga pasar sebenarnya, yang pada akhirnya menyebabkan risiko aset mangkrak (*vacancy rate*) yang tinggi [3]. Ketidakakuratan prediksi harga ini memiliki dampak ganda; *overpricing* merugikan investor karena aset tidak segera terserap pasar, sementara *underpricing* menghilangkan potensi pendapatan yang seharusnya dapat dimaksimalkan, sebagaimana ditunjukkan dalam studi kasus negosiasi aset [4] dan fenomena penentuan harga sewa kos yang intuitif [5].

Kondisi ini menegaskan perlunya pengembangan model prediksi berbasis data yang mampu mengurangi subjektivitas dan risiko kesalahan penilaian, sehingga menjadi kebutuhan mendesak untuk menciptakan transparansi harga dan meningkatkan efisiensi pasar. Dalam dekade terakhir, penerapan teknologi *Machine Learning* (ML) dalam *Property Technology* (PropTech) telah menjadi solusi untuk mengatasi keterbatasan metode penilaian tradisional. Calainho et al. (2024) menunjukkan bahwa pendekatan ML memiliki akurasi prediksi yang lebih tinggi dibandingkan metode *Ordinary Least Squares* (OLS) karena kemampuannya mempelajari pola data yang kompleks dan non-linear [6]. Hal ini diperkuat oleh Afrisca et al. (2025) yang berhasil mengembangkan *Automated Valuation Model* (AVM) untuk aset negara menggunakan algoritma *Random Forest* [7].

Berbagai penelitian terdahulu telah mengeksplorasi penggunaan algoritma ML untuk prediksi harga properti. Mulyahati (2020) menerapkan algoritma *Random Forest* untuk memprediksi harga sewa apartemen di DKI Jakarta dengan akurasi 92,12% [8]. Di sisi lain, Munawar dan Arisal (2025) menemukan bahwa XGBoost memberikan performa terbaik di Jakarta, sementara *Support Vector Regression* (SVR) unggul di Surabaya [9]. Namun, karakteristik pembentukan harga di wilayah penyangga memiliki variabel unik seperti aksesibilitas dan infrastruktur lokal [10]. Selain itu, aspek fitur properti seperti luas bangunan dan fasilitas juga memiliki pengaruh signifikan yang berbeda-beda di setiap wilayah [11].

Meskipun penelitian mengenai prediksi harga properti sudah cukup banyak, studi yang secara spesifik membandingkan kinerja algoritma ML untuk properti komersial (ruko) di wilayah suburban seperti Tangerang Selatan dengan karakteristik data terbatas masih minim. Penelitian ini bertujuan mengisi celah tersebut dengan melakukan analisis komparasi terhadap tiga algoritma: *Random Forest*, XGBoost, dan SVR. Kontribusi utama penelitian ini adalah menemukan model paling robust untuk karakteristik pasar properti Tangerang Selatan.

2. TINJAUAN LITERATUR

Penerapan *Machine Learning* dalam penilaian real estate telah berkembang pesat sebagai alternatif metode hedonik tradisional. Berbagai penelitian telah melakukan studi komparasi untuk menemukan algoritma terbaik dalam memprediksi harga properti. Rawool et al. [12] dan Gupta et al. [13] dalam penelitiannya menemukan bahwa algoritma RF secara konsisten memberikan akurasi tertinggi dibandingkan algoritma pohon keputusan tunggal (*Decision Tree*) maupun regresi linear dalam memprediksi harga jual rumah, karena kemampuannya menangani non-linearitas data. Temuan ini didukung oleh He [14] yang menganalisis harga properti di Amerika Serikat, di mana *Random Forest* terbukti menjadi model optimal dengan *Mean Squared Error* (MSE) terendah dibandingkan Lasso dan regresi linear.

Di sisi lain, algoritma berbasis boosting seperti XGBoost sering kali dilaporkan mengungguli *Random Forest* dalam *dataset* berskala besar. Li [15] membandingkan empat model (Linear, RF, ANN, XGBoost) pada *dataset* King County dan menemukan bahwa XGBoost mencapai akurasi tertinggi (88,8%), sedikit lebih unggul dibanding *Random Forest* (87,8%). Hal serupa ditemukan oleh Mahalakshmi et al. [16] dan Kee & Ho [17], di mana XGBoost menunjukkan kinerja superior dengan nilai R^2 mencapai 0,908 dibandingkan RF (0,876) dalam konteks pasar perumahan Hong Kong yang kompleks. Keunggulan XGBoost terletak pada kemampuannya menangani bias-variance trade-off secara lebih efektif pada *dataset* yang besar dan memiliki banyak fitur.

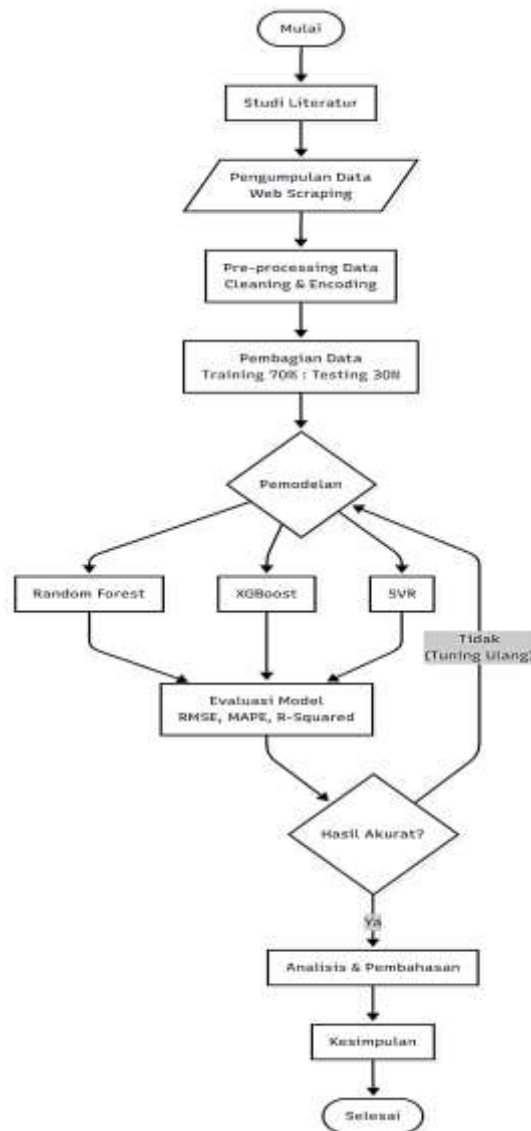
Selain metode berbasis pohon (*tree-based*), SVR juga banyak diteliti. Maulana et al. [18] dalam studi optimalisasinya menemukan bahwa meskipun CatBoost menjadi model terbaik, SVR dengan kernel polinomial mampu mendekati performa algoritma boosting jika dilakukan penyetelan parameter yang tepat. Hernes et al. [19] juga menyoroti potensi ML dalam memprediksi pasar primer properti yang memiliki dinamika berbeda dengan pasar sekunder.

Meskipun berbagai studi komparatif telah mengeksplorasi kinerja *Random Forest*, XGBoost, dan SVR dalam prediksi harga properti, sebagian besar berfokus pada *dataset* besar di wilayah urban seperti Jakarta, Yogyakarta, dan pasar internasional dengan ribuan sampel. Berbeda dengan tren tersebut di mana XGBoost cenderung mendominasi pada data berskala besar, penelitian ini secara unik menguji efektivitas

algoritma bagging (*Random Forest*) dibandingkan *boosting* (XGBoost) dan SVR pada *dataset* terbatas (275 data listing) properti komersial ruko di wilayah suburban Tangerang Selatan. Penelitian ini mengisi gap krusial dengan membuktikan algoritma mana yang paling robust untuk menangani risiko *overfitting*, heterogenitas data kecil, dan karakteristik pasar penyangga yang belum banyak diteliti, sehingga memberikan kontribusi praktis bagi manajemen aset publik di Indonesia.

3. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksperimental. Alur penelitian dirancang secara sistematis mulai dari pengumpulan data hingga evaluasi model, sebagaimana digambarkan pada Gambar 1.



Gambar 1. Alur Metodologi Penelitian

3.1 The Length of the Paper

Data yang digunakan merupakan data primer yang diperoleh melalui teknik web scraping dari situs properti Lamudi.co.id pada bulan Desember 2025 dengan fokus wilayah Kota Tangerang Selatan. Total data mentah yang dikumpulkan sebanyak 275 data listing. Variabel yang diambil meliputi: Harga Sewa (Target), Luas Bangunan (Fitur), dan Lokasi Kecamatan (Fitur). Tabel 1 adalah sampel data mentah dari penelitian.

Tabel 1. Sampel Data Mentah (Raw Data Listing)

No	Lokasi (Raw)	Harga (Raw)	Luas Bangunan (Raw)
1	Serpong, Tangerang Selatan	Rp200.000.000	145 m ²
2	Bumi Serpong Damai, Tangerang Selatan	Rp120.000.000	120 m ²

No	Lokasi (Raw)	Harga (Raw)	Luas Bangunan (Raw)
3	Bumi Serpong Damai, Tangerang Selatan	Rp150.000.000	190 m ²
4	Serpong, Tangerang Selatan	Rp85.000.000	107 m ²
5	Serpong, Tangerang Selatan	Rp15.000.000	7 m ²

3.2 Pra-Pemrosesan Data (*Preprocessing*)

Data mentah melalui tahapan pembersihan (*cleaning*) menggunakan bahasa pemrograman Python di lingkungan Google Colab:

1. Pembersihan Teks: Menghapus karakter non-numerik seperti "Rp", "Juta", "m²", dan "/tahun" agar dapat dikonversi menjadi tipe data integer/float.
2. Penanganan *Outlier*: Menghapus data dengan harga sewa atau luas bangunan yang bernilai 0 atau tidak masuk akal.
3. *Feature Encoding*: Mengubah data kategorikal "Kecamatan" (seperti Serpong, Ciputat, Pamulang) menjadi format numerik menggunakan teknik *Label Encoding* agar dapat diproses oleh algoritma.
4. Pembagian Data: *Dataset* dibagi menjadi data latih (training set) sebesar 70% dan data uji (*testing set*) sebesar 30% menggunakan fungsi `train_test_split` dari pustaka Scikit-learn.

Setelah melalui tahap pembersihan data didapatkan data bersih seperti sampel pada Tabel 2.

Tabel 2. Sampel Data Bersih

No	Kecamatan (Label)	Harga Bersih (Float)	Laba Bersih (Float)
1	Serpong	Rp200.000.000	145.02
2	Serpong	Rp120.000.000	120.0
3	Serpong	Rp155.000.000	190.0
4	Serpong	Rp85.000.000	107.0
5	Serpong	Rp15.000.000	7.0

3.3 Algoritma Pemodelan

Pemilihan algoritma didasarkan pada karakteristik data dan tinjauan literatur.

3.3.1 *Random Forest*

RF adalah metode ensemble berbasis bagging yang diperkenalkan oleh Breiman [20], dimana banyak pohon keputusan dibangun secara independen dan hasil prediksinya digabungkan untuk menghasilkan prediksi akhir. Algoritma ini dipilih karena stabilitasnya terhadap *noise*, kemampuan menangani hubungan non-linear, serta keunggulannya dalam mengurangi varians dan *overfitting*, menjadikannya dominan dalam studi penilaian massal sebagaimana dipetakan oleh Zilli et al. [21]. Prediksi regresi RF untuk input x adalah rata-rata dari K pohon, sebagaimana dijelaskan oleh Ho et al. [22]:

$$\hat{y} = \frac{1}{K} \sum_{k=1}^K h_k(x) \quad (1)$$

Dimana $\hat{y}$ adalah prediksi harga, dan $h_k(x)$ adalah *output* pohon ke- k

3.3.2 *Extreme Gradient Boosting (Xgboost)*

XGBoost, yang dikembangkan oleh Chen dan Guestrin [23], dipilih untuk mewakili metode *Boosting*. Algoritma ini membangun pohon secara sekuensial untuk memperbaiki kesalahan dari pohon sebelumnya. Keunggulan utamanya adalah kemampuan meminimalkan *spatial autocorrelation* dalam penilaian massal [24]. Fungsi objektifnya adalah:

$$Obj(\theta) = L(\theta) + \Omega(\theta) \quad (2)$$

3.3.3 *Support Vector Regression (SVR)*

SVR adalah adaptasi dari SVM untuk regresi yang diperkenalkan oleh Drucker et al. [25]. Algoritma ini dipilih karena terbukti efektif untuk *dataset* berukuran kecil hingga menengah [26]. Karena kemampuannya menangani pola data yang kompleks dan non-linear dengan margin toleransi yang ditentukan. Algoritma ini sering digunakan dalam penilaian properti dengan jumlah data terbatas. Persamaan optimasi SVR bertujuan meminimalkan error dengan batasan margin toleransi ϵ :

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (3)$$

3.4 Evaluasi Model

Evaluasi dilakukan menggunakan tiga metrik: *Root Mean Squared Error* (RMSE) untuk mengukur rata-rata kesalahan prediksi kuadrat, semakin kecil nilainya semakin baik hasil dari akurasi modelnya. *Mean Absolute Percentage Error* (MAPE) untuk menampilkan persentase kesalahan prediksi rata-rata, bisa digunakan untuk membandingkan performa model pada skala yang berbeda. Dan *R-Squared* (R^2) untuk menunjukkan seberapa besar variasi data yang dapat dijelaskan oleh model, semakin mendekati 1 maka semakin baik hasilnya

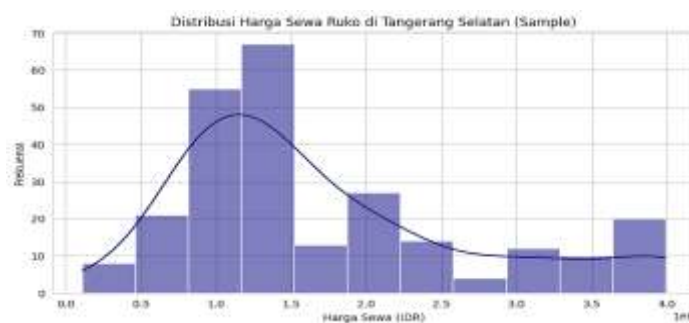
4. HASIH DAN PEMBAHASAN

4.1. Pelaksanaan Eksperimen dan Konfigurasi Model

Tahap Hyperparameter Tuning dilakukan menggunakan *Grid Search Cross-Validation*. Konfigurasi terbaik yang diperoleh adalah: RF ($n_estimators=100$, $max_depth=None$), XGBoost ($learning_rate=0.1$, $max_depth=5$), dan SVR (Kernel RBF, $C = 1.0$)

4.2. Analisis dan Deskriptif

Analisis terhadap data menunjukkan distribusi harga yang right-skewed, dengan mayoritas ruko di kisaran Rp 50-150 juta, namun terdapat *outlier* properti premium di atas Rp 300 juta (lihat Gambar 2).



Gambar 2. Distribusi Harga Sewa Ruko di Tangerang Selatan

4.3. Perbandingan Performa Model

Hasil evaluasi pada data uji (*testing set*) disajikan dalam Tabel 3.

Tabel 3. Perbandingan Akurasi Model Prediksi

Algoritma	RMSE (Rp)	MAPE (%)	R^2 Score
Random Forest	73.374.292	27,40%	0,6801
XGBoost	137.545.200	52,27%	-0,1241
SVR	155.829.140	67,05%	-0,4428

Berdasarkan Tabel 1, *Random Forest* terbukti menjadi model paling superior dengan nilai R^2 sebesar 0,6801. Hasil ini konsisten dengan temuan Hadi dan Benedict [27] yang menyatakan bahwa *Random Forest* memiliki stabilitas tinggi dalam penilaian aset properti karena mekanisme *bagging* yang mampu mengurangi varians tanpa meningkatkan bias secara signifikan, terutama pada *dataset* dengan *noise* yang tinggi.

Sebaliknya, XGBoost dan SVR menunjukkan performa yang buruk. Kegagalan ini mengindikasikan bahwa metode boosting dan kernel SVR membutuhkan jumlah data yang lebih besar untuk generalisasi pola yang efektif dibandingkan metode *bagging* pada kasus ini.

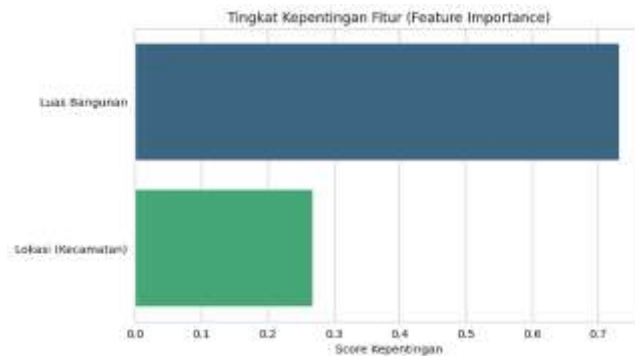
4.4. Analisis Fitur Penting

Analisis *feature importance* menggunakan model *Random Forest* menunjukkan bahwa variabel Luas Bangunan mendominasi pengaruh terhadap harga sewa dibandingkan Lokasi (Kecamatan). Tingkat kepentingan fitur dapat dilihat pada Gambar 3.

Hasil penelitian ini menunjukkan bahwa *Random Forest* lebih unggul dibandingkan XGBoost dan SVR dalam memprediksi harga sewa ruko di Tangerang Selatan. Hal ini sejalan dengan temuan Hadi dan Benedict yang menyatakan bahwa *Random Forest* memiliki stabilitas tinggi dalam penilaian aset properti karena mekanisme *bagging* yang bisa mengurangi varians tanpa meningkatkan bias secara signifikan terutama pada *dataset* dengan *noise* yang tinggi.

Sebaliknya, XGBoost dan SVR menunjukkan performa yang kurang baik. Kegagalan ini menandakan bahwa metode boosting dan kernel SVR membutuhkan jumlah data yang lebih besar untuk generalisasi pola yang efektif dibandingkan metode *bagging* pada kasus ini. Hasil ini berbeda dengan penelitian di Jakarta oleh

Munawar dan Arisal yang menunjukkan XGBoost lebih unggul, namun konteks data dan ukuran *dataset* berbeda.



Gambar 3. Tingkat Kepentingan Fitur

5. KESIMPULAN

Penelitian ini berhasil membandingkan kinerja algoritma *Machine Learning* untuk prediksi harga sewa ruko di Tangerang Selatan. Algoritma *Random Forest* terbukti menjadi metode paling akurat dengan nilai R^2 0,6801 dan MAPE 27,40%, mengungguli XGBoost dan SVR. Penelitian ini membuktikan bahwa *Random Forest* lebih robust untuk *dataset* terbatas. Penelitian ini masih menggunakan data terbatas yaitu hanya 275 listing, sehingga hasil generalisasi model masih terbatas dan belum mencerminkan seluruh dinamika pasar properti di wilayah lain yang lebih luas cakupannya. Selain itu masih belum mempertimbangkan faktor lain seperti kondisi bangunan dan aksesibilitas. Saran untuk penelitian selanjutnya adalah mengumpulkan data lebih banyak dari sumber yang lebih luas untuk meningkatkan generalisasi model. Selain itu menambahkan faktor lain seperti aksesibilitas, kondisi bangunan, dan fasilitas pendukung lainnya.

REFERENSI

- [1] Kosadata, "Pasar Properti di Kabupaten Tangerang dan Tangsel Tumbuh Signifikan," Kosadata.com, Feb. 12, 2025. [Online]. Available: [Link Berita]..
- [2] Badan Pusat Statistik Kota Tangerang Selatan, Kota Tangerang Selatan dalam angka 2025. Tangerang Selatan: BPS Kota Tangerang Selatan, 2025.
- [3] C. Lee, "Alleviating surveyor bias in real estate: An application to vacancy and property prices," *International Journal of Strategic Property Management*, vol. 28, no. 2, pp. 93–100, 2024.
- [4] D. Noviana, E. S. Ferista, B. T. Handayani, F. P. Adelia, and A. Setiawan, "Analisis strategi gaya negosiasi tarif sewa tanah PT KAI dengan warga di Kelurahan Sumbang, Kabupaten Bojonegoro," *Kalijaga: Jurnal Penelitian Multidisiplin Mahasiswa*, vol. 1, no. 2, pp. 33-40, 2024.
- [5] D. Prasdika, R. Auliyah, and A. R. Setiawan, "Menguak nilai dan makna di balik praktik penentuan harga sewa: Studi fenomenologis pada pengusaha kos-kosan," *Jurnal InFestasi*, vol. 14, no. 1, pp. 40-56, 2018.
- [6] F. D. Calainho, A. M. van de Minne, and M. K. Francke, "A machine learning approach to price indices: Applications in commercial real estate," *The Journal of Real Estate Finance and Economics*, vol. 68, pp. 624–653, 2024.
- [7] C. C. Afrisca, H. N. Rofiq, and D. D. Atmoko, "Penilaian properti: Penggunaan machine learning untuk prediksi nilai sewa," *Jurnal Manajemen Keuangan Publik*, vol. 8, no. 2, pp. 138-155, 2025..
- [8] I. L. Mulyahati, "Implementasi machine learning prediksi harga sewa apartemen menggunakan algoritma random forest melalui framework website Flask Python," Skripsi, Universitas Islam Indonesia, 2020.
- [9] Y. F. Munawar and A. Arisal, "Analisis prediksi harga sewa ruko menggunakan pendekatan machine learning," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 3, pp. 2538-2544, 2025.
- [10] A. Siddiqui, G. Thekkayil, R. Verma, V. Verma, and R. Kaushik, "Property price prediction using machine learning," *Journal of Information Systems Engineering and Management*, vol. 10, no. 41s, 2025.
- [11] B. Wisnuadhi and I. Setiawan, "Rekomendasi fitur yang mempengaruhi harga sewa menggunakan pendekatan machine learning," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 8, no. 4, pp. 673-682, 2021.
- [12] A. G. Rawool, D. V. Rogye, S. G. Rane, and V. A. Bharadi, "House Price Prediction Using Machine Learning," *Iconic Research and Engineering Journals*, vol. 4, no. 11, pp. 29-33, 2021.
- [13] V. Gupta, A. Yadav, V. Nishad, and A. Haider, "House Price Prediction Using Machine Learning,"

- International Journal of Innovative Research in Computer Science & Technology, vol. 12, no. 2, pp. 175-180, 2024.
- 14 T. He, "Analysis and Prediction of Influencing Factors of American Real Estate Price Based on Machine Learning," in Proceedings of the 2nd International Conference on Financial Technology and Business Analysis, 2024, pp. 315-323.
- [15] C. Li, "House price prediction using machine learning," in Proceedings of the 4th International Conference on Signal Processing and Machine Learning, 2024, pp. 225-237.
- [16] K. Mahalakshmi, D. J. Priyan J, D. Raj N, and A. Aravind, "Enhancing House Price Predictability: A Comprehensive Analysis of Machine Learning Techniques for Real Estate and Policy Decision-Making," Data Analytics and Artificial Intelligence, vol. 4, no. 2, pp. 68-75, 2024.
- [17] T. Kee and W. K. O. Ho, "eXplainable Machine Learning for Real Estate: XGBoost and Shapley Values in Price Prediction," Civil Engineering Journal, vol. 11, no. 05, pp. 2116-2133, 2025.
- [18] I. Maulana, A. M. Siregar, S. A. P. Lestari, and S. Faisal, "Optimal Study of Real-Estate Price Prediction Models Using Machine Learning," Jurnal Teknik Informatika (JUTIF), vol. 5, no. 4, pp. 1149-1164, 2024..
- [19] M. Hernes, P. Tutak, and M. Siewiera, "Prediction of residential real estate price on primary market using machine learning," Procedia Computer Science, vol. 246, pp. 3142–3147, 2024..
- [20] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
- [21] C. A. Zilli, L. C. Bastos, and L. R. da Silva, "Machine learning models in mass appraisal for property tax purposes: a systematic mapping study," Aestimum, vol. 84, pp. 31-52, 2024..
- [22] W. K. O. Ho, B. S. Tang, and S. W. Wong, "Predicting property prices with machine learning algorithms," Journal of Property Research, vol. 38, no. 1, pp. 48–70, 2020.
- [23] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2016, pp. 785–794..
- [24] A. A. F. Oliveira, et al., "Comparing traditional and machine learning techniques in apartments mass appraisal in Fortaleza, Brazil," Aestimum, vol. 85, pp. 21-38, 2024.
- [25] H. Drucker, C. J. C. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, "Support vector regression machines," in Advances in Neural Information Processing Systems 9, MIT Press, 1997, pp. 155–161.
- [26] A. U. Akar and S. Yalpir, "Using SVR and MRA methods for real estate valuation in the smart cities," The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLVI-4/W5-2021, pp. 21–26, 2021..
- [27] N. Hadi and J. Benedict, "Implementasi machine learning untuk prediksi harga rumah menggunakan algoritma random forest," Computatio: Journal of Computer Science and Information Systems, vol. 8, no. 1, pp. 50-61, 2024.