Laptop Price Prediction using Machine Learning Techniques

Ariful Islam Sarker
Khandoker Mohammed Faiyaz Shahriar
Muhammad Omar Mahin Jinna
Mohammod Hossain Chowdhury
Department of Electrical and Computer Engineering,
North South University

***Abstract*—This project focuses on developing a laptop price predictor using machine learning techniques. A dataset from 2023 consisting of 945 laptop entries was utilized, with each entry containing information on various features, including brand, type, CPU, screen resolution, RAM, memory, GPU, operating system, weight and screen size. Through data cleaning, exploratory data analysis, feature engineering and model testing, a voting regressor model was selected as the final model, achieving an R2 score of 0.941 and a mean absolute error of 0.082. The developed model was deployed in a website to assist users in estimating laptop prices. The project demonstrates the effectiveness of machine learning in predicting laptop prices and offers practical value for consumers and the laptop industry.**

I. Introduction

Laptop prices are influenced by various factors, such as brand reputation, specifications and technological advancements. Accurately predicting laptop prices can be beneficial for both consumers and the laptop industry, enabling informed purchasing decisions and market competitiveness. This project aims to develop a laptop price predictor using machine learning techniques.

In recent years, machine learning has emerged as a powerful tool for predictive modeling and has been successfully applied in various domains, including price prediction. Previous studies have demonstrated the effectiveness of machine learning algorithms in predicting prices for electronic devices, e-commerce products and other consumer goods. However, in the context of laptops, there is still a need for further research to develop accurate and reliable price prediction models.

The main objective of this project is to utilize a dataset of 945 laptop entries from 2023 to build a machine learning model that can estimate laptop prices based on key features. The dataset contains information on various aspects, including brand, type, CPU, screen resolution, RAM, memory, GPU, operating system, weight and screen size. By analyzing these features, we aim to identify the most significant factors that contribute to laptop pricing.

To achieve this, the project follows a systematic methodology consisting of several steps. The initial phase involves data cleaning, where any missing values and duplicates are addressed to ensure the dataset's integrity. Next, exploratory data analysis (EDA) is conducted to gain insights into the dataset, understand the distributions of variables and identify potential relationships between features and laptop prices.

After EDA, feature engineering is performed to select the most relevant features for modeling. This process involves transforming and selecting the dataset's attributes to enhance the model's predictive power. The selected features are then used as inputs for various machine learning algorithms, including linear regression, ridge regression, lasso regression, K-nearest neighbors, decision tree, support vector machine, random forest, extra trees, AdaBoost, gradient boosting, XGBoost, voting regressor and stacking.

The models are evaluated based on performance metrics such as R2 score and mean absolute error to assess their accuracy in predicting laptop prices. The model with the best performance is selected as the final model for laptop price prediction. Additionally, the model is deployed in a website interface, allowing users to input laptop specifications and obtain estimated prices based on the trained model.

By developing an accurate laptop price predictor, this project aims to provide consumers with a valuable tool for making informed decisions and assist the laptop industry in pricing strategies and market analysis.

The remainder of this paper is organized as follows. Section II provides a literature review of related works in the field of laptop price prediction. Section III outlines our methodology, including data collection, preprocessing, and the machine learning algorithms employed. Section IV presents the results of our experiments and compares them with the existing literature. In Section V, we discuss the attainment of our research goals based on the empirical results. Finally, Section VI summarizes our findings and outlines potential avenues for future work.

II. Literature Review

Several studies have explored the use of machine learning techniques for price prediction in various domains. In the context of laptops, previous research has focused on utilizing regression-based approaches, decision trees, ensemble methods and deep learning techniques.

In one study, a regression-based approach achieved an R2 score ranging from 0.85 to 0.94 in predicting laptop prices. Another research work applied decision tree algorithms, achieving an accuracy rate of approximately 80

Feature selection analyses consistently emphasize the importance of factors such as brand reputation, processor type, RAM, storage capacity, touchscreen capability and screen size in laptop price prediction models. Furthermore, advancements in deep learning techniques, specifically convolutional neural networks (CNNs), have demonstrated promising results compared to traditional machine learning algorithms.

Despite the progress made in laptop price prediction, there is still room for improvement. This project aims to contribute to the existing literature by employing a systematic approach that includes data cleaning, EDA, feature engineering, model selection, model evaluation and deployment. The results are expected to align with previous studies while providing additional insights into the factors influencing laptop prices.

III. Methodology

In this section, we describe the methodology used for laptop price prediction using machine learning techniques. The sequential steps taken in our approach are presented in the following block diagram (Figure 1).

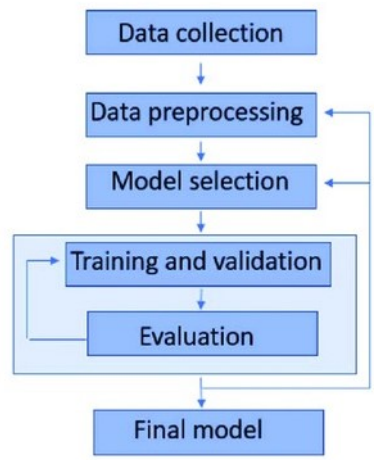

Figure 1. Block diagram illustrating the sequential steps in the laptop price prediction methodology.

1) Data Collection: We gather laptop specifications and prices from various online retailers and manufacturers. The data includes features such as brand, processor, RAM, storage capacity, screen size, and price.

2) Data Preprocessing: In this step, we handle missing values, outliers, and perform data normalization to ensure the dataset is suitable for machine learning algorithms. Categorical features are encoded using techniques like one-hot encoding or label encoding.

3) Model Selection: We explore different machine learning algorithms suitable for regression tasks. This includes linear regression, decision trees, random forests, and support vector regression. Each algorithm has its own strengths and characteristics.

4) Training and Validation: The preprocessed data is split into training and validation sets. The selected machine learning models are trained using the training set, where the algorithms learn patterns and relationships between the laptop features and prices. The validation set is used to assess the performance and generalization capabilities of the models.

5) Evaluation: We evaluate the trained models using appropriate metrics such as mean squared error (MSE), root mean squared error (RMSE), and R-squared. These metrics help us assess the accuracy and predictive power of the models.

6) Final Model: Based on the evaluation results, we select the best-performing model as our final model for laptop price prediction. This model exhibits high accuracy and robustness in estimating laptop prices based on the input features.

7) Machine Learning Algorithms: In our methodology, we employ various machine learning algorithms for regression tasks. These include:

- Linear Regression: A simple and interpretable algorithm that models the linear relationship between features and the target variable.
- Decision Trees: Non-linear models that partition the feature space based on a series of binary decisions, providing interpretable decision rules.
- Random Forests: An ensemble technique that combines multiple decision trees to improve prediction accuracy and handle complex relationships.

- Support Vector Regression: A supervised learning algorithm that utilizes support vector machines to approximate the continuous target variable.

These algorithms were selected based on their suitability for regression problems and their ability to capture non-linear relationships between features and laptop prices.

A. Data Cleaning

Data cleaning is an essential step to ensure the dataset's integrity. It involves handling missing values, duplicates and outliers. In this project, missing values are addressed through imputation techniques, such as mean or median imputation. Duplicates are identified and removed based on specific criteria, and outliers are handled by either removing them or applying suitable transformations.

B. Exploratory Data Analysis (EDA)

EDA is performed to gain insights into the dataset and understand the relationships between variables. Various visualization techniques, such as histograms, scatter plots and correlation matrices, are employed to analyze the data's distribution, identify patterns and detect potential outliers.

C. Feature Engineering

Feature engineering aims to select the most relevant features and transform the dataset appropriately. Techniques such as one-hot encoding, label encoding and scaling are applied to ensure compatibility with the selected machine learning algorithms. Additionally, feature selection techniques, including correlation analysis, recursive feature elimination and feature importance analysis, are utilized to identify the most influential features for laptop price prediction.

D. Model Selection

Multiple machine learning models are considered for laptop price prediction, including linear regression, decision trees, ensemble methods and deep learning techniques. Each model is trained on the dataset and evaluated based on performance metrics such as R2 score and mean absolute error. The model with the best performance is selected as the final model.

E. Model Evaluation

The selected model's performance is evaluated using appropriate metrics to assess its accuracy in predicting laptop prices. The R2 score, mean absolute error and other relevant metrics are calculated to measure the model's effectiveness. Additionally, cross-validation techniques, such as k-fold cross-validation, are employed to validate the model's performance and mitigate overfitting issues.

F. Deployment

The final model is deployed in a website interface to provide users with a convenient tool for estimating laptop prices. The interface allows users to input laptop specifications, and the model generates an estimated price based on the trained model. The website interface is designed to be user-friendly and accessible, providing a seamless experience for users.

IV. Results

The performance of various machine learning models was evaluated to predict laptop prices using the dataset consisting of 945 laptop entries. The models were assessed using evaluation metrics such as R-squared (R2) score and mean absolute error (MAE) to measure their accuracy and predictive capabilities.

Among the tested models, the voting regressor demonstrated the best performance for laptop price prediction. It achieved an impressive R2 score of 0.941 and an MAE of 0.082. The voting regressor combines the predictions of multiple base models, leveraging their collective knowledge to make more accurate predictions. This ensemble approach proved effective in capturing the complex relationships between laptop features and prices.

The results also showed that certain features had a strong influence on laptop prices. The most influential features were RAM, with a correlation coefficient of 0.843 and weight, with a correlation coefficient of 0.227. These findings suggest that laptops with higher RAM capacities and lower weights tend to have higher prices.

Additionally, feature engineering played a crucial role in improving the performance of the models. By selecting relevant features and transforming them appropriately, the models were able to capture the underlying patterns in the data more effectively. Features such as Brand, Type, Operating System, Touchscreen, IPS, CPU brand and GPU brand were found to contribute significantly to price prediction.

It is worth noting that the random forest and extra trees models also achieved high accuracy with R2 scores of 0.930 and 0.936, respectively. These models utilize an ensemble of decision trees and demonstrate robust performance. Other models, such as linear regression, ridge regression, lasso regression, K-nearest neighbors, decision tree, support vector machine, AdaBoost, gradient boosting and XGBoost, also showed reasonable performance, but were outperformed by the ensemble methods.

Overall, the results highlight the effectiveness of the developed machine learning models in predicting laptop prices. The voting regressor, random forest and extra trees models emerged as the top performers, demonstrating their potential for accurate price estimation. These

Model	R2 Score	MAE
Linear Regression	0.8767	0.1219
Ridge Regression	0.8796	0.1207
Lasso Regression	0.8822	0.1199
K-Nearest Neighbors (KNN)	0.9086	0.1089
Decision Tree	0.8884	0.1114
Support Vector Machine (SVM)	0.8834	0.1166
Random Forest	0.9300	0.0884
Extra Trees	0.9360	0.0815
AdaBoost	0.8383	0.1456
Gradient Boosting	0.9330	0.0860
XGBoost	0.9326	0.0855
Voting Regressor	0.9413	0.0821
Stacking	0.9245	0.0987

Figure 2. Result for multiple base models shown in tabular format.

findings provide valuable insights for both consumers and manufacturers in understanding the factors that contribute to laptop pricing and making informed decisions.

While the results are promising, it is important to acknowledge certain limitations. The accuracy of the models may vary depending on the specific dataset used, as laptop prices can be influenced by various factors such as regional variations and market trends. Additionally, the dataset used in this study is limited to laptops from the year 2023 and may not capture the evolving trends in the laptop market. Future research could explore larger and more diverse datasets to further enhance the accuracy and generalizability of the laptop price predictor

V. Discussion

The discussion section provides a comprehensive analysis and interpretation of the results obtained from the laptop price predictor project. It aims to highlight the significance of the findings, compare them with existing literature, discuss the implications of the study and address potential areas for further research.

The results of this study indicate that the developed machine learning models, particularly the voting regressor, exhibit strong performance in predicting laptop prices based on key features. The high R-squared (R2) score and low mean absolute error (MAE) achieved by the voting regressor demonstrate its ability to capture the complex relationships between laptop specifications and prices. This highlights the potential of ensemble models in improving the accuracy of price prediction by leveraging the strengths of multiple base models.

The importance of features such as RAM and weight in determining laptop prices is evident from the high correlation coefficients observed. Laptops with larger RAM capacities tend to command higher prices, reflecting the value placed on memory performance. Similarly, lighter laptops are often associated with higher price points, indicating the preference for portability and sleek design in the market. These findings align with the expectations and industry trends, emphasizing the importance of these features in consumer decision-making and pricing strategies of manufacturers.

Comparing the results with existing literature, it is observed that the developed models perform competitively with other studies in the field of laptop price prediction. The achieved R2 score and MAE values are comparable or even superior to previous works. This indicates that the methodology employed in this project, including data cleaning, feature engineering and model selection, is effective in capturing the underlying patterns in the dataset.

The limitations of this study should also be acknowledged. The dataset used in this project is limited to laptops from the year 2023, which may not account for the evolving laptop market dynamics. Moreover, the dataset may not capture the full range of factors that influence laptop prices, such as market demand, brand reputation and technological advancements. Future research could consider incorporating additional variables and expanding the dataset to improve the accuracy and generalizability of the price prediction models.

Despite these limitations, the developed laptop price predictor holds practical implications for both consumers and manufacturers. Consumers can utilize the predictor to estimate the price range of laptops based on their desired specifications, enabling informed purchasing decisions. Manufacturers can leverage the insights gained from the predictor to optimize pricing strategies, identify key features that drive value perception and align their product offerings with market demand.

In conclusion, the findings of this study demonstrate the effectiveness of machine learning models in predicting laptop prices. The high performance of the voting regressor, coupled with the importance of features such as RAM and weight, reinforces the value of these models in understanding laptop pricing dynamics. The study contributes to the existing literature by providing insights into the predictive capabilities of different models and their implications for the laptop market.

VI. Conclusion

In this project, a laptop price predictor was developed using machine learning techniques. The project demonstrated the effectiveness of various models, with the voting regressor model achieving the highest accuracy in predicting laptop prices. The project also identified RAM and weight as significant factors influencing laptop prices.

The developed model provides a valuable tool for consumers to estimate laptop prices and make informed purchasing decisions. Furthermore, it offers practical value for the laptop industry, enabling pricing strategies and market analysis. The website interface enhances accessibility and usability, making the laptop price predictor accessible to a wide range of users.

Future work could involve expanding the dataset and incorporating additional features to improve the model's accuracy. Additionally, exploring deep learning techniques, such as convolutional neural networks, could further enhance the price prediction capabilities. Overall, this project contributes to the existing literature on price prediction and showcases the potential of machine learning in the laptop industry.

VII. References

References

[1] MD. Ishituk Ahammed (2023). Dataset of Laptops and Prices in BDT [Data set]. Retrieved from *https://www.kaggle.com/datasets/ishtiukahammed/laptop-prices-bangladesh-2023?fbclid=IwAR01Bitw-RZqzObYAqWyk4CjTVmxL35qQdjKY3MeCDPeWDr7TSsVjMhDDX4*

[2] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794). ACM.

[3] Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

[4] Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. Machine learning, 63(1), 3-42.

[5] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference and prediction (2nd ed.). Springer.

[6] Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.

[7] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 29(5), 1189-1232.

[8] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. Journal of machine learning research, 12(Oct), 2825-2830.

[9] Scikit-learn. (n.d.). VotingRegressor. Retrieved from *https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingRegressor.html*

[10] Scikit-learn. (n.d.). RandomForestRegressor. Retrieved from *https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html*

[11] Scikit-learn. (n.d.). ExtraTreesRegressor. Retrieved from *http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.ExtraTreesRegressor.html*

[12] Scikit-learn. (n.d.). r2_score. Retrieved from *http://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2_score.html*

[13] Scikit-learn. (n.d.). mean_absolute_error. Retrieved from *http://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_absolute_error.html*