JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.7256

COMPARISON OF PRINCIPAL COMPONENT ANALYSIS AND RANDOM FOREST ALGORITHM FOR PREDICTING HOUSING PRICES

Dahlan Susilo[1*]; Diyah Ruswanti[1]; Supriyanta[2]; Wawan Nugroho[2]

Informatika[1]
Universitas Sahid, Surakarta, Indonesia[1]
https://www.usahidsolo.ac.id[1]
dahlan.susilo@usahidsolo.ac.id*, dyahruswanti@usahidsolo.ac.id

Sistem Informasi[2]
Universitas Bina Sarana Informatika, Surakarta, Indonesia [2]
https://www.bsi.ac.id/ubsi/index.js[2]
supriyanta.spt@bsi.ac.id, wawan.wgh@bsi.ac.id

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *House price predictions are an important thing in the property industry and are useful for buyers in making decisions. Principal Component Analysis (PCA) and Random Forest (RF) methods were used for accuracy analysis in predicting housing prices. Purpose of this research is to measure the accuracy of both methods also to compare RF method optimized with PCA and the one that has not been optimized. The data used is house prices in Karanganyar city based on data scraping results on the rumah123.com site. The analysis reveals that Jaten has the highest number of house sales, and sales of houses with land ownership certificates are also the highest. Of the 10 variables used, land area and buildings have the most influence on selling prices. The model training results show that the RF and PCA methods combination has more optimal value than only using the RF method. The error rate of the PCA method is smaller, averaging 0.0257, making its value more consistent than using only 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 only using the RF method (6099.25).*

Keywords*: housing, PCA, prediction, price, random forest.*

Intisari— Prediksi harga rumah merupakan hal penting dalam industri properti dan berguna bagi pembeli dalam mengambil keputusan. Metode Principal Component Analysis (PCA) dan Random Forest (RF) digunakan untuk analisis akurasi dalam memprediksi harga perumahan. Tujuan penelitian ini untuk mengukur akurasi kedua metode juga membandingkan metode RF yang sudah dioptimalkan dengan PCA dan yang belum. Data yang digunakan adalah harga rumah di kota Karanganyar berdasarkan hasil scraping data pada situs rumah123.com. Hasil analisis menunjukkan bahwa wilayah Jaten merupakan wilayah dengan jumlah penjualan rumah tertinggi, dan penjualan rumah yang memiliki sertifikat hak atas tanah juga merupakan yang tertinggi. Dari 10 variabel yang digunakan, luas tanah dan bangunan memiliki pengaruh paling besar terhadap harga jual. Hasil pelatihan model menunjukkan bahwa kombinasi metode RF dan PCA memiliki nilai yang lebih optimal daripada hanya menggunakan metode RF. Tingkat kesalahan pada metode PCA lebih kecil, dengan rata-rata 0,0257, sehingga nilainya lebih konsisten daripada hanya menggunakan metode RF, yang memiliki nilai kesalahan lebih besar dengan rata-rata 0,0332. Waktu pelatihan model menggunakan PCA lebih cepat (5005,75) dibandingkan hanya menggunakan metode RF (6099,25).

Kata Kunci*: perumahan, PCA, prediksi, harga, hutan acak.*

INTRODUCTION

Property is a unique product that cannot be contrasted with other commercial products due to two pricing conditions [1]. Property pricing is one of the crucial aspects in property development activities considering as it will affect the profit margin gained by developers and property purchasing decisions [2]. First, the conditions provided by the product are natural conditions inherent to the product, including aspects such as property location, landforms, physical condition, and natural resources. Second, conditions created by property developers to increase product value, such as facility development, security features, gateless access, and design [3] [4]. As a pricing strategy, property developers often incorporate relevant property development costs and established profit margins into the sales price [5].

Residential property prices rise annually, influenced by various factors such as location, lot size, facilities, and other considerations [6]. Therefore, house price prediction is an important aspect of the property industry and is useful for buyers in making decisions [7]. A system or mechanism is needed that can predict future house prices. One such mechanism is machine learning, which can improve and predict house prices with a high degree of accuracy [8]. For years, house price prediction has been a major research topic, as housing demand continues to skyrocket. It is crucial to develop an appropriate framework that enables buyers and sellers to make swift decisions regarding buying or selling property [9].

Machine learning is a method of data analysis for the automation of construction in analytical models [10]. Machine learning is a branch of artificial intelligence and is based on the idea that a system can learn from data, identify patterns on its own, and make decisions with minimal human intervention [11]. Alongside big data and high-performance computing, machine learning has developed to build new opportunities in various operational contexts for explaining, measuring, and acquiring data-intensive processes [12]. Random forest is one of the machine learning models commonly used for classification and forecasting [13]. To train machine learning algorithms and artificial intelligence models, it is essential to have a large amount of high-quality data for effective data collection [14]. System performance data is crucial for refining algorithms, improving software and hardware efficiency, evaluating user behavior, enabling pattern identification, decision-making, predictive modeling, and troubleshooting, ultimately leading to increased effectiveness and accuracy [15] [16]. Principal Component Analysis (PCA) is one of method for reducing dimensionality in a data set, and can increase interpretability without losing a lot of information [17]. This can be achieved by creating new covariates that are not related to each other. Finding these new variables, or what we call principal components, will reduce the problem of solving for eigenvalues or eigenvectors [18]. PCA can be considered an adaptive data analysis technology because the variables are developed to adapt to different data types and structures [19].

The results of previous studies that also used the random forest method made an analysis for predicting house prices in the city of Surabaya. The purpose of this study was to determine the fairness of property prices, whether the price is below or above the market or according to market prices. Several trials were conducted to achieve high prediction values; the highest prediction value was achieved using 80% of the data set for training and 20% of the data set for testing. The classifier using the random forest method produced the highest accuracy and F1 score, 88%, among all other classification methods. With this accuracy, it is hoped that homeowners can find out whether the previously determined price is too low, average, or high [20]. Other similar studies develop a web-based system to predict the selling price and purchase price of two types of houses, namely urban houses and rural houses. The development of the system uses or follows the construction stages in machine learning, and the Rational Unified Process methodology. The developed system can generate price predictions from the location and several home facilities desired by prospective consumers. The system accuracy test by 10 experts in the field of machine learning and property entrepreneurs using linear regression produced a value of 4.88 out of 5, so the accuracy is very high [21].

There were several differences among the three previous studies, including the methods that were utilized and the procedure for testing accuracy. In this study, the PCA and RF methods were used to determine the house price prediction. From the two methods used, an analysis was conducted on their accuracy in predicting house prices and to determine the effect of using PCA in optimizing the RF method. This research utilized data on Karanganyar house prices, which was scraped from the rumah123.com website. PCA method was a suitable choice as it reduced variable complexity, which in turn accelerated the computation time. In PCA, eigenvectors represent principal components direction, and eigenvalues indicate the amount of variance. By choosing

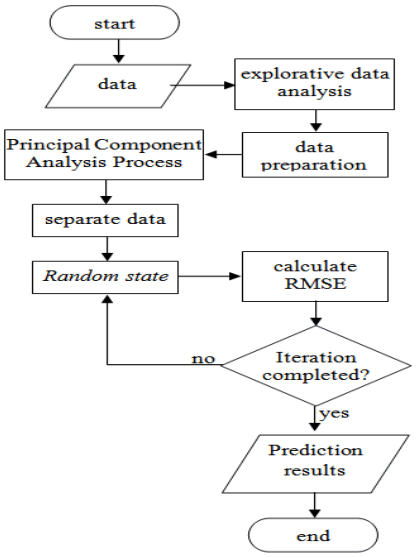
Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.7256

principal components with the largest eigenvalues, data's dimensions can be reduced for easier analysis. From the two methods used, an analysis was carried out for accuracy in generating house price predictions and to determine the effect of using PCA in optimizing the RF method. The data collected and used as research material are house prices in Karanganyar city based on data scraping results from rumah123.com site. PCA is very suitable for data exploration and dimensionality reduction, while RF is a powerful prediction algorithm. Using PCA as a preprocessing step for RF can provide a balance between computational efficiency and accuracy, especially when dealing with high-dimensional data [22].

MATERIALS AND METHODS

The research stages, starting from data collection to generating house price predictions, are shown in Figure 1.



Source: (Research Results, 2025)
Figure 1. Research Flow Diagram

Figure 1 illustrates in more detail that the exploratory data analysis phase involves the process of data acquisition, variable description, addressing outliers and missing data, as well as conducting univariate and multivariate analysis. Some of the processes included in data preparation are data encoding, dimensionality reduction using PCA, splitting the dataset, and data normalization.

Data in this study was obtained by scraping the rumah123.com site for house sales in Karanganyar city. The dataset consists of 6,130 rows and 10 columns. There are ten columns or variables presented as in table 1.

Table 1. Dataset Description

Coloum	Data type	Description
jml_kt	float	Total bedrooms
jml_km	float	Total bathrooms
luas_tn	float	land area (square meters)
address	object	house address or location
large_bg	Float	building area (square meters)
certificate	object	land certificate
interior	object	interior type
garage	float	garage size
electric	float	electric voltage (watts)
price	float	house price

Source: (Research Results, 2025)

Exploratory Data Analysis (EDA) is part of the data understanding process, which consists of five stages: data acquisition, data identification, handling outliers and missing values, univariate analysis, and multivariate analysis. The Interquartile Range (IQR) method is a way to detect outliers, which explains that outliers are values that fall outside the range from Q3 to Q1 (the interquartile). Therefore, the equation to find the IQR is:

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

[23]

The data preparation stage is data encoding, which is changing the data category into numeric. Next step is the dimension reduction process with PCA, then separating data into training and testing data. Next process is the RF method by carrying out a normalization process to change the dataset scale to approach a normal distribution [24].

There are five steps in PCA:

1. Data Normalization: data must have a scale in a uniform form so that comparing between features is not affected by differences in scale.
2. Knowing the covariance by calculating the variance and correlation between features that are already in normal form with the formula:

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

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

3. Calculate eigenvectors and eigenvalues using the formula:

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

[25].

4. Select the eigenvector that has the largest variance and data projection.
5. Data projection: Data can be compressed by projecting data into spatial subspaces defined by principal components.

0.75 random data samples will be taken to be used as a tree, in order to form several more trees to create a forest. From that data, the most

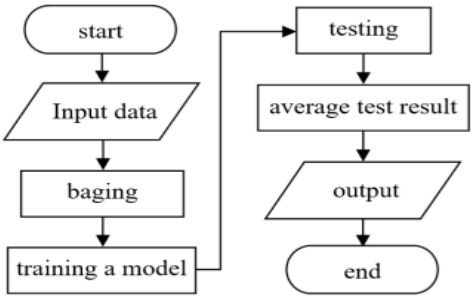
important variable (VI) is sought, which is affected by the out-of-bag (OOB) value [26]. An formula for calculating OBB is:

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

n= observation data yi= i-th data
z= sample $\hat{y}i$= i-th prediction

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

[27]



Source: (Research Results, 2025)
Figure 2. Random Forest Method Flowchart

1. The input data is then split into a training set and a testing set.
2. The training data enters the bagging or bootstrap aggregation stage to create training data samples that are different from the others.
3. The model used to train each sample using the decision tree algorithm.
4. The trained models will be tested with the test data. Because an ensemble learning system works with several methods that cooperate to perform its function, the results of each model will be averaged for regression cases, and for classification cases, the most frequent mode value will be used.

The next step is to evaluate the regression model by determining the error, which is the difference between the actual values and the predicted values. The RMSE equation is calculated repeatedly for all the data using the following equation:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n} (At - Ft)^2}{n}} \quad (7)$$

[28]

RESULTS AND DISCUSSION

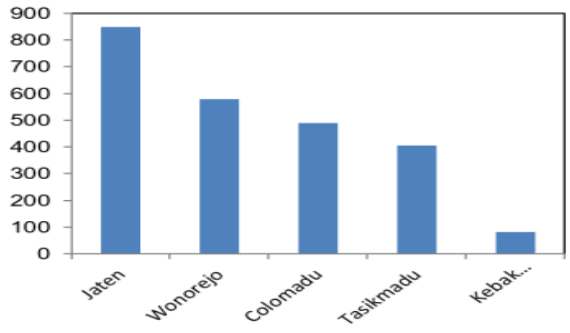
This study utilized data extracted via web scraping from the rumah123.com website, focusing on house sales in Karanganyar.

Table 2. Dataset

bedroom	bathroom	surface area	price /meter	address	building area	certificate	interior	garage	electricity	price
5	2	136	26000000	Plesungan	110	SHM	Partial	2	1300	2,90000E+08
2	1	124	31000000	Mojosongo	95	SHM	without	1	1300	2,96500E+08
3	1	117	24000000	Degan	110	SHM	partial	1	1300	2,68000E+08
4	2	70	22000000	Cengklik	70	SHM	complete	1	900	2,31500E+08
4	1	90	26700000	Plesungan	80	SHM	without	1	1300	2,17600E+08

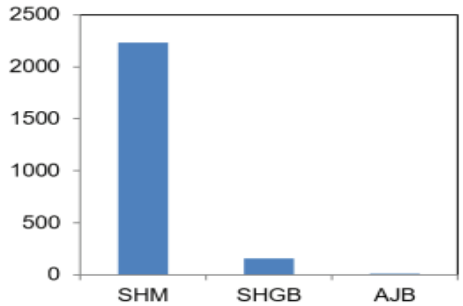
Source: (Research Results, 2025)

In univariate analysis, there are three types of data that have categorical properties, namely address, interior, and certificate.



Source: (Research Results, 2025)
Figure 3. Location Field

From Figure 3, it can be seen that Jaten area has the highest number of house sales, 849 or 35.3%, while Kebak Kramat area has the lowest number of sales, only 3.4% or 82.

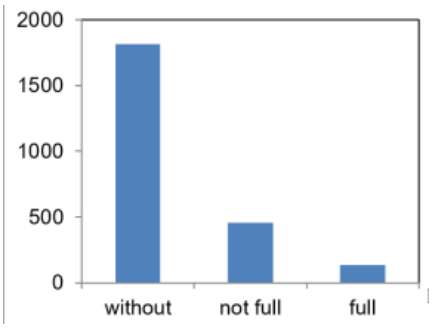


Source: (Research Results, 2025)
Figure 4. Certificate Field

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

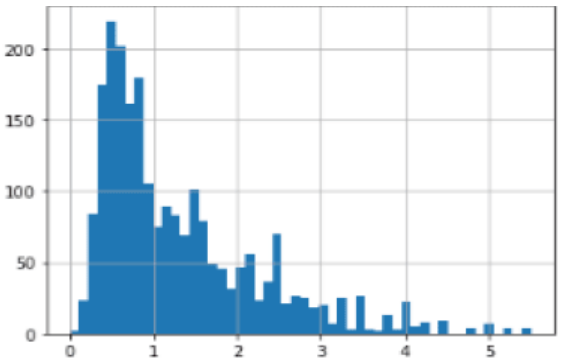
VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.7256

From Figure 4, it can be seen that the number of houses sold that have a freehold certificate (SHM) is the highest, 2234 or 92.9%, and houses with AJB certificates has the lowest number of sales, only 0.4% or 10.



Source: (Research Results, 2025)
Figure 5. Interior Field

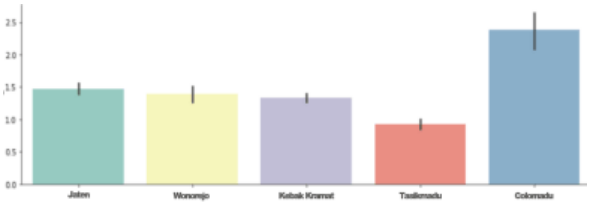
From Figure 5, it is known that the most houses without interiors were sold, with a total of 1814 (75.5%), while the fewest were houses with complete interiors, with a total of 133 (5.5%).



Source: (Research Results, 2025)
Figure 6. Location Field

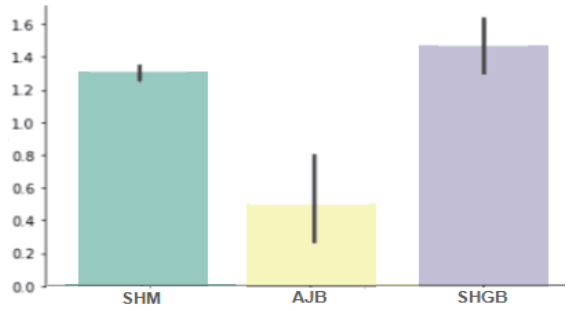
In Figure 6, it can be seen that the graph decreases as total of samples increases, so the increase in house prices is proportional to the decrease in houses or samples total.

Multivariate Analysis is useful for finding out the relationship between 2 or more variables used in this research.



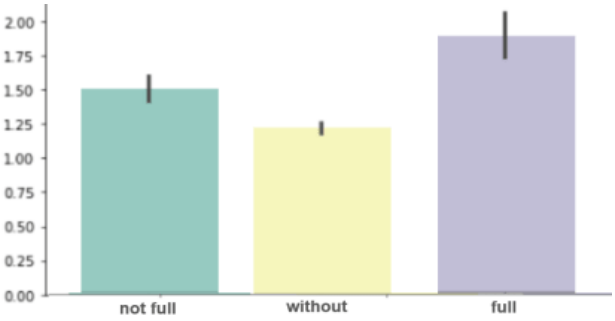
Source: (Research Results, 2025)
Figure 7. Average Price Relative to Location

In Figure 7, it can be seen that the average price in the three locations: Jaten, Wonorejo, and Kebak Kramat, is around 1.3 billion Rupiah. Colomadu is the location with the highest average of 2.4 billion Rupiah.



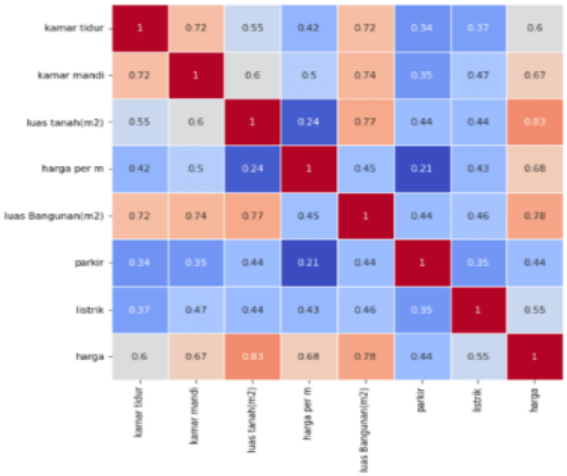
Source: (Research Results, 2025)
Figure 8. Average Price Relative to Certificate Type

The average price for the type of house certificate that has SHGB and SHM is more popular.



Source: (Research Results, 2025)
Figure 9. Average Price Relative to Interior Completeness

The average price for interior equipment is clear which has the highest price complete interior.



Source: (Research Results, 2025)
Figure 10. Numerical Feature Matrix Correlation

From Figure 10, it can be explained that there are two variables that have the most influence on housing prices, namely land area and building area. Garage is the variable that has the least influence or the lowest correlation to price.

Categorical data is converted into numerical data. Categorical data is converted into new variables. There are 5 types of address variables, 3 types of interior and 3 types of certificate types, so based on the value of this variable there are 11 new variables with values 0 and 1. The converted data is presented in Table 3.

Table 3. Dataset After Conversion

Variable	Type	Data Type
total bed room	fitur	Float
total bath room	fitur	Float
land area	fitur	Float
price per meter	fitur	Float
cut_jaten	fitur	Float
cut_wonorejo	fitur	Float
cut_colomadu	fitur	Float
cut_tasikmadu	fitur	Float
cut_kebakkramat	fitur	Float
building area	fitur	Float
cut_shgb	fitur	Float
cut_ajb	fitur	Float
cut_shm	fitur	Float
cut_without interior	fitur	float
cut_ full interior	fitur	Float
cut_ not full interior	fitur	Float
electricity	fitur	Float
price	target	Float

Source: (Research Results, 2025)

At this stage it is divided into three reduction processes, namely to reduce location features, certificate type features and interior completeness features. From each feature reduction there is a data normalization process, calculating variance and covariance values, calculating eigenvectors and eigenvalues, and finally data projection. This article presents one process of location features which is running program result.

```
array([3.20449173e-01, 2.19495270e-01, 1.68047804e-01, 2.22429467e-
       2.32677342e-31])

array([[ 0.19096969, -0.04905121, -0.00466612, -0.09719197, -0.04006039],
       [-0.04905121,  0.15451133, -0.00346741, -0.07222369, -0.02976901],
       [-0.00466612, -0.00346741,  0.01783586, -0.00687047, -0.00283186],
       [-0.09719197, -0.07222369, -0.00687047,  0.23527159, -0.05898547],
       [-0.04006039, -0.02976901, -0.00283186, -0.05898547,  0.13164673]])
```

Source: (Research Results, 2025)

Figure 11. Location Feature Variance and Covariance Values

```
array([[-0.51364018, -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 ]])
```

Source: (Research Results, 2025)

Figure 12. Location Feature Vector Eigen Values

After knowing an eigenvector values, they are sorted in descending order. The largest eigenvalue has the highest information that will be used as a basis for data reduction. An information proportion of location feature is: (0.428, 0.311, 0.24, 0.03, 0). Table 4 presents some of location feature reduction results.

Table 4. Location Feature Reduction Values

id	location	id	location	id	location
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	-0.6406
8	0.701776	17	-0.31396		
9	-0.24715	18	-0.31396		

Source: (Research Results, 2025)

Normalization is a stage to change existing values into values in the range 0-1. Table 5 is the result of normalization of some existing data.

Table 5. Normalization Results

Price	Bed room	Bath room	Land area	Price/meter	Building area	Electricity	Location	Certificate
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	0.13	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
0.263	0.125	0.200	0.227	0.605	0.156	0.539	1	0
0.509	0.500	0.400	0.578	0.460	0.631	0.279	1	0
0.500	0.750	0.800	0.397	0.675	0.578	1	0.281	0

Source: (Research Results, 2025)

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.7256

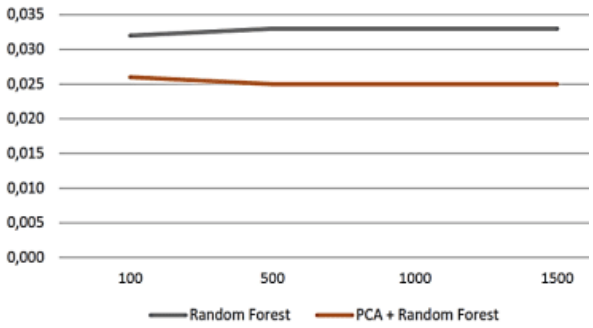
This research results are the price predictions results using the RF method only and price predictions from the results of two methods, namely RF and PCA. From the comparison of the two results, the accuracy of combining the two methods in determining house price predictions can be seen. PCA is used to reduce location and certificate variables.

To test and compare the two methods, the RMSE matrix was used using equation 7. In this study, testing was carried out with four iterations, the results of which are presented in table 6.

Table 6. Comparison Test Results of Two Methods

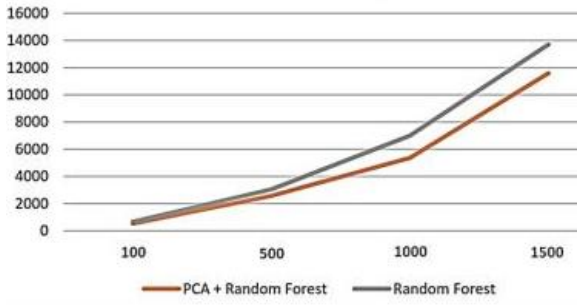
Iteration	PCA and RF		*RF*	
	Error	Time	Error	Time
100	0.027	527	0.032	657
500	0.026	2562	0.033	3047
1000	0.025	5348	0.034	6996
1500	0.025	11586	0.034	13697
Average	0.0257	5005.75	0.0332	6099.25

Source: (Research Results, 2025)



Source: (Research Results, 2025)
Figure 13. RMSE Results Graph

As shown in Figure 13, the RMSE testing indicates that the error value is 0.032 at 100 iterations. It increases slightly to 0.033 at 500 iterations, and then to 0.34 at both 1000 and 1500 iterations.



Source: (Research Results, 2025)
Figure 14. Model Training Time

From Table 5 and Figure 14, it can be explained that the RF method training time in one iteration is: 100 takes 657 milli seconds, 500 takes 3047 milli seconds, 1000 takes 6996 milli seconds, 1500 takes 13697 milli seconds. For PCA and RF in one iteration it is found: 100 takes 527 milli seconds, 500 takes 2562 milli seconds, 1000 takes 5348 milli seconds and 1500 takes 11586 milli seconds. Based on testing using RF and PCA, the error rate is smaller and has a more consistent value with an average of 0.0257. Testing only using RF (without PCA) has a higher error rate with an average of 0.0332. The PCA model produces a faster average training time of 5005.75, for without PCA the average time is 6099.25.

CONCLUSION

Based on the analysis, it is concluded that house sales are highest in the Plesungan area, and homes with freehold title certificates are also the most sold. Based on the ten facilities or variables, the analysis concluded that land area and building area are the most influential factors on the selling price, while electricity is the least significant variable, possessing the lowest correlation value.

The training results indicate that the integration of the random forest and PCA methods produces more optimal outcomes than using the random forest method alone. Testing with the PCA method resulted in a smaller and more consistent error rate, with an average of 0.0257, compared to the random forest method alone, which had a higher error rate, averaging 0.0332. The training time for the PCA model was faster (5005.75) compared to the random forest method alone (6099.25).

For subsequent research, it is necessary to use a dataset from another source or another region in Indonesia with more variables and a larger dataset. The use of other comparison methods such as XGBoost or Gradient Boosting is also needed to further assess performance.

REFERENCE

[1] F. C. K. Analisa and S. Okada, "Tiny house characteristics in Indonesia based on millennial's user preference," *Urban, Plan. Transp. Res.*, vol. 11, no. 1, pp. 1–25, 2023, doi: 10.1080/21650020.2023.2166095.

[2] A. Barlybayev, A. Sankibayev, R. Niyazova, and G. Akimbekova, "Machine learning for real estate valuation: Astana, Kazakhstan case," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 35, no. 2, pp. 1110–1121, 2024, doi: 10.11591/ijeecs.v35.i2.pp1110-1121.

[3] Y. Lu, V. Shi, and C. J. Pettit, "The Impacts of Public Schools on Housing Prices of Residential Properties: A Case Study of Greater Sydney,

Australia," *ISPRS Int. J. Geo-Information*, vol. 12, no. 7, 2023, doi: 10.3390/ijgi12070298.

[4] L. G. Perdamaian and Z. (John) Zhai, "Status of Livability in Indonesian Affordable Housing," *Architecture*, vol. 4, no. 2, pp. 281–302, 2024, doi: 10.3390/architecture4020017.

[5] A. M. Igamo, A. Azwardi, A. Saputra, R. G. Ismail, G. Gustriani, and V. D. Melliny, "Monetary Policy and Demographics: Empirical Evidence for Housing Prices in Indonesia," *Sriwij. Int. J. Dyn. Econ. Bus.*, vol. 6, no. 4, pp. 371–384, 2023, doi: 10.29259/sijdeb.v6i4.371-384.

[6] H. Sharma, H. Harsora, and B. Ogunleye, "An Optimal House Price Prediction Algorithm: XGBoost," *Analytics*, vol. 3, no. 1, pp. 30–45, 2024, doi: 10.3390/analytics3010003.

[7] E. B. Satoto, "Boosting Homeownership Affordability for Low-Income Communities in Indonesia," *Int. J. Sustain. Dev. Plan.*, vol. 18, no. 5, pp. 1365–1376, 2023, doi: 10.18280/ijsdp.180506.

[8] N. Dhaka, A. Chaudhary, D. Sisodia, M. Sharma, and S. Babu, "Prediction of House Pricing Using Machine Learning," *Tuijin Jishu/Journal Propuls. Technol.*, vol. 45, no. 2, pp. 1026–1034, 2024, doi: 10.1109/ICAC3N60023.2023.10541549.

[9] K. Srivastava, S. Verma, M. S. Khan, and A. Singh, "House Price Prediction Using Machine Learning," in *Proceedings - 2021 3rd International Conference on Advances in Computing, Communication Control and Networking, ICAC3N 2021*, 2021, pp. 203–206. doi: 10.1109/ICAC3N53548.2021.9725552.

[10] E. sakti Pramukantoro, K. Amron, V. Wardhani, and P. A. Kamila, "Implementasi Sensor Polar H10 dan Raspberry Pi dalam Pemantauan dan Klasifikasi Detak Jantung Beberapa Individu Secara Simultan dengan Pendekatan *Machine Learning* ," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 1, pp. 175–182, 2024, doi: 10.25126/jtiik.20241117716.

[11] Wiharto and F. N. Mufidah, "Early detection of coronary heart disease based on risk factors using interpretable machine learning," *Int. J. Adv. Appl. Sci.*, vol. 13, no. 4, pp. 944–956, 2024, doi: 10.11591/ijaas.v13.i4.pp944-956.

[12] E. Pitaloka, T. B. A. Hartanto, and S. Sandiwarno, "Penerapan Machine Learning Untuk Prediksi Bencana Banjir," *J. Sist. Inf. Bisnis*, vol. 14, no. 1, pp. 62–76, 2024, doi: 10.21456/vol14iss1pp62-76.

[13] R. Kosasih, Sudaryanto, and A. Fahrurozi, "Classification of six banana ripeness levels based on statistical features on machine learning approach," *Int. J. Adv. Appl. Sci.*, vol. 12, no. 4, pp. 317–326, 2023, doi: 10.11591/ijaas.v12.i4.pp317-326.

[14] Nimatul Mamuriyah, Richard, and Haeruddin, "Implementation Mean Imputation and Outlier Detection for Loan Prediction Using the Random Forest Algorithm," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 10, no. 4, pp. 937–944, 2025, doi: 10.33480/jitk.v10i4.6437.

[15] H. A. Setyadi, Supriyanta, G. S. Nurohim, P. Widodo, and Y. Sutanto, "Knowledge-Based Intelligent System for Diagnosing Three-Wheeled Motorcycle Engine Faults," *Int. J. Informatics Vis.*, vol. 8, no. 4, pp. 2472–2478, 2024, doi: 10.62527/joiv.8.4.2487.

[16] J. N. Sari, P. Madona, H. Kusryanto, M. M. Zain, and M. Valzon, "Electrocardiogram signals classification using random forest method for web-based smart healthcare," *Int. J. Adv. Appl. Sci.*, vol. 12, no. 2, pp. 133–143, 2023, doi: 10.11591/ijaas.v12.i2.pp133-143.

[17] J. M. Alyza, F. S. Utomo, Y. Purwati, B. A. Kusuma, and M. S. Azmi, "Music Recommendation System Based on Cosine Similarity and Supervised Genre Classification," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 9, no. 1, pp. 77–80, 2023, doi: 10.33480/jitk.v9i1.4324.

[18] E. A. Winanto *et al.*, "Peningkatan Performa Deteksi Serangan Menggunakan Metode PCA Dan Random Forest," vol. 11, no. 2, pp. 285–290, 2024, doi: 10.25126/jtiik.20241127678.

[19] J. Chen, F. Gong, S. Xiang, and T. Yu, "Application of principal component analysis in evaluation of epidemic situation policy implementation," *J. Phys. Conf. Ser.*, vol. 1903, no. 1, pp. 1–5, 2021, doi: 10.1088/1742-6596/1903/1/012056.

[20] R. Tanamal, N. Minoque, T. Wiradinata, Y. Soekamto, and T. Ratih, "House Price Prediction Model Using Random Forest in Surabaya City," *TEM J.*, vol. 12, no. 1, pp. 126–132, 2023, doi: 10.18421/TEM121-17.

[21] R. Jáuregui-Velarde, L. Andrade-Arenas, D. H. Celis, R. C. Dávila-Morán, and M. Cabanillas-Carbonell, "Web Application with Machine Learning for House Price Prediction," *Int. J. Interact. Mob. Technol.*, vol. 17, no. 23, pp. 85–104, 2023, doi: 10.3991/IJIM.V17I23.38073.

[22] Z. A. Jasim, Z. Zahid, A. Z. Ul-Saufie, and M. M. Mansor, "Comparison Between Principal Component Analysis and Sparse Principal Component Analysis as Dimensional Reduction Techniques for Random Forest based High Dimensional Data Classification," in *2024 IEEE International Conference on Computing, ICOCO 2024*, IEEE, 2024, pp. 7–11.

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.7256

doi: 10.1109/ICOCO62848.2024.10928248.
[23] O. Ben Ali, S. Hammami, M. Hasni, F. H'Mida, and A. N. S. Moh, "Using Machine Learning To Evaluate Industry 4.0 Maturity: A Comprehensive Analysis Highlighting Lean's Impact On Digital Transformation," *J. Eng. Technol. Ind. Appl.*, vol. 10, no. 5, pp. 156–167, 2024, doi: https://doi.org/10.5935/jetia.v10i50.1262.
[24] S. M. S. Zulkiplee, M. A. M. Shukran, M. R. M. Isa, M. A. Khairuddin, N. Wahab, and H. Hidayat, "Examining the Impact Factors Influencing Higher Education Institution (HEI) Students' Security Behaviours in Cyberspace Environment," *Int. J. Informatics Vis.*, vol. 9, no. 1, pp. 146–152, 2025, doi: 10.62527/joiv.9.1.2296.
[25] H. A. Parhusip, S. Trihandaru, A. H. Heriadi, P. P. Santosa, and M. D. Puspasari, "Data Exploration Using Tableau and Principal Component Analysis," *Int. J. Informatics Vis.*, vol. 6, no. 4, pp. 911–920, 2022, doi: 10.30630/joiv.6.4.952.
[26] P. M. Paithane, "Random Forest Algorithm Use for Crop Recommendation," *J. Eng. Technol. Ind. Appl.*, vol. 9, no. 43, pp. 34–41, 2023, doi: 10.5935/jetia.v9i43.906.
[27] A. B. Wiratman and Wella, "Personalized Learning Models Using Decision Tree and Random Forest Algorithms in Telecommunication Company," *Int. J. Informatics Vis.*, vol. 8, no. 1, pp. 318–325, 2024, doi: 10.62527/joiv.8.1.1905.
[28] M. K. A. Rahman *et al.*, "Hand Gesture Recognition Based on Continuous Wave (CW) Radar Using Principal Component Analysis (PCA) and K-Nearest Neighbor (KNN) Methods," *Int. J. Informatics Vis.*, vol. 6, no. 1–2, pp. 188–194, 2022, doi: 10.30630/joiv.6.1-2.926.