Journal Collabits
Vol. 3 No. 1 January 2026 : 26-37
E-ISSN : 3046-6709, P-ISSN : 3062-8601
https://publikasi.mercubuana.ac.id/index.php/collabits

Laptop Price Prediction Based on Specifications: A Comparison of Random Forest and Linear Regression

Bagas Pratama Putra[1*], Ilham Miftahali Mahfuzh[2], Agus Fahrizal Kurniawan[3], Ramdani Budiman[4]

[1,2,3] Informatics Engineering, Universitas Mercu Buana. Indonesia
[4] Informatics Engineering, Universitas Banten Jaya, Indonesia

*Coressponden Author: bagasputra997@gmail.com

Abstract - This study investigates the prediction of laptop prices based on hardware specifications by comparing the performance of Linear Regression and Random Forest algorithms. The dataset consists of both numerical and categorical features, including brand, processor type, RAM capacity, storage configuration, screen size, and other relevant attributes that influence pricing. Data preprocessing was conducted through data cleaning, handling missing values, and transforming categorical variables using one-hot encoding. The dataset was then divided into training and testing sets with a 70:30 ratio to evaluate model generalization. Exploratory data analysis was performed using visualizations such as average price per brand, correlation heatmaps of numerical features, and scatter plots comparing actual and predicted prices. Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2) on both training and testing data. The results indicate that the Random Forest model achieves higher predictive accuracy compared to Linear Regression, as it is more effective in capturing non-linear relationships and complex feature interactions. In contrast, Linear Regression tends to underperform due to its linear assumptions when applied to heterogeneous laptop specification data. These findings suggest that ensemble-based models are more suitable for laptop price prediction tasks involving diverse and non-linear feature patterns.

Keywords :
Laptop Price Prediction; Machine Learning; Random Forest Regression; Linear Regression; Exploratory Data Analysis;

Article History:
Received: 02-12-2025
Revised: 23-12-2025
Accepted: 10-01-2025

Article DOI : 10.22441/collabits.v3i1.37603

INTRODUCTION

Laptop pricing is shaped by a mix of technical specifications and market positioning. Components such as processor class, RAM capacity, storage type, screen size, graphics, and build characteristics often interact in ways that do not translate into a simple, linear increase in price. In addition, brand segmentation and product tiers can introduce price gaps between devices with similar specs. This complexity makes it difficult to estimate a “reasonable” price using manual comparison alone, especially when buyers and sellers must evaluate many options quickly.

To address this problem, this study builds a regression-based machine learning pipeline to predict laptop prices from specification data and compares Linear Regression with a Random Forest Regressor. The workflow includes data cleaning, missing-value handling, and one-hot encoding for categorical attributes, followed by a 70:30 train–test split to assess generalization. Exploratory visual analysis is also provided through average price by brand, a correlation heatmap for numerical features, and actual-versus-predicted scatter plots to examine prediction behavior. Model performance is evaluated using MAE, RMSE, and R^2 on both training and testing sets, allowing the comparison to highlight underfitting in linear models and potential overfitting in more flexible ensemble methods.

LITERATURE REVIEW

Laptop price prediction is commonly framed as a supervised regression problem where device specifications serve as predictors and the listing price becomes the target variable. Prior studies emphasize that laptop pricing is influenced by a mixture of numeric features (e.g., RAM size, storage capacity, screen size, CPU speed) and categorical attributes (e.g., brand, processor family, GPU type). Because the data often combine these heterogeneous variables, most workflows begin with data cleaning, handling missing values, and transforming categorical fields into numerical representations, typically through one-hot encoding. Exploratory data analysis (EDA) is also frequently reported as a necessary step to understand market patterns, such as brand-based price differences and the relationships among numeric attributes.

A number of studies adopt linear models particularly (multiple) linear regression as baseline methods due to their interpretability and low computational cost. In laptop pricing contexts, linear regression can capture broad trends and is useful for identifying which features contribute positively or negatively to price under linear assumptions. However, several papers note that linear regression performance may degrade when the mapping between specifications and price is non-linear or when interactions between components are important (for example, CPU–RAM–storage combinations). Additionally, with many categorical variables expanded through one-hot encoding, linear regression can become unstable or underfit if the true relationships are more complex than linear combinations of features.

To address these limitations, many recent works employ ensemble learning, especially Random Forest regression, because it naturally models non-linearities and feature interactions without requiring explicit functional forms. In the papers you provided, Random Forest is frequently presented as a strong candidate for laptop price prediction and is evaluated using standard error metrics such as MAE and RMSE, as well as goodness-of-fit measures such as R^2. Some studies further strengthen Random Forest performance using hyperparameter optimization techniques (for example, randomized search strategies) and report improvements in predictive accuracy compared to simpler baselines. These results align with the broader machine learning literature where tree-based ensembles are known to perform well on structured/tabular data with mixed feature types.

At the same time, the literature also highlights an important caveat: exceptionally high test performance can be misleading when the dataset contains information leakage or near-identifier attributes. For example, columns resembling IDs or fields with nearly unique values can allow models—especially flexible ensembles—to “memorize” patterns rather than learn generalizable relationships. Several studies therefore recommend careful feature inspection, removal of ID-like columns, and reporting metrics on both training and testing sets to detect overfitting. Visualization tools such as actual-vs-predicted scatter plots and residual distributions are commonly used to diagnose whether a model’s predictions generalize or merely fit the training data too closely.

Overall, the reviewed works suggest a consistent pattern: linear regression remains valuable as a transparent baseline, but it may struggle when laptop pricing depends on complex interactions and non-linear effects. Random Forest and other ensemble approaches tend to deliver lower prediction errors on heterogeneous specification datasets, provided that preprocessing is applied carefully and evaluation is conducted rigorously. These insights motivate comparative studies that combine EDA, robust preprocessing, and train–test evaluation to determine which modeling approach is most appropriate for laptop price prediction under realistic conditions.

METHODOLOGY

This study utilized Googlee Colaboratory as research platform to conduct a comparative analysis between the Linear Regrssion and Random Forest algorithms. The research process began with the collection of datasets from the kaggle repository. Once the data were gathered, a data cleaning process was performed to improve data quality and ensure that they dataset was ready for analysis.

3.1 Data Acquisition

The dataset used in this study consists of 1,303 laptop records with 22 attributes, as summarized in Figure 1. Each record represents a laptop characterized by a combination of technical specifications and categorical descriptors. The target variable is price, stored as a continuous numerical value, while an additional transformed variable (log_price) is included in the dataset but excluded from modeling to prevent information leakage. No missing values are observed across all attributes, indicating that the dataset is complete and suitable for direct analysis without imputation at the dataset level.

The features can be broadly divided into numerical and categorical groups. Numerical attributes include hardware-related specifications such as RAM size (ram), screen resolution (x_res, y_res), pixel density (ppi), storage capacities (ssd_gb, hdd_gb, total_storage), and touchscreen and IPS panel indicators, which are encoded as binary variables. Several attributes—namely weight, cpu_speed_ghz, and inches—are originally stored as object types due to formatting inconsistencies and are subsequently converted into numerical representations during preprocessing. Categorical features describe qualitative characteristics such as brand (company), laptop type (typename), operating system (opsys), CPU family and brand, as well as GPU brand and class, capturing both manufacturer-related and functional distinctions.

Additionally, the dataset contains an identifier column (laptop_id), which uniquely distinguishes each record and is therefore removed prior to modeling. Overall, the dataset provides a balanced mixture of quantitative and qualitative features, making it well-suited for evaluating both linear and non-linear regression models in the context of laptop price prediction.

Figure 1. Overview of variables in the laptop dataset

--- Column Overview ---

	column	dtype	missing_count	missing_%	unique_count	sample_values
0	laptop_id	int64	0	0.0	1303	1, 2, 3
1	price	float64	0	0.0	791	713786832.0, 478955232.0, 30636.0
2	log_price	int64	0	0.0	791	11175754549129500, 1077677731765330, 103299310...
3	weight	object	0	0.0	160	1.37, 1.34, 1.86
4	cpu_speed_ghz	object	0	0.0	25	2026-03-02 00:00:00, 2026-08-01 00:00:00, 2026...
5	company	object	0	0.0	19	apple, hp, acer
6	inches	object	0	0.0	18	2026-03-13 00:00:00, 2026-06-15 00:00:00, 2026...
7	total_storage	float64	0	0.0	18	128.0, 0.0, 256.0
8	ppi	float64	0	0.0	15	193.52, 108.85, 141.21
9	x_res	int64	0	0.0	13	2560, 1440, 1920
10	cpu_family	object	0	0.0	12	i5, i7, amd a series
11	ssd_gb	float64	0	0.0	10	128.0, 0.0, 256.0
12	y_res	int64	0	0.0	10	1600, 900, 1080
13	ram	int64	0	0.0	9	8, 16, 4
14	opsys	object	0	0.0	9	macos, no os, windows 10
15	typename	object	0	0.0	6	ultrabook, notebook, netbook
16	gpu_brand	object	0	0.0	4	intel, amd, nvidia
17	cpu_brand	object	0	0.0	3	intel, amd, other
18	hdd_gb	float64	0	0.0	3	0.0, 1024.0, 2048.0
19	gpu_class	object	0	0.0	2	integrated, dedicated
20	touchscreen	int64	0	0.0	2	0, 1
21	ips_panel	int64	0	0.0	2	1, 0

3.2 Data Cleaning and Feature Preparation

Data preparation was carried out using the original raw dataset to identify and correct issues that could affect regression performance. Records with missing or invalid target values were removed, and only observations with non-negative prices were retained. During inspection, several numerical attributes—particularly inches, cpu_speed_ghz, and weight—were found to be stored in inconsistent formats (e.g., object strings or date-like values) due to spreadsheet import behavior. These variables were systematically converted into valid numerical representations to ensure that the feature values reflect their intended magnitudes and can be processed by the learning algorithms.

To prevent biased evaluation caused by information leakage, variables that do not represent genuine predictive signals were excluded from the modeling feature set. Specifically, laptop_id (a unique identifier) and log_price (a direct transformation of the target) were removed prior to training. An interquartile range (IQR)–based filtering strategy was applied to the price variable to reduce the influence of extreme outliers while preserving the majority of observations. After cleaning, features were prepared through a unified preprocessing pipeline to ensure consistent transformations on both training and testing data. Numerical features were imputed using the median and standardized to improve stability for linear regression, while categorical attributes were imputed using the most frequent category and converted to numerical form via one-hot encoding with unknown-category handling. This combined cleaning and feature preparation process produces a reliable input matrix for model training and supports a fair comparison between linear and ensemble regression methods.

3.3 Exploratory Data Analysis Procedure

Exploratory Data Analysis (EDA) was conducted as a preliminary step to understand the characteristics of the laptop dataset and to guide subsequent modeling decisions. The main objective of EDA in this study is to identify basic patterns in the target variable (price), assess potential relationships among input features, and detect structural issues (such as extreme values or unusual distributions) that may influence model performance. The EDA process is designed to remain descriptive and does not involve model training; all model-related outputs are presented later in the Results and Discussion section.

Three complementary analyses were performed. First, a univariate inspection of the target variable was carried out by computing summary statistics and visualizing the distribution of price to evaluate its spread and the presence of potential outliers. Second, a brand-level analysis was applied by aggregating prices by company and visualizing the mean price per brand. This step provides an overview of systematic price differences that may be driven by brand positioning and market segmentation. Third, a correlation analysis was performed on numerical features (e.g., RAM, storage capacities, resolution variables, and pixel density) using a Pearson correlation matrix and a heatmap visualization. This correlation inspection helps reveal linear associations between numeric predictors and identifies variables that may carry overlapping information. The figures generated from these EDA procedures are reported and interpreted in Section 4 to support the discussion of pricing behavior and model results.

3.4 Data Spliting

After data cleaning and feature preparation, the dataset was divided into training and testing subsets to evaluate the generalization ability of the proposed models. A hold-out validation strategy was applied, where 70% of the data were allocated to the training set and the remaining 30% to the testing set. The training set was used exclusively for model fitting, while the testing set was reserved for performance evaluation on unseen data.

The split was performed randomly with a fixed random seed to ensure reproducibility of the experimental results. This data splitting strategy allows the study to assess whether the learned relationships between laptop specifications and prices generalize beyond the training data. By reporting evaluation metrics on both training and testing sets, the approach also facilitates the detection of underfitting and overfitting behaviors in linear and ensemble-based regression models

3.5 Model Development

This study develops and compares two supervised regression models to predict laptop prices from the cleaned specification dataset: Linear Regression and Random Forest Regressor. Both models were trained using the same training set and identical preprocessing steps to ensure a fair comparison. Categorical variables (e.g., company, CPU and GPU descriptors, opsys, and typename) were transformed using one-hot encoding, while numerical variables (e.g., ram, storage capacities, screen and resolution attributes, and ppi) were standardized after median-based imputation. These transformations were implemented within a single pipeline so that the same feature mappings were consistently applied to both the training and testing data.
Linear Regression was selected as a baseline model due to its simplicity and interpretability. It estimates laptop price as a linear combination of the encoded input features and provides a reference point for evaluating whether more flexible models offer meaningful improvements. In contrast, Random Forest Regressor was used to capture non-linear relationships and feature interactions that are common in laptop pricing (for example, the combined effect of processor family, RAM, and storage type). To reduce the risk of overly optimistic results, leakage-prone variables such as laptop_id and log_price were excluded from the predictors, and the Random Forest configuration was constrained using regularization-related hyperparameters (e.g., limiting tree depth and enforcing minimum samples per split/leaf) to mitigate overfitting. After training, both models were used to generate predictions for the training and testing sets, which were then assessed using the evaluation metrics described in the following section.

RESULTS AND DISCUSSION

Table 1. Training and Testing Data distribution

Data Type	Number of Records	Prencentage
Training Data	723	70%
Testing Data	311	30%
Total	**1,034**	**100%**

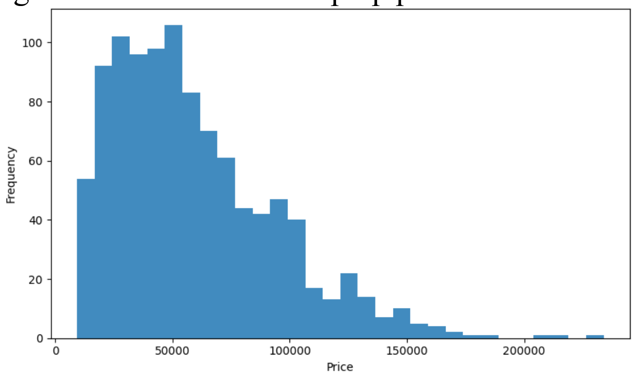
Following the data cleaning and preprocessing stages, the number of observations used in this study was reduced from 1,303 to 1,034 laptop records. This reduction resulted from the application of an interquartile range (IQR)–based filtering on the price variable, which was intended to remove extreme outliers and produce a more representative and stable dataset for regression modeling.

The refined dataset was subsequently divided into training and testing subsets using a 70:30 ratio. A total of 723 records were assigned to the training set for model development, while 311 records were reserved for testing purposes. This data partitioning strategy enables the models to be trained on sufficient data while allowing an objective evaluation of their generalization performance on unseen samples.

4.1Exploratory Data Analysis and Visualization

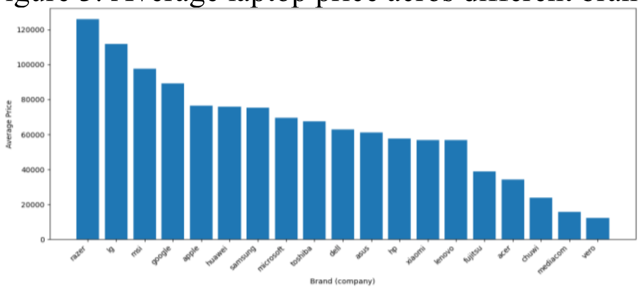
The exploration data analysis was conducted to examine the underlying characteristics of laptop prices and their relationships with key attributes prior to model interpretation. Figure 1 presents the distribution of laptop prices in the dataset. The price distribution shows a right-skewed pattern, indicating that most laptops are concentrated in the lower to mid-price range, while a smaller number of devices fall into higher price categories. This distribution suggests the presence of variability and potential outliers in pricing, which supports the use of preprocessing steps such as outlier handling and motivates the consideration of non-linear regression models for prediction.

Figure 2. Distribution of laptop prices in the dataset



To further explore the influence of categorical attributes, brand-level analysis was performed by calculating the average price for each manufacturer. As shown in Figure 2, noticeable differences exist in the mean prices across laptop brands. Certain brands consistently exhibit higher average prices, reflecting market positioning and perceived product value, while others are concentrated in more affordable segments. This observation highlights the importance of brand-related features in the predictive modeling process and justifies the inclusion of categorical variables such as company in the regression models.

Figure 3. Average laptop price acros different brand



In addition to categorical analysis, relationships among numerical features were examined using a correlation heatmap, as illustrated in Figure 3. The heatmap reveals moderate correlations between several hardware-related attributes, including storage capacity, screen resolution, and pixel density. However, no single numerical feature shows a dominant linear correlation with price. This finding suggests that laptop pricing is influenced by combinations of specifications rather than individual attributes alone, indicating that linear models may be limited in capturing the full complexity of pricing behavior.

Figure 4. Correlation Heatmap

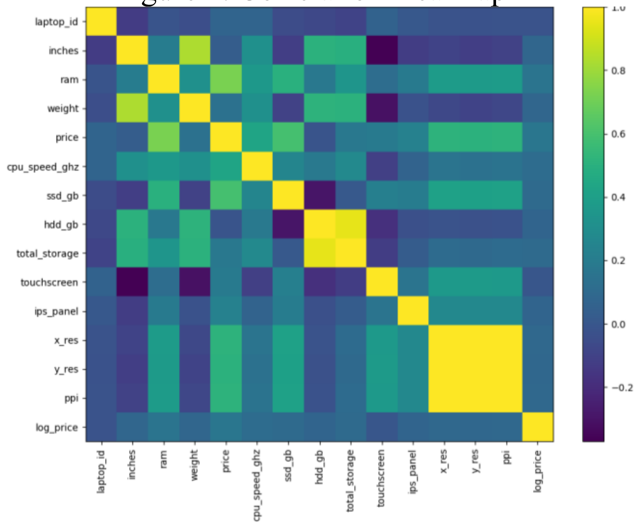


Figure 5. Scatter Plot Linear Regression Train

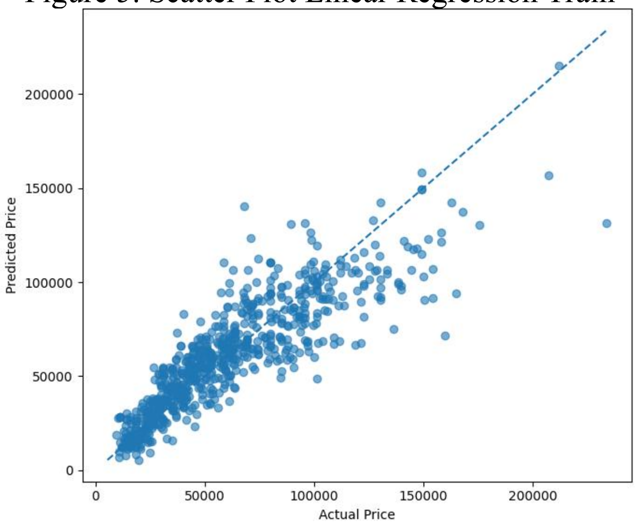
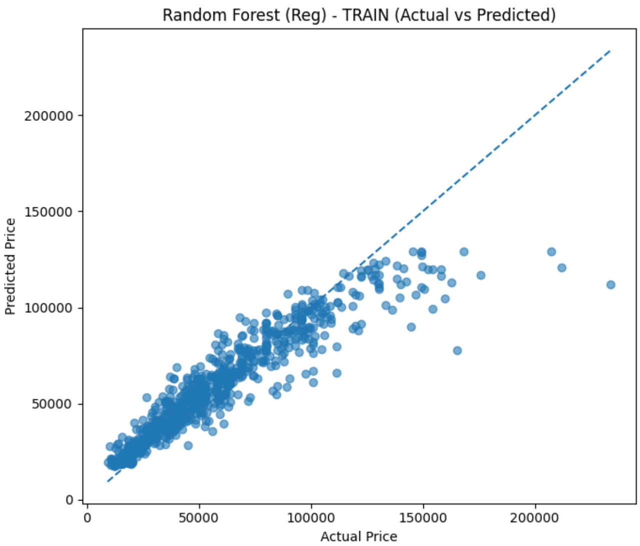


Figure 6. Scatter Plot Random Forest Train



4.2 Model Performance Results

Model performance was evaluated using MAE, RMSE, and R^2 on both the training and testing sets to assess predictive accuracy and generalization. Reporting metrics on both splits is important because it reveals whether a model learns meaningful patterns (good training performance) while maintaining consistent performance on unseen data (good testing performance). Table 2 summarizes the results for Linear Regression and Random Forest under the same preprocessing pipeline and the same 70:30 train–test split.

Table 2. Performance comparison of Linear Regression and Random Forest on training and testing sets (MAE, RMSE, R^2)

Model	MAE	RMSE	R2
Linear Regersion	11,711	16,579	0.7421
Random forest test	9,682.7087		

Based on Table 2, the Linear Regression model serves as a baseline and provides a reference for how well a linear relationship between laptop specifications and price can explain the target variation. The difference between training and testing metrics indicates the degree of generalization; when the gap is small, the model tends to be stable but may still underfit if both scores are relatively low. In contrast, Random Forest typically achieves stronger predictive performance because it can capture non-linear relationships and interactions between features (for example, how CPU family and RAM jointly affect pricing). However, unusually high scores—especially on the training set—may indicate that the model is overly complex or that the dataset contains features that allow memorization. For this reason, the Random Forest configuration in this study is constrained through regularization-related hyperparameters (e.g., limiting tree depth and enforcing minimum sample sizes per node) to reduce overfitting and produce more realistic test performance.

Overall, the comparative results demonstrate the trade-off between interpretability and flexibility. Linear Regression remains easy to interpret and computationally efficient, but its predictive capacity is limited when pricing behavior is influenced by non-linear patterns. Random Forest provides improved modeling capacity for heterogeneous specification data, and the train–test comparison confirms whether this improvement reflects genuine generalization rather than overfitting. These findings motivate a closer inspection of prediction behavior and error patterns, which are discussed in the following section.

4.3 Prediction Behavior and Error Analysis

To further examine how each model behaves beyond summary metrics, prediction patterns were analyzed using scatter plots of actual versus predicted prices and residual-based diagnostics. Figure 4 visualizes the relationship between true prices and Linear Regression predictions. The points show a wider dispersion around the diagonal reference line, indicating larger deviations between predicted and actual values. This pattern suggests that the linear model struggles to represent complex pricing structures, particularly when the relationship between specifications and price is not well approximated by a single linear function.

Figure 7. Scatter Plot Linear Regression Test

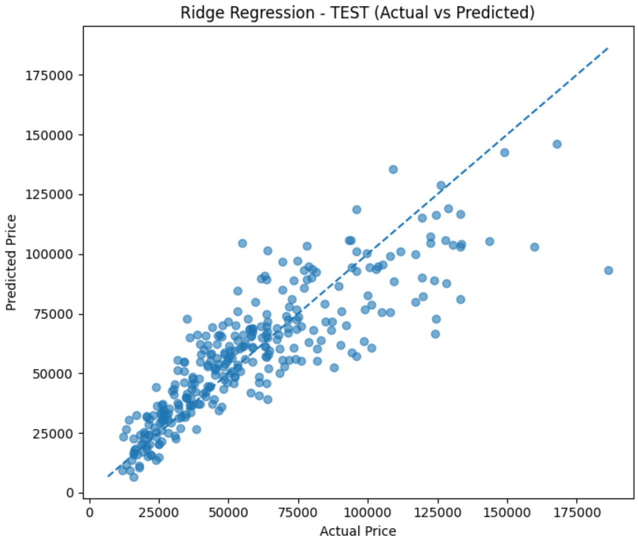
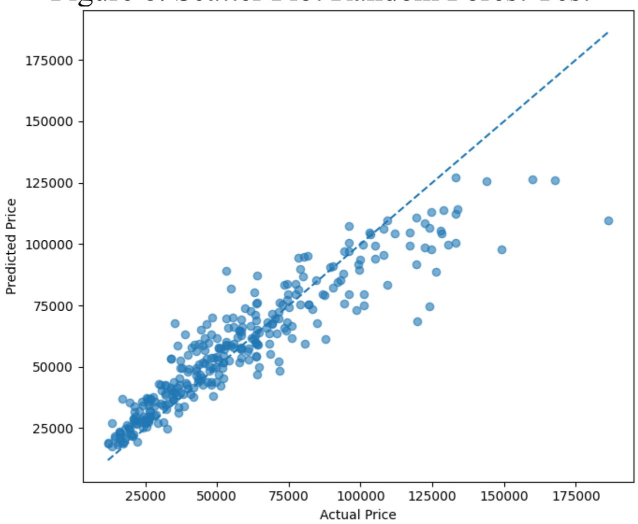


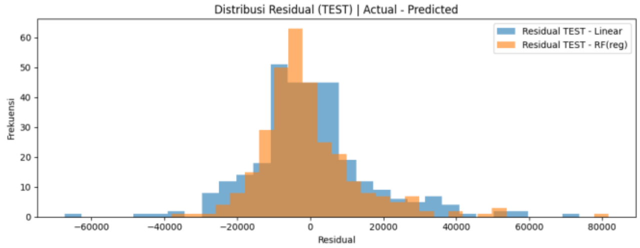
Figure 5 presents the same analysis for the Random Forest model. Compared with Linear Regression, the predictions are generally closer to the diagonal line, reflecting improved capability in capturing non-linear effects and interactions among features. However, noticeable deviations remain for higher-priced laptops, where prediction errors tend to increase. This behavior is consistent with the pricing characteristics of premium devices, which are often influenced by brand prestige and specific high-end configurations that may be less frequent in the dataset.

Figure 8. Scatter Plot Random Forest Test



Residual distributions were then examined to assess the magnitude and direction of prediction errors. Figure 6 compares residual histograms for both models on the evaluation set, where residuals are computed as **Actual − Predicted**. The Linear Regression residuals exhibit a broader spread, indicating larger variability in errors, while the Random Forest residuals are more concentrated around zero, suggesting more consistent predictions. Despite this improvement, both models still show increased residual variance in the upper price range, implying that extreme values remain more difficult to predict accurately. Overall, the scatter and residual analyses support the metric-based comparison by demonstrating that Random Forest provides better fit and stability, whereas Linear Regression is limited by its linear assumptions.

Figure 9. Distribusi Residual



4.4 Discussion Summary and Key Findings.

Overall, the results from exploratory analysis and model evaluation consistently indicate that laptop pricing is driven by a combination of brand effects and interacting hardware specifications. The EDA results (Section 4.1) show clear brand-based price segmentation and non-trivial relationships among numerical features, suggesting that the mapping from specifications to price is not strictly linear. This observation is reflected in the predictive behavior (Section 4.3), where the Linear Regression model tends to produce broader deviations from the ideal prediction line, while the Random Forest model generally yields predictions that are closer to the actual values.

From the model comparison (Section 4.2), the test-set metrics provide the most reliable indication of real-world predictive capability. The Random Forest model is able to capture non-linear relationships and feature interactions, resulting in more stable prediction patterns and a tighter residual distribution. However, both models still exhibit larger errors for higher-priced laptops, implying that premium segments remain challenging due to factors such as brand premium, limited representation of high-end configurations, and additional market-driven pricing effects not fully explained by specifications alone. In summary, the findings suggest that non-linear ensemble models are more suitable for this dataset, while linear models can still serve as interpretable baselines but may underrepresent complex pricing behavior.

CONCLUSION

This study evaluated machine learning approaches for predicting laptop prices using a structured dataset of laptop specifications. A complete pipeline was implemented, including data cleaning, type correction for inconsistent numeric fields, leakage prevention by removing identifier-like variables, and feature preparation using one-hot encoding for categorical attributes and standardization for numerical variables. The dataset was split using a 70:30 train–test strategy to assess generalization performance.

The exploratory analysis indicated that laptop pricing is influenced by both brand segmentation and combinations of hardware characteristics rather than a single dominant numerical feature. In model evaluation, Linear Regression provided an interpretable baseline but showed limitations when pricing relationships were non-linear. Random Forest demonstrated more stable predictive behavior and improved capability to capture complex interactions among features, although prediction errors increased for higher-priced laptops in both models. Overall, the findings suggest that ensemble-based regression models are more suitable for laptop price prediction on heterogeneous specification data, while linear models remain useful for baseline comparison and interpretability.

For future work, performance may be improved by expanding the dataset to better represent premium devices, applying systematic hyperparameter optimization, and incorporating additional market-related variables (e.g., release year, warranty status, and regional pricing factors). Further evaluation using k-fold cross-validation and feature importance analysis could also strengthen model robustness and interpretability.

REFERENCES

[1] Siburian, A. D., Sitompul, D. R. H., Sinurat, S. H., Situmorang, A., Ruben, Ziegel, D. J., & Indra, E. (2022). Laptop price prediction with machine learning using regression algorithm. JUSIKOM PRIMA (Jurnal Sistem Informasi dan Ilmu Komputer Prima), 6(1), 87–91.

[2] Sari, N. N., Anisah, T. T., & Fitriani, R. (2024). Implementasi machine learning untuk prediksi harga laptop menggunakan algoritma regresi linear berganda. Jurnal Manajemen Informatika (JAMIKA), 14(2), 162-177 .

[3] Fahri, M. (2025). Penerapan algoritma random forest untuk deteksi phishing pada website. *JITSI: Jurnal Ilmiah Teknologi Sistem Informasi*, *6*(2), 186–194.

[4] Wardani, S. N., & Nurmalitasari. (2023). Penerapan regresi linier berganda untuk memprediksi harga laptop dengan menggunakan software python. Prosiding Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB), 377–382.

[5] Darma, I. W. A. S., Wirawan, I. M. A., & Darmawiguna, I. G. M. (2023). Comparative analysis of machine learning algorithms for laptop price prediction. Jurnal Pendidikan Teknologi dan Kejuruan, 20(1), 47–56.

[6] Anggoro, R., Al Hakim, R. R., Saptiningtyas, R. P. S., Pradanaputra, S. A. G. G., Maulana, H., Muslih, A., & Solihin, M. A. (2022). Laptop price prediction using random forest regressor and linear regression algorithms. JST (Jurnal Sains dan Teknologi), 11(2), 221–229.

[7] Darma, I. W. A. S., Pradnyana, I. M. A., & Darmawiguna, I. G. M. A. (2024). Penerapan algoritma random forest untuk prediksi harga laptop. JITSI: Jurnal Ilmiah Teknologi Sistem Informasi, 5(2), 70–76.

[8] Amnur, A., Hadi, R., Mooduto, H. A., Idmayanti, R., Syaljumairi, R., & Asri, E. (2025). Prediksi harga laptop berdasarkan spesifikasi menggunakan algoritma k-nearest neighbour. JITSI: Jurnal Ilmiah Teknologi Sistem Informasi, 6(3), 267–273.

[9] Fahri, M. (2025). Penerapan algoritma random forest untuk deteksi phishing pada website. JITSI: Jurnal Ilmiah Teknologi Sistem Informasi, 6(2), 186–194.

[10] Harahap, R., Karpen, Yenni, H., & Jamaris, M. (2025). Implementasi algoritma regresi linear untuk memprediksi harga laptop. Jurnal Ilmiah BETRIK, 16(02), 157–171.

[11] Septiyanah, S., & Athalina, G. (2025). Implementasi algoritma random forest regression dalam prediksi harga laptop. Jurnal MNEMONIC, 8(1), 70–75.

[12] Tian, P. (2024). Research on laptop price predictive model based on linear regression, random forest and xgboost. Highlights in Science, Engineering and Technology, 85, 265–271.

[13] Wardani, S. N., & Nurmalitasari. (2023). Penerapan regresi linier berganda untuk memprediksi harga laptop dengan menggunakan software python. Prosiding Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB), 377–382

[14] Hadi, N., & Benedict, J. (2024). Implementasi machine learning untuk prediksi harga rumah menggunakan algoritma random forest. Computatio: Journal of Computer Science and Information Systems, 8(1), 50–61.

[15] Hidayat, R., Saputra, H. T., Husnah, M., Nabila, Hidayatullah, M. B., Nazhmi, M. N., Azra, J., & Rana, A. (2025). Implementasi algoritma random forest regression untuk memprediksi

penjualan produk di supermarket. Jurnal Sistem Informasi dan Sistem Komputer, 10(1), 101–110.

[16] Ilmi, I. N., Mustafid, & Suparti. (2025). Penerapan model regresi random forest untuk prediksi harga laptop berdasarkan fitur laptop. Jurnal Gaussian, 14(2), 269–279

[17] Utami, P., Jundi, M., Rahmaddeni, & Sinaga, L. (2025). Property price prediction using the random forest regression algorithm. SITEKIN: Jurnal Sains, Teknologi dan Industri, 22(2), 208–215.

[18] Amnur, A., Hadi, R., Mooduto, H. A., Idmayanti, R., Syaljumairi, R., & Asri, E. (2025). Prediksi harga laptop berdasarkan spesifikasi menggunakan algoritma k-nearest neighbour. JITSI: Jurnal Ilmiah Teknologi Sistem Informasi, 6(3), 267–273.

[19] Asnur, H., Yunita, R., & Marsa, A. R. (2025). Algoritma random forest untuk prediksi perkerasan tanah. Jurnal Pustaka AI, 5(2), 381–387

[20] Gulo, S. H. T., Rumapea, H., & Simanullang, H. G. (2025). Penerapan data mining dalam pemilihan laptop mahasiswa menggunakan metode random forest (Studi kasus: Universitas Methodist Indonesia). Methosisfo: Jurnal Ilmiah Sistem Informasi, 5(1), 55–64.

[21] Nugroho, M. W. (2025). Analisis performa algoritma random forest dalam mengatasi overfitting pada model prediksi. Program Studi Teknik Informatika, Universitas Kristen Satya Wacana.

[22] Srinivas, K. S., Sai, K. T., Sriram, K. M. S., & Rao, P. S. V. (2024). Laptop price prediction in machine learning using random forest classifier technique. International Journal of Innovative Science and Research Technology, 9(2), 1251–1256.

[23] Wijaya, S., & Fauziah. (2023). Analysis of the comparison between linear regression, random forest, and logistic regression methods in predicting crude palm oil (CPO) price. Brilliance: Research of Artificial Intelligence, 3(2), 243–251.

[24] Adrianty, S., & Maspiyanti, F. (2024). Laptop price prediction using extreme gradient boosting algorithm. Journal of Applied and Research Computer Science and Information Systems, 2(1), 132–148.

[25] Bindu, K. S., & Ashwini, C. (2024). Laptop price prediction using machine learning. International Journal of Advances in Engineering and Management (IJAEM), 6(07), 492–495.