International Journal of Informatics and Information Systems
Vol. 7, No. 4, December 2024, pp. 189-199

A Comparative Analysis of Linear Regression and XGBoost Algorithms for Predicting GPU Prices Using Technical Specifications

Dendi Putra Prakoso[1,*], Muhammad Irfan[2], Quba Siddique[3]

[1]*Program of Information System, Faculty of Computer Science,Universitas Amikom Purwokerto, Indonesia*

[2]*Institute of Banking and Finance, Bahauddin Zakariya University Bosan Road Multan, Multan, Pakistan*

(Received March 10, 2024; Revised June 18, 2024; Accepted September 30, 2024; Available online December 1, 2024)

Abstract

This study investigates and compares the predictive performance of Linear Regression and XGBoost algorithms in estimating Graphics Processing Unit (GPU) prices based on their technical specifications. GPU prices are known for their high volatility, influenced not only by hardware characteristics—such as memory capacity, clock speed, and bandwidth—but also by external market factors including demand from the gaming industry, machine learning applications, and cryptocurrency mining activities. The dataset used in this research comprises 475 GPU units from three leading manufacturers—NVIDIA, AMD, and Intel Arc—featuring 15 technical attributes obtained from publicly accessible data sources. Adopting an experimental quantitative approach, the dataset was divided into training and testing subsets using an 80:20 ratio. The data preprocessing phase involved handling missing values, detecting outliers through the Interquartile Range (IQR) method, performing data normalization, and encoding categorical features. The models were evaluated using four performance metrics: the Coefficient of Determination (R^2), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The results demonstrate that XGBoost outperforms Linear Regression, achieving an R^2 of 0.8129, MAE of 85.07 USD, RMSE of 122.03 USD, and MAPE of 35.23%. In comparison, the Linear Regression model recorded an R^2 of 0.7629, MAE of 106.59 USD, RMSE of 137.38 USD, and MAPE of 56.04%. The superior performance of XGBoost can be attributed to its ability to model non-linear relationships and capture complex feature interactions among GPU specifications.

Keywords: GPU, XGBoost, Linear Regression, Price Prediction, Machine Learning, Technical Specifications

1. Introduction

The rapid development of technology has increased the need for hardware or hardware that supports high performance with qualified specifications, one of which is the Graphics Processing Unit (GPU). GPUs are currently not only used in gaming needs[1], but also used for heavy work such as graphic rendering[2], machine learning[3], and crypto mining which a few years ago was widely discussed[4]. Thus, GPUs have experienced a significant increase in demand, so the price of GPUs has also fluctuated quite sharply. As demand for GPUs has increased, both from general consumers and industry, GPU prices have experienced significant volatility in recent years. This situation is reinforced by several factors such as the surge in interest in crypto coin mining which causes GPU stocks on the market to drop dramatically[5], economic disruption during the COVID-19 pandemic[6], and the global scarcity of semiconductor chips[7]. As a result, GPU prices no longer follow a stable pattern and are difficult to predict based solely on release year and product class.

On the other hand, GPUs have various technical specifications that directly reflect their capabilities and performance[8]. Some of the main specifications that are often used as a reference are the number of processing cores, for example in NVIDIA, the CUDA cores and in AMD, the Stream Processor, besides that there is also the capacity of video memory (VRAM), base core clock and boost core clock, bus width, and memory bandwidth. The general assumption about GPU prices is that the higher the GPU specification, the higher the GPU price [9]. However, this assumption has not been fully validated quantitatively, especially with many non-technical factors that can affect the price. The complexity of the relationship between technical specifications and GPU price requires an appropriate analytical approach to uncover hidden patterns. This relationship is not always linear, as there are interactions between

*Corresponding author: Dendi Putra Prakoso (dendipp624@gmail.com)
DOI: https://doi.org/10.47738/ijiis.v7i4.228

specifications that can affect the overall value of the product. In addition, various external factors such as market conditions, stock availability, and brand positioning also contribute to the final price.

To address these challenges, this study uses two different machine learning approaches to predict GPU prices based on specifications. First, linear regression is used as the baseline model to measure the linear relationship between GPU specifications and its selling price. This method was chosen due to its ease of interpretation and effectiveness in identifying linear patterns between quantitative variables[10]. Secondly, XGBoost (Extreme Gradient Boosting) is used as an ensemble model that is able to capture complex non-linear relationships between specifications and price, and can better handle interactions between features[11]. XGBoost was chosen because of its several advantages, including the ability to handle non-linear relationships and complex feature interactions, robustness to outliers, ability to provide clear feature importance, and high prediction performance[12]. Meanwhile, linear regression provides a baseline that is easy to understand and can be used as a comparison to measure how much accuracy improvement is obtained from more complex approaches. By comparing the two approaches, this research hopes to provide an understanding of the best method for predicting GPU prices based on specifications, while identifying the specification factors that have the most influence on pricing. The comparison will be made based on several evaluation metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2-score to measure the prediction accuracy of both models. The results of this analysis are expected to be a reference for consumers in making optimal purchasing decisions, helping industry players in pricing strategies, and contributing to the field of technology device price prediction.

2. Literature Review

2.1. Research Related to GPU Price Prediction

GPU performance and price are closely related to technical specifications such as the number of cores, VRAM capacity, and clock speed. Research [12] examined the performance of GPUs as hardware for price forecasting, although their focus was on stock prices. They compared the performance of LSTM model training on TPU and GPU hardware using stock price datasets such as S&P 500, FTSE100, and HKEX. The main findings show that GPUs provide shorter computation time with comparable accuracy on small-sized datasets, while TPUs can achieve lower error (RMSE) on very large datasets. These results indicate that GPUs with high computational specifications have excellent computational efficiency in small data scenarios. Although not the main focus, this study highlights how the technical specifications of GPUs (e.g. number of parallel cores and high frequency) affect model performance in price forecasting [12].

Research [13] specifically studied GPU price changes since the COVID-19 pandemic. They used historical daily price data of NVIDIA RTX 3090 Founder Edition GPUs to predict prices 8, 16, and 30 days ahead with a Transformer-based deep learning model. The comparison results showed that the Transformer model provided superior performance: at the 30-day horizon, the Transformer achieved a correlation coefficient of 0.8743, RMSE of 34.68, and MAPE of 0.82, better than RNN or LSTM. The study concluded that the Transformer model was effective and efficient in predicting GPU prices, demonstrating that advanced machine learning approaches can capture GPU price trends with high accuracy [13].

Research [14] predicted the price of NVIDIA graphics cards by considering the influence of cryptocurrency prices using the Support Vector Regression (SVR) algorithm. This research applies the Knowledge Discovery in Data (KDD) process and optimizes SVR parameters with a grid search algorithm. The dataset used includes six GPU models (GTX 1050, 1050 Ti, 1060, RTX 3070, 3080, and 3090) along with the prices of two cryptocurrencies. The experimental results show high prediction accuracy across all models, especially for the RTX 30 GPU series: for example, the RMSE for RTX 3070 is only 0.03178, which is much lower than the RMSE for GTX 1060 (0.07629) and GTX 1050 (0.2028). This study concludes that SVR with RBF kernel is able to model the relationship between crypto prices and GPU prices effectively. This finding is relevant as it confirms that external variables such as cryptocurrency prices can affect GPU price fluctuations, and that regression methods are effective for forecasting GPU prices [14].

2.2. Related Research on Linear Regression in Price Prediction

Linear regression has been widely used as a baseline model in electronic device price prediction due to its ease of interpretation and ability to identify linear patterns between quantitative variables. Research [15] predicted the price of used electronic devices, especially smartphones, using various machine learning algorithms. They collected used smartphone price data for the past five years using web scraping techniques, and then modeled the relationship between the device's features and its selling price. Their experiments involved linear regression, multilayer perceptron (MLP), and Random Forest. The results showed that the Random Forest model provided the lowest prediction error compared to linear regression and MLP, confirming the limitations of linear regression in capturing the non-linear complexity of broad-spectrum data[15].

Although linear regression is often used as a point of departure due to its ease of interpretation of coefficients, [16] showed that this model is only able to capture simple linear associations between variables such as property area and sales price, so when faced with data with complex feature interactions, its accuracy decreases significantly. In addition, assumption violations such as multicollinearity between technical features or heteroscedasticity can result in biased coefficient estimates, triggering systematic over or under prediction[16] .

Tian [17] extended linear regression testing on a laptop dataset that included 13 technical attributes, and compared it with Random Forest and XGBoost; the experimental results showed linear regression provided a reasonable baseline but still fell short in terms of RMSE and R^2, emphasizing its limitations in modeling non-linear patterns and overlapping technical feature interactions. Thus, linear regression remains valuable as a diagnostic and benchmark tool, but needs to be supplemented or replaced with more adaptive techniques when dealing with modern data complexity[17].

2.3. Related Research on XGBoost in Price Prediction

XGBoost as an ensemble learning method has shown significant advantages in various price prediction tasks. Researchs conducted a comparative analysis of several machine learning models to predict laptop prices, using a dataset containing 1,303 laptop data with 11 attributes (including brand, processor, RAM, GPU, and other technical specifications). This study evaluated three regression methods: Linear Regression, Histogram Gradient Boosting, and XGBoost. The results showed that XGBoost significantly outperformed, with the highest coefficient of determination R^2 (0.93559 on training data) and significantly lower RMSE error (9,334.9 on training data) than Linear Regression. Correlation analysis in the study identified RAM and processor specification variables as the most dominant factors determining laptop price, confirming that ensembles such as XGBoost are very effective in modeling complex relationships between features [18].

Adrianty and Maspiyanti [19] tested XGBoost on two large datasets from Indonesian laptop e-commerce Bhinneka and Pemmz finding an R^2 of 0.98 with low RMSE in both data sources; this achievement shows that XGBoost effectively maps non-linear interactions between features such as RAM, GPU, and storage capacity, while maintaining prediction stability despite wide-scale price distributions after careful preprocessing[19] .

Gautam [20] utilized a hybrid LSTM+XGBoost architecture to predict crypto prices, where the XGBoost component dampens noise and corrects residual bias from the LSTM, bringing the MAPE down to sub-1% industry level; this reinforces XGBoost's flexibility in handling highly volatile time-series data influenced by many external factors such as market sentiment. Overall, the literature shows that XGBoost not only excels on traditional tabular data, but also as a vital component in hybrid models for complex forecasting[20].

2.4. Advantages and Limitation of Methods

Based on the literature reviewed, there are clear characteristic differences between linear regression and ensemble methods such as XGBoost. Linear Regression offers several advantages, including ease of interpretation, fast training time, minimal data requirements, and the ability to provide clear insights into the effect of each variable. However, it is limited to modeling linear relationships, is sensitive to outliers, and performs poorly when dealing with complex, non-linear patterns. In contrast, XGBoost is capable of capturing non-linear relationships, is robust to outliers, can model interactions between features effectively, and generally achieves higher prediction accuracy. Despite these advantages, XGBoost has some limitations, such as being more complex to interpret, requiring careful parameter tuning, and having a tendency to overfit when applied to small datasets.

 ISSN 2579-7069

2.5. Research Gap

Although numerous studies have demonstrated the effectiveness of machine learning methods in predicting electronic device prices, several important gaps remain. First, most existing research focuses on general devices such as smartphones or laptops, with limited attention given to GPUs, which have distinct market and technical characteristics. Second, few studies have conducted a systematic comparison between linear regression and XGBoost specifically for GPU price prediction based on technical specifications. Third, prior research often limits the analysis to a single vendor, such as NVIDIA, thereby lacking a multi-vendor perspective that includes Intel, AMD, and NVIDIA. Lastly, the trade-off between model interpretability (as seen in linear regression) and prediction accuracy (as achieved by XGBoost) has not been explicitly examined in GPU pricing studies. This research aims to address these gaps by performing a systematic comparison between linear regression and XGBoost for GPU price prediction using a dataset that encompasses products from three major vendors—Intel, AMD, and NVIDIA—thus providing more comprehensive insights into the most effective method for GPU price estimation.

3. Method

This research uses a quantitative approach with an experimental design to compare the performance of linear regression and XGBoost algorithms in predicting GPU prices based on technical specifications. The research framework includes the stages of data collection, preprocessing, modeling, and comparison evaluation.

3.1. Dataset and Data Source

The research dataset is obtained from a public data collection available on Google Sheets shared by the tech community. The dataset includes technical specification and price information of GPUs from three major vendors namely Intel Arc, AMD Radeon, and NVIDIA GeForce with a period up to 2025. The total dataset consists of approximately 475 GPU units with 15 technical specification attributes that include GPU Clockspeed,Memory Clockspeed,Memory Size (MB),Buswidth (bits),Max Pixel Fillrate (MP/s),Max Texel Fillrate (MT/s),Max Shader Performance (MFLOPS),Max Bandwidth (MB/s),Memory Type,DirectX Gen,Process Size,Transistors,Ext. Power, Price, Year of Release, Vendor.

The independent variables in this study include GPU technical specifications such as GPU Clockspeed, Memory Clockspeed, Memory Size (MB), Buswidth (bits), Max Pixel Fillrate (MP/s), Max Texel Fillrate (MT/s), Max Shader Performance (MFLOPS), Max Bandwidth (MB/s), Memory Type, DirectX Gen, Process Size, Transistors, Ext. Power, Year of Release, Vendor. While the dependent variable is the GPU Price in USD units. The selection of these variables is based on the literature which shows that these technical specifications have a significant correlation with GPU performance and price [8][9].

3.2. Data Preprocessing

The preprocessing stage is carried out to ensure data quality before modeling. This process begins with the identification and handling of missing values using median imputation techniques for numerical variables and mode for categorical variables. Duplicate data was identified based on a combination of GPU name and key specifications, then removed to avoid bias in the model. Outlier detection was performed using the Interquartile Range (IQR) method with thresholds Q_1 - 1.5×IQR and Q_3 + 1.5×IQR. Extreme outliers that could undermine model performance were removed from the dataset, while moderate outliers were retained as they reflect the natural variation in GPU price data. To address scale differences between variables, normalization was performed using StandardScaler for the XGBoost model and MinMaxScaler for linear regression. Categorical variables such as GPU brand were transformed using One-Hot Encoding to preserve non-ordinal information, while variables with natural hierarchy such as memory type used Label Encoding. Several additional features were created to improve the predictive power of the model, including performance per watt ratio and memory efficiency.

3.3. Dataset Division and Validation

The dataset was divided into three parts using stratified random sampling to ensure a balanced representation of each brand and price category. The division was done with a proportion of 80% for training set, 20% for validation set, and 20% for test set. Stratification was done based on brand distribution and price range to avoid bias in model evaluation.

3.4. Modeling

A multiple linear regression model was implemented as the baseline model using the Ordinary Least Squares (OLS) method. This model assumes a linear relationship between GPU technical specifications and price, with the equation:

$$\text{Price} = \beta 0 + \beta 1 \cdot \text{Spec1} + \beta 2 \cdot \text{Spec2} + \ldots + \varepsilon. \tag{1}$$

where $X_1, X_2, \ldots, X_n$ are GPU technical specifications, β_0 is the intercept, $\beta_1, \beta_2, \ldots, \beta_n$ are regression coefficients, and ε is the error term. Before modeling, linear regression assumptions such as linearity, residual normality, homoscedasticity, and error independence were tested to ensure the validity of the model.

XGBoost was chosen as the ensemble model due to its ability to capture non-linear relationships and complex interactions between features. The model was implemented with an initial configuration that used the objective function 'reg:squarederror' for the regression problem. Hyperparameter tuning is performed using Grid Search with 5-fold cross-validation to optimize parameters such as max_depth, learning_rate, n_estimators, subsample, and colsample_bytree. The hyperparameter optimization process aims to find the combination of parameters that gives the best performance on the validation set while avoiding overfitting. Feature importance is analyzed using gain-based importance from XGBoost to identify the technical specifications that have the most influence on price prediction.

3.5. Model Evaluation

The performance of both models was evaluated using four key metrics. Mean Absolute Error (MAE) is used to measure the average prediction error in dollar terms, while Root Mean Squared Error (RMSE) penalizes larger errors. The Mean Absolute Percentage Error (MAPE) shows the relative percentage error, and the coefficient of determination (R^2) measures the proportion of price variance explained by the model.

$$\text{coefficient of determination } R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{2}$$

$$\text{Mean Absolute Error (MAE) } MAE = \frac{n}{1}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3}$$

$$\text{Root Mean Squad Error (RMSE) } RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4}$$

A 5-fold cross validation was performed to test the stability and generalizability of the model. Residual analysis was performed to verify model assumptions, including plots of residuals versus predicted values and Q-Q plots to test residual normality. To provide deeper insight, evaluation was also conducted on different data segments based on GPU brand and price category.

3.6. Comparative Analysis

The performance comparison of the two models was done quantitatively using paired t-test to test the significance of differences in evaluation metrics. Effect size was calculated using Cohen's d to measure the magnitude of practical differences between the two models. In addition to the quantitative aspect, a qualitative comparison was also conducted by considering the model's interpretability, computational complexity, and ease of implementation. The feature importance analysis of the two models was compared to identify consistency in the determination of technical specifications that have the most influence on GPU price. The results of this analysis provide practical insights for stakeholders in understanding the factors that determine GPU prices in the market. All experiments were conducted using Python 3.8 with scikit-learn library for linear regression, XGBoost for gradient boosting, and pandas and numpy for data manipulation. Reproducibility was guaranteed by using a consistent random seed at each stage of the experiment.

4. Results and Discussion

4.1. Dataset description

The final dataset used in this study consists of 475 GPU units from three major vendors: NVIDIA, AMD, and Intel Arc. The price distribution ranges from $23 to $1,999 with a median of $249.5, indicating a right-skewed distribution typical of technology products where high-end models command premium prices while entry-level options remain accessible to budget-conscious consumers.

Table 1 shows the descriptive statistics of the main variables in the dataset. The distribution of GPU prices shows considerable variation, reflecting a diverse market segmentation ranging from entry-level GPUs to high-end GPUs designed for professional rendering, machine learning applications, and cryptocurrency mining. This data provides a comprehensive representation of the current GPU market conditions, with the three major vendors contributing significantly to the dataset. The wide price range creates its own challenges in the prediction process, as each price segment has different characteristics and patterns that traditional linear models may struggle to capture effectively.

Table 1. Descriptive Statistic of Main Variable

Variable	Mean	Min	Max	Median
Price (USD)	348,5	23	1999	249,5
Memory Size (MB)	3240,58	80	32768	10240
GPU Clockspeed (MHz)	1023	40	3130	900
Memory Clockspeed (MHz)	1040	100	3750	900
Buswidth (bits)	280	64	8192	192

The dataset encompasses GPUs released over multiple years, capturing the evolution of GPU technology and corresponding price adjustments. This temporal dimension adds complexity to the prediction task, as older high-end GPUs may have similar prices to newer mid-range models despite having different performance characteristics. The dataset includes various memory types (GDDR5, GDDR6, HBM), process node sizes (ranging from 28nm to 7nm), and different architectural approaches, providing a rich feature space for machine learning algorithms.

Preprocessing identified and handled 81 missing values using median imputation for numerical variables and mode imputation for categorical variables, ensuring that the dataset maintained its statistical properties while addressing gaps in the data. Outlier detection was performed using the Interquartile Range (IQR) method with thresholds Q_1 - 1.5×IQR and Q_3 + 1.5×IQR, which identified 15 extreme outliers that were carefully examined before removal. These outliers included prototype GPUs, limited edition models, and GPUs with pricing anomalies that could distort the learning process, resulting in a final dataset of 385 samples.Descriptive Statistics of the GPU Dataset

4.2. Model Performance Evaluation

To evaluate the performance of the Linear Regression and XGBoost models in predicting GPU prices, four main evaluation metrics were systematically employed: coefficient of determination (R^2), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). These metrics were selected to provide a comprehensive view of model performance from different perspectives, including explained variance, absolute error magnitude, error sensitivity to outliers, and relative error percentage. Table 2 and Figure 1 present the comprehensive evaluation results of both models, revealing significant performance differences that highlight the superiority of the XGBoost approach. The evaluation results demonstrate not only statistical significance but also practical significance in terms of prediction accuracy improvement.

Table 2. Model Performance Evaluation Result

Evaluation Matrix	Linear Regression	XGBoost
R^2	0,7629	0,8129
MAE (USD)	106,59	85,07
RMSE (USD)	137,38	122,03
MAPE (%)	56,04%	35,23%

ISSN 2579-7069

XGBoost consistently outperformed linear regression across all evaluation metrics, with improvements ranging from moderate to substantial depending on the specific metric considered.evaluate the performance of the Linear Regression and XGBoost models in predicting GPU prices, four main evaluation metrics were used: coefficient of determination (R^2), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). Table 2 and Figure1 shows the evaluation results of the two models. The evaluation results show significant performance differences between the two models. XGBoost not only excels in prediction accuracy but also demonstrates better consistency in handling complex price variations. This finding confirms that the non-linear characteristics in GPU pricing require a more sophisticated approach. The MAPE difference of 20.81% between the two models shows that XGBoost provides substantial improvement in GPU price prediction practices.

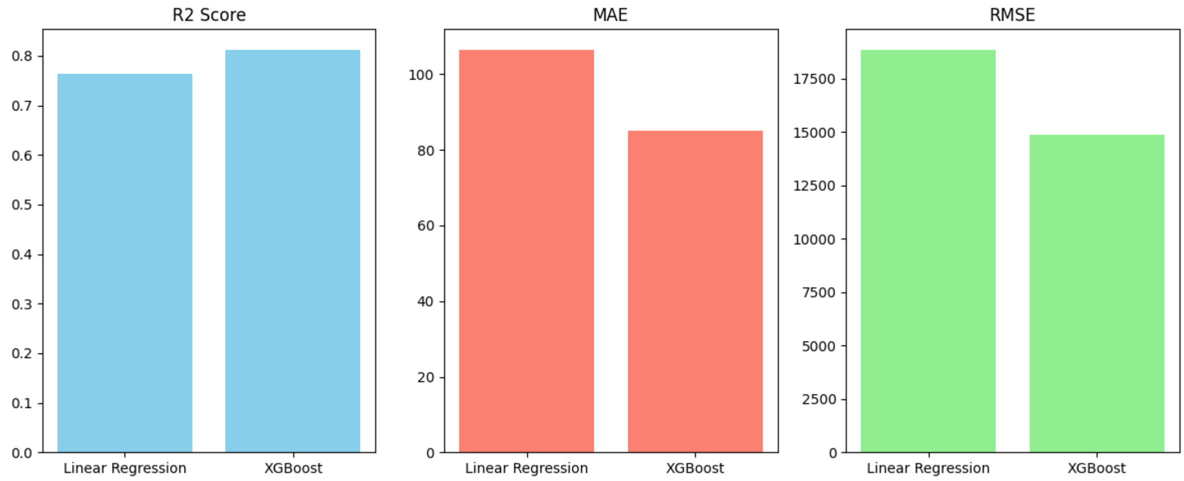

Figure 1. Comparison of R^2, MAE, and RMSE

4.3. Discussion

Based on the evaluation results, the XGBoost model performed better than linear regression on all evaluation metrics. The visual comparison further supports the quantitative findings. The bar plots of R^2, MAE, and RMSE clearly highlight the performance superiority of XGBoost over linear regression. XGBoost achieves a higher R^2 score, indicating better explained variance, and lower error values in both MAE and RMSE. These visualizations reinforce the numerical evaluation and make the performance gap between the two models more intuitive. The advantages of XGBoost have important practical implications for the industry. With higher prediction accuracy, this model can help consumers make more informed purchasing decisions and assist retailers in developing more effective pricing strategies. Additionally, XGBoost's ability to identify the most influential technical features provides valuable insights for the development of future GPU products. The finding that features such as the number of transistors and bandwidth have the greatest influence on price offers insights that can be used for product design optimization and more targeted marketing strategies.

In addition, XGBoost produces a lower MAE (85.07 USD) than linear regression (106.59 USD), which means the average prediction error of XGBoost is smaller. The same can be seen in the RMSE value, where XGBoost has a smaller value (122.03 USD) than linear regression (137.38 USD), indicating that XGBoost is more resistant to large outliers.The MAPE value of 35.23% in XGBoost is also much better than linear regression (56.04%), indicating that the XGBoost model has a lower percentage error rate and is more stable. Overall, the superiority of the XGBoost model can be attributed to its ability to capture the non-linear relationship between the GPU specification and price variables. Meanwhile, linear regression can only model linear relationships and has limitations in handling interactions between features. In terms of feature importance, the XGBoost model identifies “Transistors”, “Max Bandwidth (MB/s)”, and “Ext. Power” as the most influential variables in determining GPU price. These results provide useful insights for manufacturers and consumers alike. Knowing that these features have strong weight in price prediction allows stakeholders to better assess product value and justify pricing strategies. To reinforce the quantitative analysis that has been conducted, this study also presents three additional visualizations to provide a deeper understanding of the model's behavior. First, the feature importance visualization of the XGBoost model (Figure 2) shows that attributes such as

 ISSN 2579-7069

Max Shader Performance (MFLOPS), Max Bandwidth (MB/s), and the number of Transistors have the greatest influence in determining GPU prices, which aligns with the logic of technical performance affecting product value.

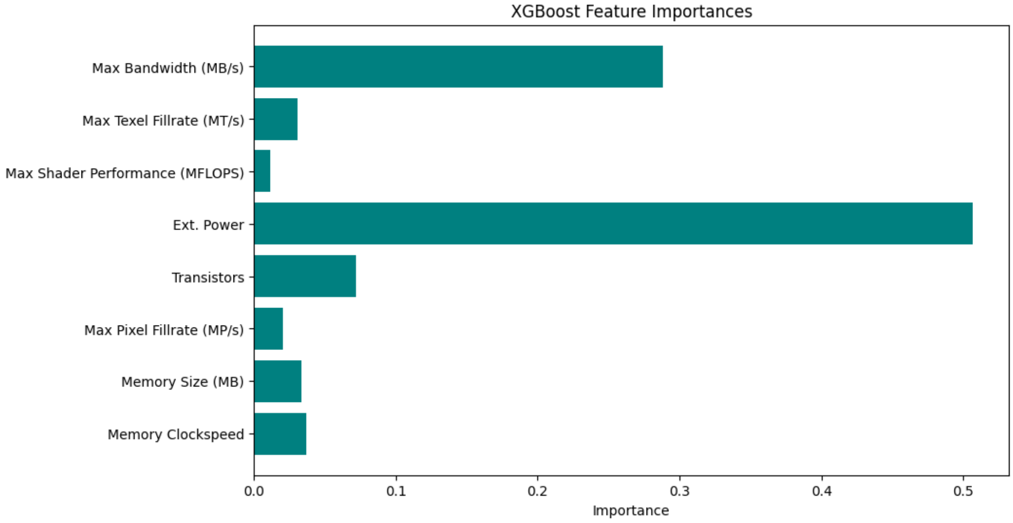

Figure 2. Feature Importance of XGBoost Model

Second, the residual plot shows (Figure 3) that the linear regression model exhibits a pattern of widening residual dispersion, indicating heteroscedasticity, while XGBoost produces residuals that are more concentrated around zero, reflecting more stable prediction errors.

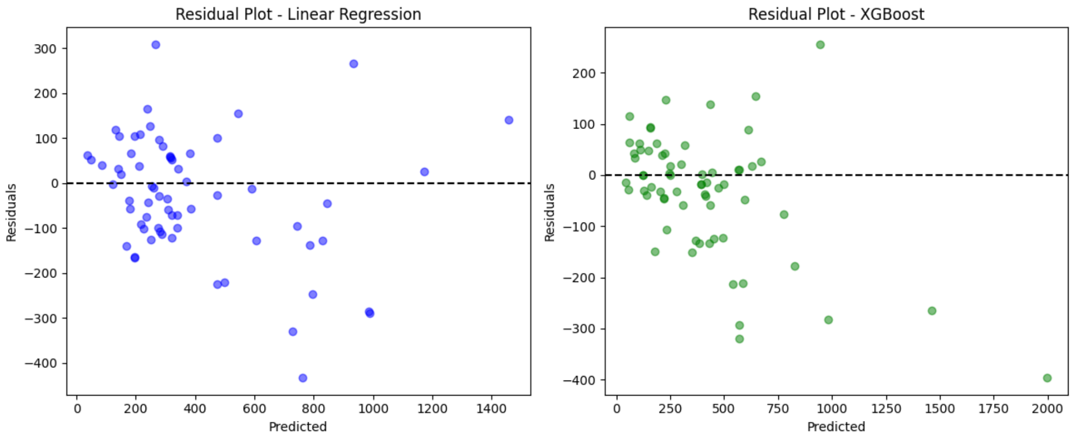

Figure 3. Residual Plot Comparison: Linear Regression vs XGBoost.

Third, the visualization comparing actual prices and predictions (Figure 4) shows that the points on the XGBoost model are closer to the diagonal line compared to linear regression, meaning that the XGBoost model has better prediction accuracy both visually and numerically. These three visualizations support the previous evaluation results that XGBoost excels at capturing non-linear patterns and complex interactions between GPU technical specification features.

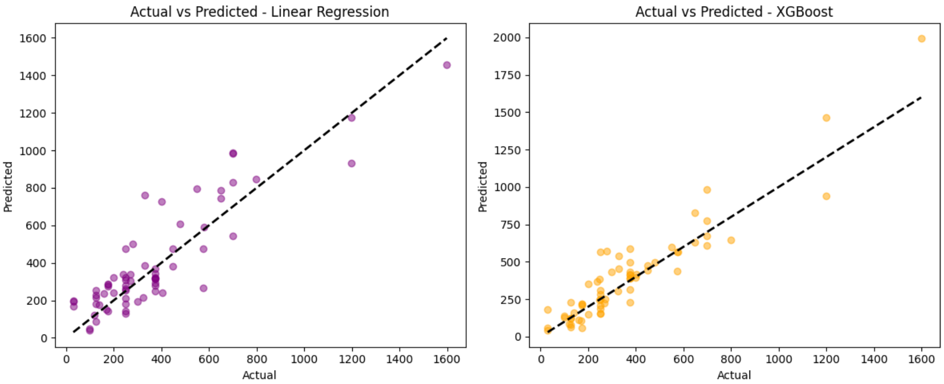

Figure 4. Actual vs Predicted Prices: Linear Regression vs XGBoost

5. Conclusion

This study successfully demonstrated the superiority of XGBoost in predicting GPU prices compared to Linear Regression by a significant margin. XGBoost achieved significantly better performance with an R^2 of 0.8129, MAE of 85.07 USD, RMSE of 122.03 USD, and MAPE of 35.23%, while Linear Regression only achieved an R^2 of 0.7629, MAE of 106.59 USD, RMSE of 137.38 USD, and MAPE of 56.04%. This performance difference indicates that XGBoost can explain 81.29% of GPU price variation, compared to Linear Regression, which can only explain 76.29%. The main findings of this study confirm that the relationship between GPU technical specifications and price is non-linear and complex, requiring a more advanced machine learning approach than simple linear regression. Feature importance analysis shows that features such as the number of transistors, Max Bandwidth (MB/s), and Max Shader Performance (MFLOPS) have the most significant influence on determining GPU prices. This aligns with the logic that higher technical specifications reflect better computational capabilities, which in turn determine the product's value proposition in the market.

Practically, this study makes a significant contribution to various stakeholders in the GPU industry. For consumers, the developed XGBoost model can assist in making more informed purchasing decisions by providing accurate price estimates based on desired technical specifications. For retailers and distributors, this model can be used to optimize pricing strategies and inventory management. Meanwhile, for manufacturers, insights into the features that most influence price can assist in the product development process and strategic positioning in the market. However, this research has several limitations that must be acknowledged. First, the dataset used is limited to a specific period and may not cover price fluctuations caused by external factors such as cryptocurrency market conditions, the global pandemic, or semiconductor shortages. Second, this study has not considered non-technical factors such as brand value, marketing strategy, and availability, which can also influence GPU prices in the market. Third, the developed model has not been tested on real-time data to assess its predictive performance in dynamic market conditions.

For future research, several directions for development are suggested. First, integrating external factors such as cryptocurrency indices, global economic conditions, and market sentiment into the prediction model to improve accuracy. Second, using a broader dataset with a longer time range to capture long-term trends and seasonality in GPU prices. Third, exploring more advanced ensemble learning techniques such as stacking or blending to combine the strengths of various algorithms. Fourth, the implementation of a real-time prediction system that can automatically adapt to changes in market conditions. Finally, the development of a model that can predict not only prices but also price trends over a specific time period to provide more comprehensive insights for stakeholders.

6. Declarations

6.1. Author Contributions

Author Contributions: Conceptualization, D.P.P., M.I., and Q.S.; Methodology, D.P.P. and M.I.; Software, Q.S. and M.I.; Validation, M.I. and Q.S.; Formal Analysis, D.P.P.; Investigation, Q.S. and M.I.; Resources, M.I. and Q.S.; Data Curation, Q.S.; Writing—Original Draft Preparation, D.P.P.; Writing—Review and Editing, M.I. and Q.S.; Visualization, M.I. All authors have read and agreed to the published version of the manuscript.

ISSN 2579-7069

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] H. Va, M. H. Choi, and M. Hong, "Efficient Simulation of Volumetric Deformable Objects in Unity3D: GPU-Accelerated Position-Based Dynamics," *Electronics (Switzerland)*, vol. 12, no. 10, May 2023, doi: 10.3390/electronics12102229.

[2] R. Holla M and D. Suma, "An effective GPU-based random grid secret sharing using an autoencoder image super-resolution," *Cogent Eng*, vol. 11, no. 1, 2024, doi: 10.1080/23311916.2024.2390134.

[3] D. C. Youvan, "Parallel Precision: The Role of GPUs in the Acceleration of Artificial Intelligence", doi: 10.13140/RG.2.2.21937.76641.

[4] A. Belkhiri and M. Dagenais, "Analyzing GPU Performance in Virtualized Environments: A Case Study," *Future Internet*, vol. 16, no. 3, Mar. 2024, doi: 10.3390/fi16030072.

[5] T. Cao, Y. Pan, H. Chen, J. Zheng, and T. Hu, "PPChain: A Blockchain for Pandemic Prevention and Control Assisted by Federated Learning," *Bioengineering*, vol. 10, no. 8, Aug. 2023, doi: 10.3390/bioengineering10080965.

[6] Q. Zhang, X. Li, and P. Gao, "Forecasting Sales in Live-Streaming Cross-Border E-Commerce in the UK Using the Temporal Fusion Transformer Model," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 20, no. 2, p. 92, May 2025, doi: 10.3390/jtaer20020092.

[7] J. Jang, Y. Son, and S. Lee, "A Numerical Study of an Ellipsoidal Nanoparticles under High Vacuum Using the DSMC Method," *Micromachines (Basel)*, vol. 14, no. 4, Apr. 2023, doi: 10.3390/mi14040778.

[8] D. ; Galván-González, S. R. ; Herrera-Sandoval, N. D. ; Guzman-Avalos, P. ; Pacheco-Ibarra, and J. J. ; Domínguez-Mota, "Citation: Molinero-Hernández," 2024, doi: 10.3390/computation.

[9] R. Limas Sierra, J. D. Guerrero-Balaguera, J. E. R. Condia, and M. Sonza Reorda, "Exploring Hardware Fault Impacts on Different Real Number Representations of the Structural Resilience of TCUs in GPUs †," *Electronics (Switzerland)*, vol. 13, no. 3, Feb. 2024, doi: 10.3390/electronics13030578.

[10] K. M S, H. Rajaguru, and A. R. Nair, "Evaluation and Exploration of Machine Learning and Convolutional Neural Network Classifiers in Detection of Lung Cancer from Microarray Gene—A Paradigm Shift," *Bioengineering*, vol. 10, no. 8, Aug. 2023, doi: 10.3390/bioengineering10080933.

[11] H. Wei, Q. Song, C. Dan, Z. He, H. Li, and M. Pu, "Performance Evaluation of ARIMA, Autoformer, and Symmetric LSTNFCL Models for Traffic Accident Emergency Prediction," *Symmetry (Basel)*, vol. 17, no. 5, p. 639, Apr. 2025, doi: 10.3390/sym17050639.

[12] T. O. Kehinde, S. H. Chung, and F. T. S. Chan, "Benchmarking TPU and GPU for Stock Price Forecasting Using LSTM Model Development," in *Lecture Notes in Networks and Systems*, Springer Science and Business Media Deutschland GmbH, 2023, pp. 289–306. doi: 10.1007/978-3-031-37717-4_20.

[13] R. F. Hadi, S. Sa'adah, and D. Adytia, "Forecasting of GPU Prices Using Transformer Method," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 12, no. 1, pp. 136–144, Mar. 2023, doi: 10.32736/sisfokom.v12i1.1569.

[14] M. N. Pangestu, M. Jajuli, and U. Enri, "Prediksi Harga Kartu Grafis NVIDIA Berdasarkan Pengaruh Harga Cryptocurrency Menggunakan Support Vector Regression," *Jurnal Ilmiah Wahana Pendidikan*, vol. 8, no. 17, pp. 280–287, 2022, doi: 10.5281/zenodo.7076540.

[15] M. H. Abbasi *et al.*, "Predicting The Price Of Used Electronic Devices Using Machine Learning Techniques." [Online]. Available: https://www.researchgate.net/publication/377526585

[16] L. Yan, "Predicting House Prices with a Linear Regression Model," 2024, doi: 10.54254/2755-2721/114/2024.18220.

[17] P. Tian, "Research On Laptop Price Predictive Model Based on Linear Regression, Random Forest and Xgboost," 2024.

[18] S. Ballamudi, "International Journal of Robotics and Machine Learning Technologies Journal homepage: www.sciforce.org Comparative Analysis of Machine Learning Models for Laptop Price Prediction An Evaluation of Linear Regression, Histogram Gradient Boosting, and XGBoost Approaches", doi: 10.55124/jmms.v1i1.234.

[19] S. Adrianty and F. Maspiyanti, "Laptop Price Prediction Using Extreme Gradient Boosting Algorithm," *Journal of Applied Research In Computer Science and Information Systems*, vol. 2, no. 1, pp. 132–148, Jun. 2024, doi: 10.61098/jarcis.v2i1.173.

[20] M. Gautam, "crypto price prediction using lstm+xgboost," Jun. 2025, [Online]. Available: http://arxiv.org/abs/2506.22055

[21] J. Shi and Y. Elkhatib, "Accurate GPU Memory Prediction for Deep Learning Jobs through Dynamic Analysis," Apr. 2025, [Online]. Available: http://arxiv.org/abs/2504.03887

[22] K. Yeung, "ARIMA Model Application in Predicting NVIDIA's Stock Price," 2024, doi: 10.54254/2754-1169/128/2024.18620.