
**JMBI (Journal of Marketing and Business Intelligence)**
JMBI, Volume 03, Number 02, November 2025
EISSN: 2987-0879; PISSN: 2987-0860, pp. 102-111
https://journal.yhmm.or.id/index.php/JMBI

# HOUSE PRICE PREDICTION USING MACHINE LEARNING WITH THE LINEAR REGRESSION ALGORITHM

**Siti Mukodimah[1], Trisnawati[2], Sri Ipnuwati[3]**
Department of Information Systems, Institut Bakti Nusantara
PSDKU Jl. Wisma Rini No.09 Pringsewu, Lampung, Indonesia
E-Mail: mukodimah97@gmail.com, trisnawatiibn@gmail.com, nengachie@gmail.com

**Article history:**
Received: Oct 16, 2025
Revised: Nov 13, 2025
Accepted: Nov 25, 2025

Corresponding authors
*mukodimah97@gmail.com

**Keywords:**
house price prediction;
linear regression;
orange data mining;
pearson correlation.

**Abstract**
Data analysis enables accurate house price prediction by utilizing historical data and machine learning techniques. Determining house prices is essential for sellers, buyers, and property developers; however, it remains challenging due to the influence of various physical and socioeconomic factors. This study aims to predict house prices based on the variables CRIM (crime rate), LSTAT (percentage of the lower socioeconomic population), RAD (road accessibility), and RM (average number of rooms), with MEDV as the target variable. The Linear Regression algorithm was implemented using Orange Data Mining software, while the model performance was evaluated using the coefficient of determination ($R^2$), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The results indicate that the model achieved an $R^2$ value of 0.624, an MAE of 3.904, an RMSE of 5.574, and a MAPE of 0.202. Pearson correlation analysis revealed that RM has a strong positive correlation (+0.791) with house prices, whereas LSTAT has a strong negative correlation (−0.731). Meanwhile, CRIM (−0.389) and RAD (−0.384) exhibit moderate negative correlations with house prices. These findings suggest that both physical factors, such as the number of rooms, and socioeconomic factors, including crime rate and the proportion of the low-income population, significantly influence house prices. This study provides valuable insights for property developers and home buyers in considering these factors when determining property prices or making purchasing decisions. Future research is recommended to incorporate additional variables, more advanced machine learning algorithms, and larger datasets to improve prediction accuracy.

# I. INTRODUCTION

The demand for adequate housing continues to increase in line with population growth, urbanization, and economic development. These conditions have made house prices increasingly dynamic and influenced by various factors, including land area, building size, the number of bedrooms, the number of bathrooms, building age, accessibility, and geographical location. The complex relationships among these factors make the process of determining house prices no longer feasible using experience-based or conventional approaches alone. Therefore, accurate house price prediction has become essential for various stakeholders, including property developers, financial institutions, governments, and prospective homebuyers, as it supports more objective and efficient decision-making. According to data from Real Estate Indonesia (REI), housing demand has continued to increase, with demand reaching 1,700 housing units in 2021 and rising significantly to 3,000 units in 2022. As housing demand continues to grow, accurate house price prediction can provide valuable insights for prospective buyers and other stakeholders, enabling them to make more informed and appropriate decisions.

Previous studies have demonstrated the effectiveness of various machine learning methods in prediction tasks. Study (Rachmat, 2020) estimated pepper commodity prices using the K-Nearest Neighbors (KNN) algorithm combined with the Forward Selection feature selection method. The results showed that Forward Selection outperformed Backward Elimination in identifying the most relevant features for improving prediction performance. Study (Muhammad Makmun Effendi, 2020) applied the Naïve Bayes algorithm using RapidMiner to predict housing loan eligibility. The experimental results achieved an accuracy of 96.23%, indicating excellent predictive performance and demonstrating the suitability of the method for housing credit eligibility prediction. Furthermore, Study (Fitri, 2023) compared several regression algorithms for house price prediction and found that Random Forest Regression achieved the highest prediction accuracy of 81.5%, outperforming both Linear Regression and Gradient Boosted Trees Regression.

This study focuses on identifying patterns and linear relationships between housing characteristics and property prices that can be used to predict future house prices. The primary objective is to develop an accurate and reliable house price prediction model using a machine learning approach based on the Linear Regression algorithm. Recent advances in information technology, particularly in the field of machine learning, have introduced new approaches for solving prediction problems involving large-scale datasets and complex relationships. Machine learning enables systems to learn patterns from historical data and generate predictions for unseen data without being explicitly programmed. In the real estate market, machine learning techniques can identify relationships between housing characteristics and market values, thereby producing more consistent prediction models than conventional manual estimation methods. Consequently, the application of machine learning to house price prediction has become an increasingly important research topic with significant relevance in both academic research and industrial applications.

Among the various machine learning algorithms, Linear Regression is one of the most widely used techniques for continuous value prediction problems. The algorithm models the linear relationship between independent variables, which represent influencing factors, and the dependent variable, which represents the target value to be predicted. Linear Regression offers several advantages, including model simplicity, ease of interpretation, computational efficiency, and the ability to quantify the contribution of individual predictor variables to the prediction results. Furthermore, it is commonly employed as a baseline model in predictive analytics studies because of its robust performance when the underlying relationships among variables are predominantly linear.

Nevertheless, the predictive performance of a Linear Regression model is highly dependent on data quality and the effectiveness of data preprocessing procedures. These procedures include handling missing values, data normalization, feature selection, and appropriate partitioning of training and testing datasets. In addition, model performance should be evaluated using established metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination ($R^2$) to assess the accuracy of house price predictions. Therefore, developing a reliable prediction model requires not only the implementation of an appropriate algorithm but also effective data management practices to ensure strong generalization performance.

Based on these considerations, this study aims to develop a house price prediction model using a machine learning approach with the Linear Regression algorithm. The model is constructed using housing characteristics as predictor variables to estimate property prices more objectively. The findings are expected to contribute to the development of decision support systems in the real estate sector, provide a useful reference for future research on house price prediction, and demonstrate the effectiveness of the Linear Regression algorithm in modeling the relationship between housing characteristics and property values. Accordingly, this study offers both practical value in supporting house price determination and academic value by enriching the body of knowledge on the application of machine learning to real estate price prediction.

Table titles are written in Times New Roman 10-point, placed above the table, without ending the dot. The table should not be decapitated, unless it is not possible to be typed in a single page. On the next page the table lists the table numbers and is written an Advanced word without title. Charts, graphs, maps, photos, all called images. The title of the picture writes in Times New Roman 10 point, just below the image, without ending by the dot. The picture description is written in an empty space on the same page. The scale and units on the graph should be as clear as possible. Each table and picture should be referenced in the paper.

# II. RESEARCH METHODS

## 2.1 Data Collection Method

Data collection is a crucial component of any research study, as it determines how researchers obtain, process, and verify the validity of the information required to address the research objectives. In other words, data collection methods provide a systematic and structured framework that guides researchers in gathering relevant data in accordance with the research design and objectives. In this study, several data collection methods were employed based on their suitability for the research objectives and the characteristics of the study. These methods include:

a. Literarure Riview
The literature review was conducted through an in-depth examination of relevant and credible scientific publications, including the identification of key findings, critical analysis, and synthesis of previous studies (Al-Musaylh et al., 2020). This approach enabled the researchers to establish a solid theoretical foundation while gaining a comprehensive understanding of recent developments in the related research fields. In

this study, the literature review focused on publications concerning data mining, machine learning, and linear regression methods. These references provided both empirical and methodological support for the selection of the proposed approach and the validation of the predictive model employed in this study (Rahman & Islam, 2022)(Binti Ahmad & Hassan, 2020).

b. Documentation

   The data used in this study were obtained from publicly available digital sources, including online data repositories and data analysis software such as Orange Data Mining. The use of digital datasets has become a common practice in data mining and machine learning research because it provides accessible, reliable, and well-documented data for model development and evaluation. This approach demonstrates that the utilization of digital data and analytical software constitutes a relevant and widely accepted practice in predictive modeling research.

## 2.2 Mechine Learning

Machine learning is a branch of computer science and artificial intelligence (AI) that focuses on the development of statistical algorithms and computational models that enable computers to learn from data without being explicitly programmed for every specific task (Novita et al., 2023)(Sembiring & Wulandari, 2024). Instead of relying on predefined instructions, machine learning systems identify patterns and relationships within historical data and use these patterns to make predictions or support decision-making on new data (Sun et al., 2022)(Ali et al., 2023). By processing large volumes of data, machine learning algorithms can improve predictive performance and generate more accurate outcomes for a wide range of applications.

Machine learning is generally categorized into three main paradigms: supervised learning, unsupervised learning, and reinforcement learning (Sharma & Gupta, 2021)(Rahayu et al., 2022). Supervised learning is applied when the training dataset contains labeled data, allowing the model to learn the relationship between input variables and the corresponding target outputs, as commonly found in regression and classification tasks (Mu'tashim et al., 2021)(Haryanto et al., 2023). In contrast, unsupervised learning is designed to discover hidden structures or patterns in unlabeled data, such as through clustering and dimensionality reduction techniques (Putri et al., 2023). Reinforcement learning, meanwhile, involves an intelligent agent that learns optimal actions through continuous interaction with its environment by maximizing cumulative rewards.

In the context of house price prediction, supervised learning particularly the Linear Regression algorithm is widely employed because it effectively models the relationship between housing characteristics and property prices. The primary advantages of machine learning include its ability to process large-scale datasets, identify complex relationships among variables, and improve prediction accuracy as additional data become available and prediction models are further refined.

## 2.3 Linear Regression Algorithm

Linear regression is one of the most widely used statistical techniques for modeling and predicting the relationship between a dependent variable and one or more independent variables (Novianty et al., 2021). The relationship between these variables is represented by a linear equation, enabling researchers to estimate changes in the dependent variable based on variations in the independent variables. As a result, linear regression serves as an effective method for both explanatory analysis and predictive modeling.

In regression analysis, variables are classified into two categories: independent variables (predictors), which represent the influencing factors, and the dependent variable (response), which represents the outcome to be predicted (Purwadi et al., 2019). The objective of linear regression is to determine the best-fitting line that minimizes the difference between the observed and predicted values, thereby capturing the underlying linear relationship between the variables.

Within the field of machine learning, the Linear Regression algorithm is a supervised learning method used to model the relationship between input variables (X) and a target variable (Y). The algorithm estimates the optimal linear function that best represents the observed data and uses this relationship to predict the values of the dependent variable for new observations (18)(19). The general form of the linear regression equation is expressed as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_n X_n \in$$

Where:

Y : Dependent variable

$X_{1,} X_2, .. X_n$ : Independent (predictor) variables

$\beta_0$ : Intercept

$\beta_1, \beta_2, .. \beta_n$ : Slope atau koefisien regresi (Perubahan Y setiap 1 satuan perubahan X)

$\in$ : Error term (residual)

In multiple linear regression, the regression coefficients $\beta_0, \beta_1, \beta_2, .. \beta_n$ yare estimated using the Ordinary Least Squares (OLS) method, which minimizes the sum of squared residual errors between the observed and predicted values (Mujahid et al., 2023). The Ordinary Least Squares (OLS) method estimates the regression coefficients by minimizing the sum of squared residuals between the observed and predicted values, which is expressed as:

$$Minimize \sum_{i=1}^{n} \left(Y_i - (\beta_0 + \beta_1 X_i)\right)^2$$

Multiple Linear Regression is an extension of simple linear regression that models the relationship between one dependent variable ((Y)) and two or more independent variables ($X_1$, $X_2$, …, $X_n$).

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_n X_n \in$

## III. RESULTS

### 3.1. Discussion

This study utilized the housing.tab dataset, which was processed using the Orange Data Mining software. The independent variables included CRIM (crime rate), LSTAT (percentage of the lower socioeconomic population), RAD (accessibility to radial highways), and RM (average number of rooms per dwelling), while MEDV (median value of owner-occupied homes) was used as the target variable. The dataset consisted of 506 observations, which were divided into training and testing sets using an 80:20 ratio, with 80% of the data used for model training and the remaining 20% used for model evaluation.

The analysis was conducted to identify the factors influencing house prices (MEDV) and to evaluate the performance of the proposed prediction model. The discussion focuses on interpreting the relationships among the variables through Pearson correlation analysis and assessing the regression model using the coefficient of determination ($R^2$), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). The study employed the **housing.tab** dataset, which consists of 506 observations. The dataset was divided into training and testing sets using an 80:20 ratio, with 80% of the data used for model training and the remaining 20% used for model testing. The prediction model was then constructed as follows.

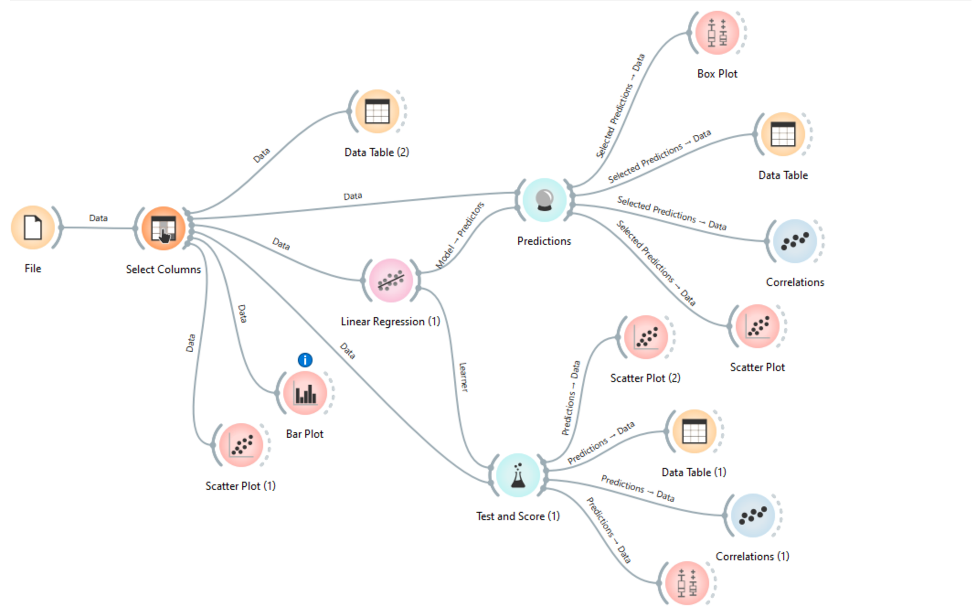


Figure 1. House Price Prediction Model

The prediction model in this study was developed using the Linear Regression algorithm implemented in the Orange Data Mining software to predict house prices (MEDV) based on the selected input variables, namely CRIM, LSTAT, RAD, and RM. The modeling process began by loading the housing.tab dataset into Orange Data Mining. Subsequently, relevant features were explored using visualization tools, including the Box Plot and Scatter Plot widgets, to examine the distribution of the data and the relationships among the variables. After the exploratory data analysis was completed, the selected variables were connected to the Linear Regression widget, which served as the primary prediction model.

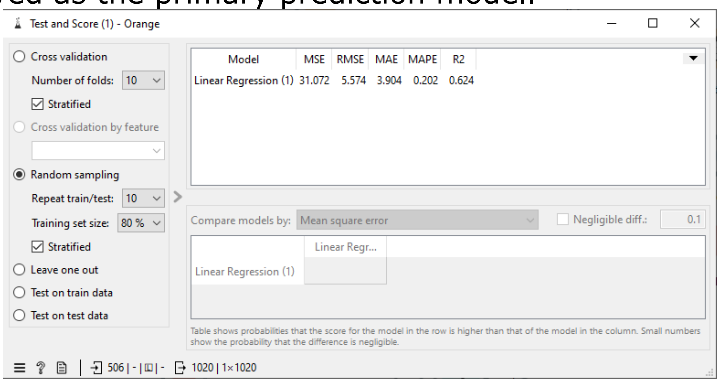


Figure 2. Test and Score Results

As shown in Figure 2, the RM variable (average number of rooms per dwelling) has the strongest positive influence on house prices (MEDV). This finding indicates that houses with a greater number of rooms tend to have higher predicted prices, which is consistent with market trends where larger homes and more complete facilities are generally associated with higher property values. In contrast, LSTAT (percentage of the lower socioeconomic population) exhibits a significant negative relationship with house prices, suggesting that areas with lower socioeconomic conditions tend to have lower property values. The CRIM variable (crime rate) also demonstrates a negative relationship with house prices, although its effect is weaker than that of LSTAT. This result suggests that neighborhood safety remains an important consideration for homebuyers but is not the primary determinant of property value. Meanwhile, RAD (accessibility to radial highways) shows a relatively weaker and more variable influence on house prices, which may reflect regional differences in the perceived value of transportation accessibility.

The regression model achieved an $R^2$ value of 0.624, indicating that approximately 62.4% of the variation in house prices can be explained by the selected predictor variables. The remaining 37.6% is likely attributable to other factors not included in the model, such as construction quality, proximity to commercial centers, neighborhood characteristics, and the availability of public facilities. Furthermore, the relatively low MAE and RMSE values indicate that the prediction errors remain within an acceptable range. The MAPE value of 0.202 (20.2%) suggests that the model's average prediction error is approximately 20.2%, indicating a satisfactory level of predictive accuracy for house price estimation.

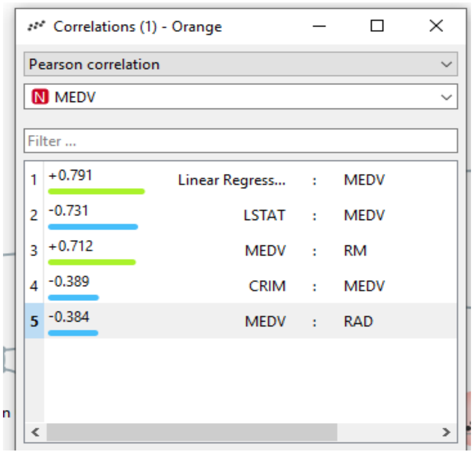


Figure 3. Pearson Correlation Matrix

The Pearson correlation analysis revealed varying degrees of association between the independent variables and house prices (MEDV). The RM variable (average number of rooms per dwelling) exhibited a strong positive correlation ($r = +0.791$), indicating that houses with a greater number of rooms tend to have higher market values. This finding is consistent with trends in the real estate market, where larger homes with more complete facilities are generally associated with higher property prices.

In contrast, LSTAT (percentage of the lower socioeconomic population) showed a strong negative correlation ($r = -0.731$), suggesting that neighborhoods with lower socioeconomic conditions tend to have lower house prices. The CRIM variable (crime rate) demonstrated a moderate negative correlation ($r = -0.389$), indicating that although neighborhood safety influences property values, its effect is less substantial than that of socioeconomic conditions or housing characteristics. Similarly, RAD (accessibility to radial highways) exhibited a moderate negative correlation ($r = -0.384$), suggesting that proximity to major highways does not necessarily increase property values and may, in some cases, reduce residential desirability because of factors such as traffic noise and environmental pollution. Overall, these findings indicate that the physical characteristics of a house, particularly the number of rooms, together with the socioeconomic conditions of the surrounding neighborhood, have a stronger association with house prices than crime rate and transportation accessibility. These results highlight the importance of considering both housing characteristics and neighborhood conditions when developing house price prediction models.

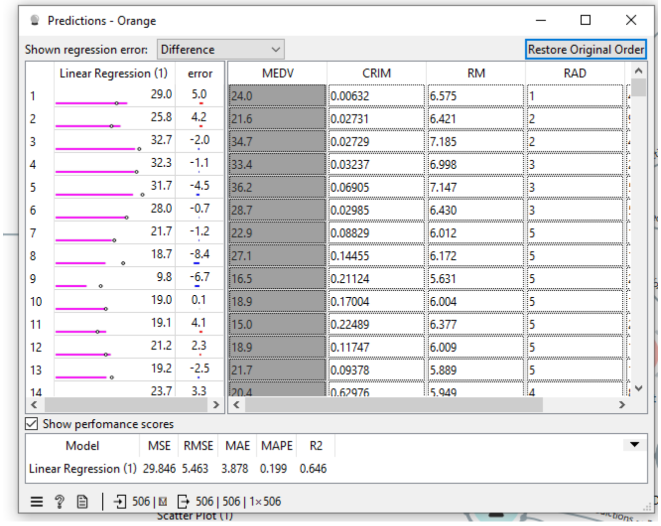


Figure 4. Prediction Results

As shown in Figure 4, the Linear Regression model evaluated on the house price prediction dataset achieved a Mean Squared Error (MSE) of 29.846, indicating that the average squared difference between the actual and predicted values was relatively low. This result suggests that the model was able to generate predictions that closely approximated the observed house prices. The model obtained a Root Mean Squared Error (RMSE) of 5.463, indicating that the average prediction error was approximately 5.46 house price units. In addition, the Mean Absolute Error (MAE) was 3.878, demonstrating that the average absolute prediction error remained relatively small, which reflects good predictive performance. The Mean Absolute Percentage Error (MAPE) of 0.199 (19.9%) indicates that the average relative prediction error was below 20%, suggesting that the model achieved an acceptable level of accuracy for house price prediction. Furthermore, the model achieved a coefficient of determination ($R^2$) of 0.646, indicating that approximately 64.6% of the variation in house prices could be explained by the independent variables included in the model. Overall, the combination of these evaluation metrics demonstrates that the proposed Linear Regression model provides satisfactory predictive performance. Nevertheless, there remains potential for further improvement through enhanced feature engineering, additional data preprocessing, or the application of more advanced machine learning algorithms.

### 3.2. Results Analysis

The results analysis indicates that the Linear Regression model developed using the CRIM, LSTAT, RAD, and RM variables was able to predict house prices (MEDV) with satisfactory accuracy. The model achieved an $R^2$ value of 0.624, indicating that approximately 62.4% of the variation in house prices could be explained by the selected predictor variables, while the remaining variation was attributable to factors not included in the model. Furthermore, the Mean Absolute Error (MAE) of 3.904 and the Root Mean Squared Error (RMSE) of 5.574 indicate that the average difference between the predicted and actual house prices ranged from approximately 3 to 6 thousand USD, which is relatively small compared with the average house prices in the dataset. The Pearson correlation analysis revealed that RM exhibited a strong positive correlation with MEDV (r = +0.791), indicating that houses with a greater average number of rooms tend to have higher market values. Conversely, LSTAT showed a strong negative correlation (r = –0.731), suggesting that neighborhoods with a higher proportion of lower socioeconomic populations generally have lower house prices. The CRIM (r = –0.389) and RAD (r = –0.384) variables demonstrated moderate negative correlations, indicating that crime rate and highway accessibility are associated with house prices, although their effects are less pronounced than those of housing characteristics and neighborhood socioeconomic conditions. Overall, the findings demonstrate that the physical characteristics of a house, particularly the average number of rooms, together with the socioeconomic characteristics of the surrounding neighborhood, play a dominant role in determining house prices. In contrast, crime rate and transportation accessibility provide additional explanatory value but are not the primary determinants of property prices. These findings may serve as a useful reference for property developers, investors, and prospective homebuyers when evaluating property values by considering both physical housing characteristics and neighborhood socioeconomic conditions.

## IV. CONCLUSION

The findings of this study demonstrate that the Linear Regression model developed using the CRIM, LSTAT, RAD, and RM variables is capable of predicting house prices (MEDV) with satisfactory accuracy. The model achieved an $R^2$ value of 0.624, a Mean Absolute Error (MAE) of 3.904, a Root Mean Squared Error (RMSE) of 5.574, and a Mean Absolute Percentage Error (MAPE) of 0.202, indicating that the proposed model provides reliable predictive performance for house price estimation. The analysis further revealed that RM had the strongest positive relationship with house prices, whereas LSTAT exhibited the strongest negative relationship. The CRIM and RAD variables also showed negative associations with house prices, although their correlations were of moderate strength. These findings suggest that house prices are primarily influenced by the physical characteristics of the property, particularly the average number of rooms, as well as the socioeconomic conditions of the surrounding neighborhood. In contrast, crime rate and transportation accessibility contribute to price variation but are less influential than the primary determinants. Overall, this study highlights the effectiveness of the Linear Regression algorithm as a baseline machine learning approach for house price prediction. The proposed model may serve as a useful reference for property developers, investors, financial institutions, and prospective homebuyers by supporting more objective and data-driven property valuation. Future research is recommended to incorporate additional predictor variables, larger and more diverse datasets, and advanced machine learning algorithms to further improve prediction accuracy and model generalizability.

## REFERENCES


Al-Musaylh, M., Deo, R. C., Adamowski, J., & Li, Y. (2020). Machine learning algorithms for rainfall prediction: A review. *Neural Computing and Applications*, *32*, 16129–16148.

Ali, R., Khan, M., & Ahmed, S. (2023). Housing price prediction using hybrid machine learning models. *Journal of Big Data*, *10*(1), 45. https://doi.org/10.1186/s40537-023-00785-4

Binti Ahmad, N., & Hassan, Z. (2020). Predictive modeling using KNN and regression techniques: A performance comparison. *Procedia Computer Science*, *178*, 381–388.

Fitri, E. (2023). Analisis Perbandingan Metode Regresi Linier, Random Forest Regression dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah. *Journal of Applied Computer Science and Technology*, *4*(1), 58–64. https://doi.org/10.52158/jacost.v4i1.491

Haryanto, C., Rahaningsih, N., & Basysyar, F. M. (2023). Komparasi Algoritma Machine Learning dalam Memprediksi Harga Rumah. *JATI: Jurnal Mahasiswa Teknik Informatika*. https://doi.org/10.36040/jati.v7i1.6343

Mu'tashim, M. L., Muhayat, T., Damayanti, S. A., Zaki, H. N., & Wirawan, R. (2021). Analisis Prediksi Harga Rumah Sesuai Spesifikasi Menggunakan Multiple Linear Regression. *Informatik : Jurnal Ilmu Komputer*. https://doi.org/10.52958/iftk.v17i3.3635

Muhammad Makmun Effendi, E. H. (2020). Penerapan Data Mining Menggunakan Metode Algoritma Naive Bayes Untuk Menentukan Kelayakan Kredit Rumah Bersubsidi. *Jurnal Teknologi Pelita Bangsa Vol.*, *11*(2), 101–108.

Mujahid, P. E., Sembiring, J. Y., & Sembiring, A. P. (2023). Penerapan Data Mining untuk Memprediksi Inflasi Listrik dan Bahan Bakar Rumah Tangga Menggunakan Metode Regresi Linear. *Jurnal Tekinkom*, *7*(2), 145–152.

Novianty, D., Palasara, N. D., & Qomaruddin, M. (2021). Algoritma Regresi Linear pada Prediksi Permohonan Paten yang Terdaftar di Indonesia. *Jurnal Sistem Dan Teknologi Informasi (Justin)*, *9*(2), 81. https://doi.org/10.26418/justin.v9i2.43664

Novita, I. H., Rohana, T., Nurlaelasari, E., & Elvira Awal, E. (2023). Pemodelan Prediksi Ekspor Kopi Indonesia Berbasis Algoritma Machine Learning. *Jurnal Ilmu Manajemen Dan Sistem Informasi (JUMIN)*, *5*(3), 112–120.

Purwadi, P., Ramadhan, P. S., & Safitri, N. (2019). Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang. *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika Dan Komputer)*, *18*(1), 55. https://doi.org/10.53513/jis.v18i1.104

Putri, I. Z., Rosyada, M., Widyadhana, S. E., Virda, S. A., & Arifin, M. (2023). Perbandingan Model Regresi Linier dan Random Forest Regressor dalam Estimasi Harga Jual Rumah Berdasarkan Data Properti di Yogyakarta. *JRIIN : Jurnal Riset Informatika Dan Inovasi*.

Rachmat. (2020). Prediksi Kelayakan Kredit Kepemilikan Rumah Menggunakan Metode K-NN Berbasis Forward Selection. *Jessi*, *1*(2), 68–73. https://ojs.unm.ac.id/JESSI/index

Rahayu, M. I., Dewi, N. H. R., & Gunawan, H. (2022). Penerapan Pembelajaran Mesin Menggunakan Algoritma Regresi Linier Untuk Prediksi Harga Rumah. *Jurnal Penelitian Dan Pengembangan Teknologi Informasi Dan Komunikasi*. https://doi.org/10.58761/jurtikstmikbandung.v13i1.812

Rahman, T., & Islam, M. (2022). Commodity price forecasting using regression-based machine learning models. *Journal of Big Data*, *9*, 1–17.

Sembiring, M. A., & Wulandari, F. (2024). Analisa Kinerja Model Regresi dalam Machine Learning untuk Memprediksi Harga Beras. *Journal of Information System, Informatics, and Computing (JOISIE)*, *5*(1), 22–31.

Sharma, A., & Gupta, R. (2021). Real estate price prediction using supervised learning models. *Procedia Computer Science*, *192*, 4519–4528. https://doi.org/10.1016/j.procs.2021.09.227

Sun, Y., Wang, J., & Li, H. (2022). Comparison of machine learning models for house price prediction. *Expert Systems with Applications*, *198*, 116804. https://doi.org/10.1016/j.eswa.2022.116804