

PREDIKSI HARGA PERUMAHAN MENGGUNAKAN METODE *PRINCIPAL COMPONENT ANALYSIS* DAN *RANDOM FOREST* REGRESI

Yusuf Sutanto¹, Budi Al Amin², Heribertus Ary Setyadi³, Bambang Eka Purnama⁴

¹Universitas Dharma Adi Unggul Bhirawa, Surakarta, ^{2,3}Universitas Bina Sarana Informatika, Jakarta Pusat

⁴Politeknik Kesehatan Bhakti Mulia, Sukoharjo

Email: ¹yusuf.sutanto@stie-aub.ac.id, ²budi.bdm@bsi.ac.id, ³heribertus.hbs@bsi.ac.id,
⁴bambang@poltekkesbhaktimulia.ac.id

*Penulis Korespondensi

(Naskah masuk: 04 September 2024, diterima untuk diterbitkan: 08 Desember 2025)

Abstrak

Penentuan harga merupakan salah satu aspek krusial dalam kegiatan pengembangan properti mengingat hal tersebut akan mempengaruhi margin keuntungan yang diperoleh pengembang dan pilihan pembelian properti. Selama bertahun-tahun, prediksi harga rumah telah menjadi topik penelitian utama, karena permintaan rumah terus meroket. Sangat penting untuk mengembangkan kerangka kerja yang sesuai yang memungkinkan pembeli dan penjual untuk membuat keputusan cepat dalam hal membeli atau menjual properti. Dalam penelitian ini menggunakan metode *Principal Component Analysis* (PCA) dan *Random Forest* (RF), dengan tujuan untuk melakukan analisis akurasi penggunaan kedua metode dalam prediksi harga rumah dan untuk mengetahui pengaruh penggunaan PCA dalam mengoptimalkan metode *random forest*. Data yang digunakan adalah harga rumah di kota Surakarta berdasarkan hasil *scraping* data di situs propertygurugroup.com. Hasil analisis menunjukkan bahwa jumlah penjualan rumah tertinggi adalah daerah Plesungan, dan penjualan rumah yang memiliki sertifikat hak milik juga paling tinggi. Dari sepuluh variabel yang ada, luas tanah dan bangunan paling berpengaruh terhadap harga jual. Hasil pelatihan model menunjukkan bahwa peggabungan metode RF dan PCA memiliki nilai yang lebih optimal dibanding hanya menggunakan metode RF saja. Tingkat kesalahan dalam metode PCA lebih kecil, dengan rerata 0,0257 maka nilainya lebih konsisten dibanding hanya menggunakan metode RF yang nilai kesalahannya lebih besar dengan rerata 0,0332. Waktu pelatihan model menggunakan PCA lebih cepat (5005,75) dibanding hanya menggunakan metode RF (6099,25).

Kata kunci: Prediksi, Harga Perumahan, PCA, *Random Forest*

HOUSING PRICE PREDICTION USING PRINCIPAL COMPONENT ANALYSIS AND RANDOM FOREST REGRESSION METHODS

Abstract

Determining prices is one of the crucial aspects in property development activities considering that this will affect the profit margin obtained by developers and property purchase choices. Over the years, home price prediction has been a major research topic, as demand for homes continues to increase. It is important to develop a suitable framework that allows buyers and sellers to make quick decisions when it comes to buying or selling a property. This research uses the Principal Component Analysis (PCA) and Random Forest (RF) methods, with the aim of accuracy analyzing using both methods in predicting housing prices and to determine the effect of using PCA in optimizing the random forest method. The data used is house prices in Surakarta city based on data scraping results on propertygurugroup.com site. The analysis results show that the highest house sales is in Plesungan area, and houses sale with ownership certificates is also the highest. Of the ten variables, land area and building have the most influence on selling price. Model training results show that combination of RF and PCA methods has a more optimal value than using only RF method. Error rate in the PCA method is smaller, with an average of 0.0257, so the value is more consistent than just using the RF method, which has a larger error value with an average of 0.0332. The model training time using PCA is faster (5005.75) than just using the RF method (6099.25).

Keywords: *Prediction, Housing Price, PCA, Random Forest*

1. PENDAHULUAN

Properti merupakan produk unik yang tidak dapat dikontraskan dengan produk komersial lainnya

karena adanya dua kondisi harga. Penentuan harga properti merupakan salah satu aspek krusial dalam kegiatan pengembangan properti mengingat hal tersebut akan mempengaruhi margin keuntungan

yang diperoleh pengembang dan pilihan pembelian properti (Analisa and Okada, 2023). Pertama, kondisi yang disediakan produk merupakan kondisi alamiah yang melekat pada produk, yang meliputi hal-hal seperti lokasi properti, bentuk tanah, keadaan fisik, dan sumber daya alam. Kedua, kondisi yang diciptakan oleh pengembang properti untuk meningkatkan nilai produk, seperti pengembangan fasilitas, fitur keamanan, akses tanpa gerbang, dan desain (Rahadi et al., 2020). Sebagai strategi penetapan harga, pengembang properti sering kali memasukkan biaya pengembangan properti yang relevan dan margin keuntungan yang telah ditetapkan ke dalam harga jual (Lu, Shi and Pettit, 2023).

Setiap orang ingin membeli dan tinggal di rumah dengan fitur yang sesuai dengan gaya hidup mereka dan menawarkan fasilitas yang memenuhi kebutuhan mereka, tetapi memprediksi harga rumah sangat sulit, karena terus berubah. Oleh karena itu, menetapkan harga untuk pembelian dan penjualan rumah adalah proses yang harus dianalisis secara mendalam (Perdamaian and Zhai, 2024). Harga di mana rumah dipasarkan secara langsung memengaruhi profitabilitas agen real estat. Dalam hal itu, harga rumah memainkan peran penting dalam perekonomian dan sangat penting bagi berbagai pemangku kepentingan yang berinteraksi, termasuk pemilik rumah, pembeli, bank, pengembang real estat, agen real estat, dan lainnya (Igamo et al., 2023). Selain itu, karena harga merupakan faktor penting dalam penjualan rumah, disarankan untuk menetapkan harga beli dan jual yang tepat sejak awal untuk mendapatkan hasil yang diinginkan (Cahyani Putri and Arianto, 2024).

Harga rumah mengalami kenaikan setiap tahunnya, dikarenakan beberapa faktor yang mempengaruhi seperti lokasi, luas wilayah, fasilitas, dan lain sebagainya. Oleh karena itu, prediksi harga rumah merupakan salah satu hal yang penting dalam industri properti dan bermanfaat bagi para pembeli dalam mengambil keputusan (Utama, 2022). Oleh karena itu, diperlukan suatu sistem atau mekanisme yang dapat memprediksi harga rumah di masa yang akan datang, salah satu mekanismenya adalah ML untuk meningkatkan dan memprediksi harga rumah dengan tingkat akurasi yang tinggi (Dhaka et al., 2024). Selama bertahun-tahun, prediksi harga rumah telah menjadi topik penelitian utama, karena permintaan rumah terus meroket. Sangat penting untuk mengembangkan kerangka kerja yang sesuai yang memungkinkan pembeli dan penjual untuk membuat keputusan cepat dalam hal membeli atau menjual properti (Srivastava et al., 2021).

Machine learning (pembelajaran mesin) adalah suatu metode analisis data yang melakukan otomatisasi konstruksi model analitis. Pembelajaran mesin adalah cabang dari kecerdasan buatan dan berdasarkan pada suatu gagasan bahwa suatu sistem dapat melakukan pembelajaran dari data, identifikasi pola sendiri, dan membuat suatu keputusan melalui

sedikit campur tangan manusia (Pramukantoro et al., 2024). Bersama dengan teknologi big data dan komputasi berperforma tinggi, pembelajaran mesin telah berkembang untuk membangun peluang baru dalam berbagai pengaturan operasi untuk menjelaskan, mengukur, dan mendapatkan proses yang intensif data (Pitaloka, Hartanto and Sandiwarno, 2024). *Random forest* adalah model pembelajaran mesin yang digunakan dalam klasifikasi dan peramalan. Untuk melatih algoritma pembelajaran mesin dan model kecerdasan buatan, sangat penting untuk memiliki sejumlah besar data berkualitas tinggi untuk pengumpulan data yang efektif (Prasetyo et al., 2024). Data kinerja sistem sangat penting untuk menyempurnakan algoritma, meningkatkan efisiensi perangkat lunak dan perangkat keras, mengevaluasi perilaku pengguna, memungkinkan identifikasi pola, pengambilan keputusan, pemodelan prediktif, dan pemecahan masalah, yang pada akhirnya menghasilkan peningkatan efektivitas dan akurasi. *Principal Component Analysis* (PCA) adalah metode untuk mengurangi dimensionalitas kumpulan data tertentu. Meningkatkan interpretabilitas tetapi tanpa kehilangan banyak informasi (Winanto et al., 2024). Hal ini dicapai dengan membuat kovariat baru yang tidak saling terkait. Menemukan variabel baru tersebut, atau yang kita sebut komponen utama, akan mengurangi masalah solusi nilai eigen atau vektor eigen. PCA dapat dikatakan sebagai teknologi analisis data adaptif karena variabel teknologi dikembangkan untuk beradaptasi dengan berbagai jenis dan struktur data (Parhusip et al., 2022).

Hasil penelitian terdahulu yang juga menggunakan metode *random forest* membuat analisa untuk peridksi harga rumah di kota Surabaya. Tujuan penelitian ini adalah untuk menentukan kewajaran harga properti, harganya di bawah atau di atas pasaran atau sesuai harga pasaran. Beberapa uji coba dilakukan untuk mencapai nilai prediksi yang tinggi; nilai prediksi tertinggi dicapai dengan menggunakan 80% data untuk pelatihan dan 20% data untuk pengujian. Pengklasifikasi menggunakan metode *random forest* menghasilkan akurasi dan skor F1 terbesar, 88%, diantara semua metode klasifikasi lainnya. Dengan akurasi tersebut, diharapkan pemilik rumah dapat mengetahui harga yang telah ditentukan sebelumnya terlalu rendah, rata-rata, atau tinggi (Tanamal et al., 2023).

Penelitian sejenis lainnya mengembangkan sistem berbasis web untuk memprediksi harga jual juga harga beli dari dua jenis rumah yaitu rumah perkotaan dan rumah pedesaan. Pengembangan sistem menggunakan atau mengikuti tahapan konstruksi dalam pembelajaran mesin, dan metodologi *Rational Unified Process*. Sistem yang dikembangkan dapat menghasilkan prediksi harga dari lokasi dan beberapa fasilitas rumah yang diinginkan calon konsumen. Uji akurasi sistem oleh 10 pakar bidang pembelajaran mesin dan pengusa

properti menggunakan regresi linier menghasilkan nilai 4,88 dari 5, sehingga akurasinya sangat tinggi (Jáuregui-Velarde et al., 2023).

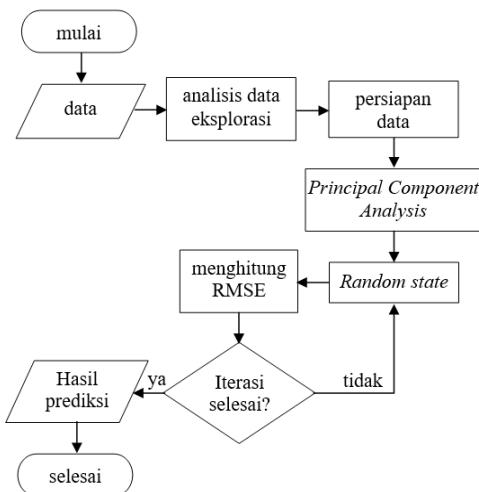
Hasil penelitian yang diterbitkan dalam bentuk jurnal melakukan prediksi untuk penjualan kendaraan listrik. Algoritma *backpropagation* digunakan untuk memprediksi harga penjualan dari tahun 2015 sampai 2022. Pengujian akurasi menggunakan *metode MAPE* menghasilkan nilai 6,25% dan termasuk dalam kategori prediksi yang akurat (Ramadhanti, Hairani and Innuddin, 2023).

Dari tiga penelitian terdahulu, terdapat beberapa perbedaan, antara lain metode yang digunakan dan cara pengujian akurasi. Terdapat satu penelitian terdahulu hanya menggunakan metode *random forest* saja, tidak digabungkan dengan PCA. Dalam menentukan prediksi harga rumah yang dilakukan dalam penelitian ini menggunakan metode PCA dan RF. Dari kedua metode yang digunakan dilakukan analisis untuk akurasi dalam menghasilkan prediksi harga rumah dan untuk mengetahui pengaruh penggunaan PCA dalam mengoptimalkan metode RF. Data yang dikumpulkan dan digunakan sebagai bahan penelitian adalah harga rumah di kota Surakarta berdasarkan hasil *scraping* data dari situs propertygurugroup.com.

2. METODE PENELITIAN

2.1. Alur Penelitian

Tahapan penelitian yang dimulai dari mengumpulkan data sampai menghasilkan prediksi harga rumah disajikan pada Gambar 1.



Dari gambar 1 dapat dijelaskan lebih detil bahwa dalam langkah analisis data eksplorasi terdapat proses pengambilan data, deskripsi variabel, menangani nilai yang sangat bereda dan data yang tidak ada, analisis univariat dan multivariat. Beberapa proses yang berada di dalam persiapan data adalah *encoding* data, reduksi dimensi menggunakan PCA, membagi dataset, dan normalisasi data.

2.2. Pengumpulan Data

Data dalam penelitian ini adalah hasil *scraping* dari situs propertygurugroup.com untuk penjualan rumah di kota Surakarta. Terdapat sepuluh kolom atau variabel yang disajikan seperti pada tabel 1.

Tabel 1. Deskripsi Dataset

kolom	tipe data	keterangan
jml_kt	float	jumlah kamar tidur
jml_km	float	jumlah kamar mandi
luas_tn	float	luas tanah (meter persegi)
alamat	object	alamat atau lokasi rumah
luas_bg	float	luas bangunan(meter persegi)
sertifikat	object	sertifikat tanah
interior	object	jenis interior
garasi	float	ukuran garasi
listrik	float	tegangan listrik (watt)
harga	float	harga rumah

2.3. Analisis Data Eksplorasi

Analisis data eksplorasi termasuk dalam proses *data understanding* yang terdiri dari lima tahapan, yaitu: pengambilan data, identifikasi data, menangani *outliner* dan *missing value*, analisis univariat, dan analisis multivariat. Metode *Interquartile Range* (IQR) merupakan salah satu cara untuk deteksi *outliner*, yang menjelaskan bahwa yang termasuk *outliner* adalah nilai yang tidak masuk dalam rentang Q3 sampai Q1 (interquartil) sehingga persamaan untuk mengetahui IQR adalah:

$$IQR = Q3 - Q1 \quad (1)$$

(Saputra, Saadah and Yunanto, 2021)

2.4. Persiapan Data

Tahapan ini berupa *data encoding*, yaitu merubah kategori data menjadi numerik. Selanjutnya proses reduksi dimensi dengan PCA, kemudian memisah data menjadi *data training* dan *testing*. Proses berikutnya adalah metode RF dengan melakukan proses normalisasi untuk merubah skala pada dataset agar mendekati distribusi normal.

2.5. PCA

Terdapat beberapa langkah dalam PCA:

1. Normalisasi Data: data harus memiliki skala dalam bentuk yang seragam sehingga dalam membandingkan antar fitur tidak terpengaruh oleh perbedaan skala.
2. Mengetahui kovariansi dengan menghitung variansi dan korelasi antar fitur yang sudah dalam bentuk normal dengan persamaan:

$$\text{var}(A_1) = \frac{\sum_{i=1}^n (x_i - \bar{x})}{(n - 1)} \quad (2)$$

$$\text{cov}(A_1, A_2) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n - 1)} \quad (3)$$

3. Menghitung *eigen vector* dan nilai *eigen* dengan persamaan:

$$C_v = \lambda_v \quad (4)$$

(Saputra et al., 2023)

4. Memilih *eigen vector* yang memiliki variansi terbesar dan proyeksi data.

2.6. Random Forest

Diambil 0,75 data acak yang akan digunakan sebagai pohon, untuk membentuk beberapa pohon lagi sehingga menjadi hutan. Dari data tersebut dicari variabel paling penting (VI) yang dipengaruhi oleh nilai *out of bag* (OBB). Persamaan untuk menghitung OBB adalah:

$$\text{errOBB} = \frac{1}{n-z} \sum_{i=1}^{n-z} (y_i - \bar{y}_i)^2 \quad (5)$$

n=data obeservasi
z=sample

y_i =data kei-i
 \hat{y}_i =prediksi ke-i

$$VI(x^j) = \frac{1}{2} \sum_t^s errOBB_t^j - errOBB_i \quad (6)$$

(Wiratman and Wella, 2024)

2.7. Menghitung RMSE dan Evaluasi

Melakukan evaluasi model regresi dengan mengetahui *error* yaitu perbedaan nilai asli dengan nilai prediksi. Persamaan RMSE dilakukan berulang kali sebanyak data yang ada dengan persamaan:

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (A_t - F_t)^2}{n}} \quad (7)$$

(Rahman et al., 2022)

3. HASIL DAN PEMBAHASAN

3.1. Data yang dikumpulkan

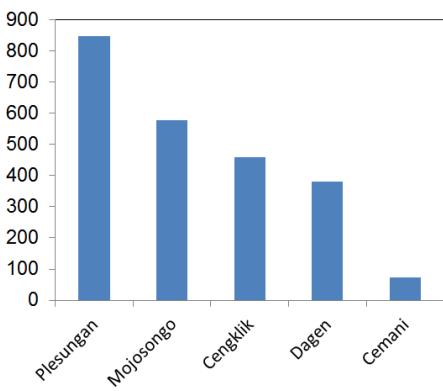
Beberapa data yang digunakan pada penelitian ini yang diambil dari hasil *scraping* situs propertygurugroup.com untuk penjualan beberapa rumah di kota Surakarta

kamar tidur	kamar mandi	luas tanah	harga per m ²	alamat	luas bangunan	sertifikat	interior	garasi	listrik	harga
5	2	136	2860000	Plesungan	110 SHM	sebagian	2	1300	2.90000E+08	
2	1	124	3100000	Mojosongo	95 SHM	tanpa	1	1300	2.98500E+08	
3	1	117	2400000	Dagen	110 SHM	sebagian	1	1300	2.88000E+08	
4	2	70	3250000	Cenglik	70 SHM	lengkap	1	900	2.31500E+08	
4	1	90	2670000	Plesungan	80 SHM	tanpa	1	1300	2.17600E+08	

Gambar 2. Dataset

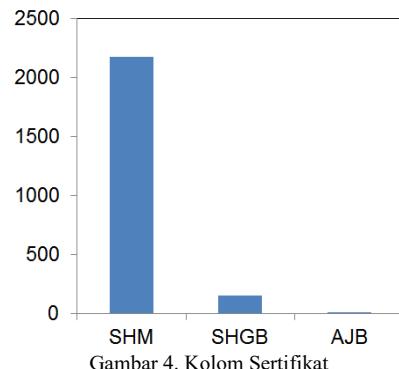
3.2. Analisis Data Eksplorasi

1. Analisis Univariat



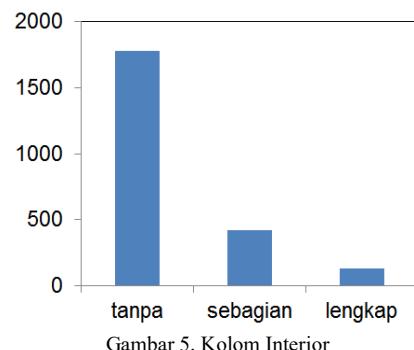
Gambar 3. Kolom Lokasi

Terdapat tiga jenis data yang memiliki sifat kategorik yaitu alamat, interior dan sertifikat. Dari gambar 3 dapat diketahui bahwa daerah plesungan jumlah penjualan rumahnya terbanyak yaitu 847 atau 36,2% sedangkan penjualan daerah cemani paling sedikit, hanya 3,2% atau sebanyak 74.



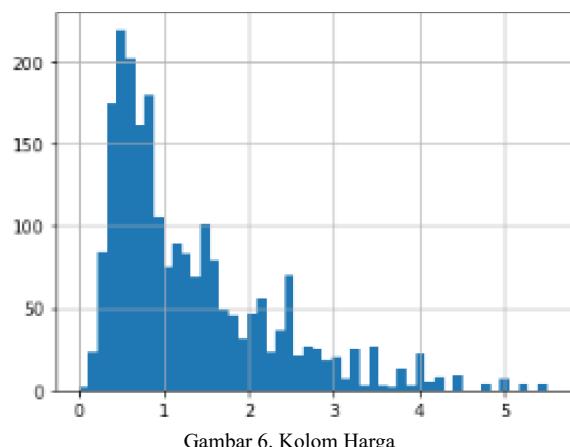
Gambar 4. Kolom Sertifikat

Dari gambar 4 didapat informasi bahwa rumah yang memiliki sertifikat hak milik (SHM) jumlah penjualannya paling banyak yaitu 93,1% dan rumah bersertifikat AJB hanya 0,4% atau 10.



Gambar 5. Kolom Interior

Dari gambar 5 diketahui untuk rumah tanpa interior paling banyak terjual dengan jumlah 1783 (76,2%) sedangkan paling sedikit adalah rumah dengan interior lengkap sebanyak 134 (5,8%).

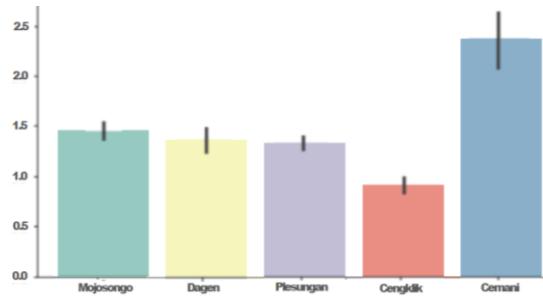


Pada gambar 6 dapat dilihat bahwa grafik yang semakin menurun seiring bertambahnya jumlah

sampel, jadi kenaikan harga rumah sebanding dengan turunnya jumlah rumah atau *sample*.

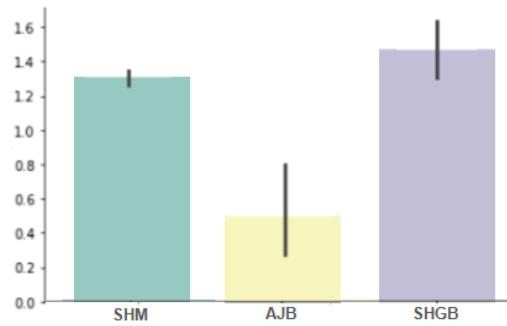
2. Analisis Multivariat

Analisis ini berguna untuk mengetahui relasi antara 2 variabel atau lebih yang digunakan dalam penelitian.



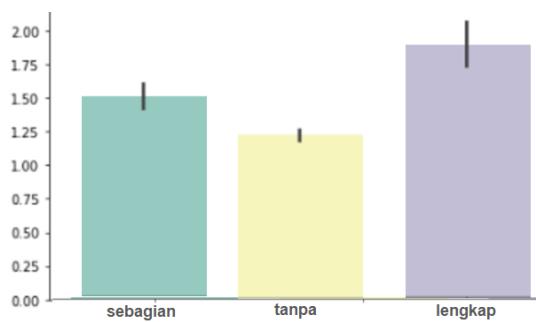
Gambar 7. Rerata Harga Relatif Terhadap Lokasi

Rerata harga di ketiga lokasi yaitu mojosongo, dagen dan plesungan berkisar di 1,3 Miliar Rupiah. Cemani merupakan lokasi dengan rerata paling tinggi sebesar 2,4 Miliar Rupiah.



Gambar 8. Rerata Harga Relatif Terhadap Jenis Sertifikat

Rerata harga terhadap jenis sertifikat rumah yang memiliki SHGB dan SHM lebih banyak diminati.



Gambar 9. Rerata Harga Relatif Terhadap Kelengkapan Interior

Rerata harga terhadap kelengkapan interior jelas yang memiliki interior lengkap yang berharga paling tinggi.

Corelation Matrik untuk fitur Numerik

	kamar tidur	kamar mandi	luas tanah(m2)	harga per m	luas Bangunan(m2)	parkir	listrik	harga	
kamar tidur	1	0.72	0.55	0.42	0.72	0.34	0.37	0.6	
kamar mandi	0.72	1	0.6	0.5	0.74	0.35	0.47	0.67	
luas tanah(m2)	0.55	0.6	1	0.24	0.77	0.44	0.44	0.83	
harga per m	0.42	0.5	0.24	1	0.45	0.21	0.43	0.68	
luas Bangunan(m2)	0.72	0.74	0.77	0.45	1	0.44	0.46	0.78	
parkir	0.34	0.35	0.44	0.21	0.44	1	0.35	0.44	
listrik	0.37	0.47	0.44	0.43	0.46	0.35	1	0.55	
harga	0.6	0.67	0.83	0.68	0.78	0.44	0.55	1	
	kamar tidur	kamar mandi	luas tanah(m2)	harga per m	luas Bangunan(m2)	parkir	listrik	harga	

Gambar 10. Korelasi Matrik Fitur Numerik

Dari gambar 10 dapat dijelaskan bahwa terdapat dua variabel yang paling berpengaruh terhadap harga perumahan, yaitu tanah luas tanah dan luas bangunan. Darasi adalah variabel yang paling tidak berpengaruh atau korelasi terhadap harga paling rendah.

3.3. Persiapan Data

Data kategorikal dikonversi menjadi data numerikal. Yang termasuk data kategorikal dikonversi menjadi variabel baru. Variabel alamat ada 5 tipe, interior ada 3 tipe, dan jenis sertifikat ada 3 tipe, jadi berdasarkan nilai variabel ini ada 11 variabel baru dengan nilai 0 dan 1. Bentuk data yang telah dikonversi disajikan pada Tabel 2.

Tabel 2. Dataset Setelah Dikonversi

variabel	tipe	tipe data
jumlah kamar tidur	fitur	float
jumlah kamar mandi	fitur	float
luas tanah	fitur	float
harga per meter	fitur	float
cut_dagen	fitur	float
cut_cengklik	fitur	float
cut_cemani	fitur	float
cut_plesungan	fitur	float
cut_mojosongo	fitur	float
luas bangunan	fitur	float
cut_shgb	fitur	float
cut_ajb	fitur	float
cut_shm	fitur	float
cut_tanpa interior	fitur	float
cut_interior lengkap	fitur	float
cut_interior sebagian	fitur	float
listrik	fitur	float
harga	target	float

3.4. PCA

Pada tahapan ini dibagi menjadi tiga proses reduksi yaitu untuk mereduksi fitur lokasi, fitur jenis sertifikat dan fitur kelengkapan interior. Dari setiap reduksi masing-masing fitur terdapat proses normalisasi data, menghitung nilai varian dan kovarian, menghitung eigen vektor dan nilai eigen,

dan yang terakhir adalah proyeksi data. Dalam artikel ini disajikan satu proses dari fitur lokasi yang merupakan hasil dari *running program*.

array([3.20449173e-01, 2.19495270e-01, 1.68047804e-01, 2.22429467e-01, 2.32677342e-31])
array([[0.19096969, -0.04905121, -0.00466612, -0.09719197, -0.04006039], [-0.04951121, 0.15451133, -0.0346741, -0.07222369, -0.02976981], [-0.00466612, -0.0346741, 0.01783586, -0.00687047, -0.00283186], [-0.09719197, -0.07222369, -0.00687047, 0.23527159, -0.05898547], [-0.04006039, -0.02976981, -0.00283186, -0.05898547, 0.13164673]])

Gambar 11. Nilai Varian dan Kovarian Fitur Lokasi

array([[-0.51364818, -0.18709157, -0.00762386, 0.8286399, -0.12028428], [-0.67390454, 0.65784316, 0.00886913, -0.23409121, 0.24128345], [-0.1836778, -0.52973951, 0.00770395, -0.11434964, 0.820063], [-0.21974737, -0.22718974, 0.89431754, -0.21326366, -0.23411677], [0.4472136, 0.4472136, 0.4472136, 0.4472136, 0.4472136]])

Gambar 12. Nilai Eigen Vektor Fitur Lokasi

Setelah diketahui nilai *Eigen Vektor*, kemudian diurutkan secara *descending*. Nilai *eigen* yang paling besar memiliki informasi yang paling tinggi yang akan dijadikan dasar dalam reduksi data. Proporsi informasi dari fitur lokasi adalah : (0,428, 0,311, 0,24, 0,03, 0). Pada tabel 3 disajikan beberapa hasil reduksi dari fitur lokasi.

Tabel 3. Nilai Reduksi Fitur Lokasi

Id	Lokasi	Id	lokasi	Id	lokasi
1	-0,6405	10	-0,24715	19	-0,31396
2	-0,24715	11	-0,24715	20	-0,31396
3	0,701776	12	0,701776	21	0,701776
4	0,701776	13	-0,31396	22	-0,6405
5	0,701776	14	-0,6405	23	-0,31396
6	0,701776	15	0,701776	24	-0,6405
7	0,701776	16	0,701776	25	-0,64056
8	0,701776	17	-0,31396		
9	-0,24715	18	-0,31396		

Normalisasi merupakan tahapan untuk merubah nilai yang ada menjadi nilai dengan rentang 0-1. Gambar 13 adalah hasil normalisasi dari beberapa data yang ada.

harga	kamar tidur	kamar mandi	luas tanah...	harga per m	luas Bangun...	listrik	alamat	sertifikat
0.536	0.625	1	0.649	0.430	0.947	0.574	0	0
0.134	0.125	0.200	0.181	0.360	0.124	0.279	0.287	0
0.268	0.125	0.200	0.193	0.727	0.124	0.279	1	0
0.089	0.250	0	0.147	0.282	0.124	0.279	1	0
0.071	0.125	0	0.147	0.215	0.082	0.279	1	0
0.089	0.250	0	0.147	0.282	0.124	0.279	1	0
0.081	0.125	0	0.193	0.184	0.093	0.279	1	0
0.223	0.500	0.400	0.274	0.412	0.420	0.574	0.287	0
0.107	0.125	0	0.209	0.237	0.124	0.279	0.287	0
0.143	0.250	0.200	0.186	0.378	0.146	0.279	0.287	0
0.253	0.125	0.200	0.227	0.605	0.156	0.539	1	0
0.509	0.500	0.400	0.578	0.480	0.631	0.279	1	0
0.500	0.750	0.800	0.397	0.675	0.578	1	0.281	0

Gambar 13. Hasil Normalisasi

3.5. Random State dan Menghitung RMSE

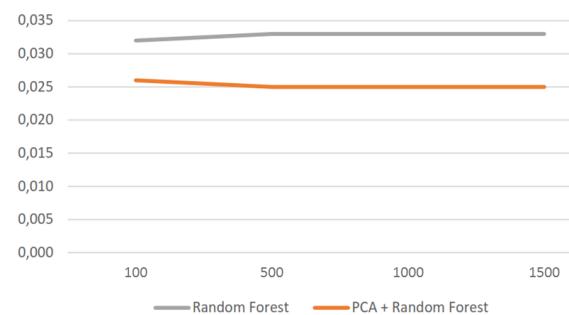
Hasil penelitian ini adalah hasil prediksi harga yang menggunakan metode RF saja dan prediksi harga dari hasil dua metode yaitu RF dan PCA. Dari perbandingan kedua hasil tersebut dapat diketahui akurasi penggabungan dua metode dalam menentukan prediksi harga rumah. PCA digunakan untuk reduksi variabel lokasi dan sertifikat.

Untuk melakukan pengujian dan perbandingan dua metode, digunakan matrik RMSE dengan menggunakan persamaan 7. Dalam penelitian ini

dilakukan pengujian dengan empat iterasi yang hasilnya disajikan dalam tabel 4.

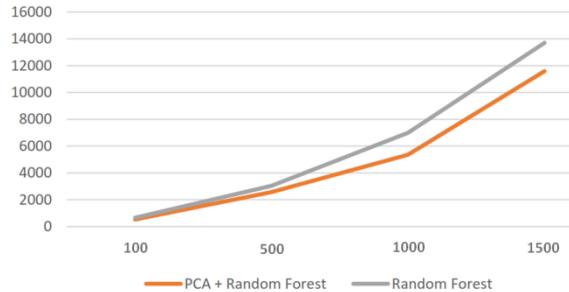
Tabel 4. Hasil Pengujian Perbandingan Dua Metode

Iterasi	PCA dan RF		RF	
	Kesalahan	Waktu	Kesalahan	Waktu
100	0,027	527	0,032	657
500	0,026	2562	0,033	3047
1000	0,025	5348	0,034	6996
1500	0,025	11586	0,034	13697
Rerata	0,0257	5005,75	0,0332	6099,25



Gambar 14. Grafik Hasil RMSE

Dari gambar 14 dapat dijelaskan dari pengujian menggunakan RMSE didapat nilai *error* 0,032 pada iterasi 100, pada iterasi ke 500 nilai *error*=0,033 dan nilai *error* 0,034 pada iterasi 1000 juga 1500.



Gambar 15. Waktu Pelatihan Model

Dari tabel 4 dan gambar 15 dapat dijelaskan waktu pelatihan metode RF pada suatu iterasi adalah: 100 butuh waktu 657 mili detik, 500 butuh waktu 3047 mili detik, 1000 butuh waktu 6996 mili detik, 1500 butuh waktu 13697 mili detik. Untuk PCA dan RF pada suatu iterasi didapatkan: 100 butuh waktu 527 mili detik, 500 butuh waktu 2562 mili detik, 1000 butuh waktu 5348 mili detik dan 1500 butuh waktu 11586 mili detik. Berdasar pengujian menggunakan RF dan PCA hasil tingkat *error* lebih kecil dan bernilai lebih konsisten dengan rerata 0,0257. Pengujian menggunakan RF saja (tanpa PCA) tingkat kesalahan lebih besar dengan rerata 0,0332. Waktu pelatihan dengan model PCA lebih cepat dengan rerata 5005,75, untuk tanpa PCA waktu rerata sebesar 6099,25.

4. KESIMPULAN

Kesimpulan hasil analisis menunjukkan bahwa jumlah penjualan rumah tertinggi adalah daerah Plesungan, dan penjualan rumah yang memiliki sertifikat hak milik juga paling tinggi. Dari sepuluh fasilitas atau variabel yang ada, disimpulkan bahwa

luas tanah dan bangunan paling berpengaruh terhadap harga jual, dan listrik merupakan variabel yang paling tidak berpengaruh karena memiliki nilai korelasi paling rendah.

Kesimpulan dari hasil pelatihan model menunjukkan bahwa peggabungan metode *random forest* dan PCA menghasilkan nilai yang lebih optimal dibanding hanya menggunakan metode *random forest* saja. Tingkat kesalahan dalam metode PCA lebih kecil, dengan rerata 0,0257 maka nilainya lebih konsisten dibanding hanya menggunakan metode *random forest* yang nilai kesalahannya lebih besar dengan rerata 0,0332. Waktu pelatihan model menggunakan PCA lebih cepat (5005,75) dibanding hanya menggunakan metode *random forest* (6099,25).

DAFTAR PUSTAKA

- ANALISA, F.C.K. and OKADA, S., 2023. Tiny house characteristics in Indonesia based on millennial's user preference. *Urban, Planning and Transport Research*, [online] 11(1). <https://doi.org/10.1080/21650020.2023.2166095>.
- CAHYANI PUTRI, N.A. and ARIANTO, D.B., 2024. Komparasi Penggunaan Information Gain Pada Machine Learning untuk Memprediksi Harga Rumah di Jabodetabek. *Jurnal Sains dan Teknologi*, 5(3), pp.756–762. <https://doi.org/10.55338/saintek.v5i3.2052>.
- DHAKA, N., CHAUDHARY, A., SISODIA, D., SHARMA, M. and BABU, S., 2024. Prediction of House Pricing Using Machine Learning. *Tujin Jishu/Journal of Propulsion Technology*, 45(2), pp.1026–1034. <https://doi.org/10.1109/ICAC3N60023.2023.10541549>.
- IGAMO, A.M., AZWARDI, A., SAPUTRA, A., ISMAIL, R.G., GUSTRIANI, G. and MELLINY, V.D., 2023. Monetary Policy and Demographics: Empirical Evidence for Housing Prices in Indonesia. *Sriwijaya International Journal of Dynamic Economics and Business*, 6(4), pp.371–384. <https://doi.org/10.29259/sijdeb.v6i4.371-384>.
- JÁUREGUI-VELARDE, R., ANDRADE-ARENAS, L., CELIS, D.H., DÁVILA-MORÁN, R.C. and CABANILLAS-CARBONELL, M., 2023. Web Application with Machine Learning for House Price Prediction. *International Journal of Interactive Mobile Technologies*, 17(23), pp.85–104. <https://doi.org/10.3991/IJIM.V17I23.38073>.
- LU, Y., SHI, V. and PETTIT, C.J., 2023. The Impacts of Public Schools on Housing Prices of Residential Properties: A Case Study of Greater Sydney, Australia. *ISPRS International Journal of Geo-Information*, 12(7). <https://doi.org/10.3390/ijgi12070298>.
- PARHUSIP, H.A., TRIHANDARU, S., HERIADI, A.H., SANTOSA, P.P. and PUSPASARI, M.D., 2022. Data Exploration Using Tableau and Principal Component Analysis. *International Journal on Informatics Visualization*, 6(4), pp.911–920. <https://doi.org/10.30630/jiov.6.4.952>.
- PERDAMAIAN, L.G. and ZHAI, Z. (John), 2024. Status of Livability in Indonesian Affordable Housing. *Architecture*, 4(2), pp.281–302. <https://doi.org/10.3390/architecture4020017>.
- PITALOKA, E., HARTANTO, T.B.A. and SANDIWARNO, S., 2024. Penerapan Machine Learning Untuk Prediksi Bencana Banjir. *Jurnal Sistem Informasi Bisnis*, 14(1), pp.62–76. <https://doi.org/10.21456/vol14iss1pp62-76>.
- PRAMUKANTORO, E., SAKTI, AMRON, K., WARDHANI, V. and KAMILA, P.A., 2024. Implementasi Sensor Polar H10 dan Raspberry Pi dalam Pemantauan dan Klasifikasi Detak Jantung Beberapa Individu Secara Simultan dengan Pendekatan Machine Learning. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(1), pp.175–182. <https://doi.org/10.25126/jtiik.2024117716>.
- PRASETYO, T., PUTRI, R.A., RAMADHANI, D., ANGRAINI, Y. and NOTODIPUTRO, K.A., 2024. Perbandingan Kinerja Metode Arima, Multi-Layer Perceptron, Dan Random Forest Dalam Peramalan Harga Logam Mulia Berjangka Yang Mengandung Penculan. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 11(2), pp.265–274. <https://doi.org/10.25126/jtiik.20241127392>.
- RAHADI, R.A., WIRYONO, S.K., NAINGGOLAN, Y.A., AFGANI, K.F., YAMAN, R., AZMI, A.S.M., ISMAIL, F.Z., SAPUTRA, J., RAHMAWATI, D. and MOULYNI, A., 2020. Determining the factors influencing residential property price: A comparative study between Indonesia and Malaysia. *Desision science letters*, 9, pp.421–438. <https://doi.org/10.5267/dsl.2022.6.002>.
- RAHMAN, M.K.A., RASHID, N.E.A., ISMAIL, N.N., ZAKARIA, N.A.Z., KHAN, Z.I., RAHIM, S.A.E.A. and ISA, F.N.M., 2022. Hand Gesture Recognition Based on Continuous Wave (CW) Radar Using Principal Component Analysis (PCA) and K-Nearest Neighbor (KNN) Methods. *International Journal on Informatics Visualization*, 6(1–2), pp.188–194. <https://doi.org/10.30630/jiov.6.1-2.926>.
- RAMADHANTI, R., HAIRANI, H. and INNUDDIN, M., 2023. Electric Vehicle Sales-Prediction Application Using Backpropagation Algorithm Based on Web. *International Journal of Engineering and*

- Computer Science Applications (IJECSA), 2(2), pp.73–80. <https://doi.org/10.30812/ijecsa.v2i2.3388>.
- SAPUTRA, E.P., NURAJIZAH, S., MAULIDAH, M., HIDAYATI, N. and RAHMAN, T., 2023. Komparasi Machine Learning Berbasis Pso Untuk Prediksi Tingkat Keberhasilan Belajar Berbasis E-Learning. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 10(2), pp.321–328. <https://doi.org/10.25126/jtiik.20231026469>.
- SAPUTRA, I.N.Y., SAADAH, S. and YUNANTO, P.E., 2021. Analysis of Random Forest, Multiple Regression, and Backpropagation Methods in Predicting Apartment Price Index in Indonesia. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, 7(2), p.238. <https://doi.org/10.26555/jiteki.v7i2.20997>.
- SRIVASTAVA, K., VERMA, S., KHAN, M.S. and SINGH, A., 2021. House Price Prediction Using Machine Learning. In: *Proceedings - 2021 3rd International Conference on Advances in Computing, Communication Control and Networking, ICAC3N 2021*. pp.203–206. <https://doi.org/10.1109/ICAC3N53548.2021.9725552>.
- TANAMAL, R., MINOQUE, N., WIRADINATA, T., SOEKAMTO, Y. and RATIH, T., 2023. House Price Prediction Model Using Random Forest in Surabaya City. *TEM Journal*, 12(1), pp.126–132. <https://doi.org/10.18421/TEM121-17>.
- UTAMA, A.A.G.S., 2022. The Best Model And Variables Affecting Housing Values Of Big Cities In Indonesia. *Galaxy International Interdisciplinary Research Journal (GIIRJ)*, [online] 10(6), pp.782–793. Available at: <<https://internationaljournals.co.in/index.php/giirj/article/view/2238>>.
- WINANTO, E.A., NOVIANTO, Y., WIJAYA, I.S., JUSIA, P.A., BANGSA, U.D., KORESPONDENSI, P., FORCE, B., ANALYSIS, P.C. and DETECTION, T., 2024. Peningkatan Performa Deteksi Serangan Menggunakan Metode PCA Dan Random Forest. 11(2), pp.285–290. <https://doi.org/10.25126/jtiik.20241127678>.
- WIRATMAN, A.B. and WELLA, 2024. Personalized Learning Models Using Decision Tree and Random Forest Algorithms in Telecommunication Company. *International Journal on Informatics Visualization*, 8(1), pp.318–325. <https://doi.org/10.62527/jiov.8.1.1905>.