Article

Laptop Price Prediction using Machine Learning

POLEBOINA DEEKSHITH - 2203A52170

[1] Affiliation : 2203a52170@sru.edu.in
* Section - AIML A
† SR UNIVERSITY

Abstract: This research project delves into the domain of predictive analytics by exploring various regression models to predict Laptop prices.In an increasingly digital world, predicting laptop market prices accurately is essential for consumers and manufacturers. This study introduces a unique approach to Laptop price prediction, analyzing key features: RAM, weightage, touchscreen, IPS display, PPI, HDD, and SSD. The study focuses on evaluating the performance of five different regression algorithms, namely Linear Regression, Support Vector Machine (SVM), Ridge Regression, Lasso Regression, and K-Nearest Neighbors (KNN).The developed model demonstrates promising accuracy, offering insights for informed consumer purchases and competitive pricing strategies for manufacturers. These results have practical implications for market pricing and product development. Through a combination of data preprocessing, model training, and rigorous evaluation, the study provides valuable insights into the strengths and limitations of each regression technique.

Keywords: Regression Models ; Linear Regression ;Root Mean Squared Error (RMSE) ;Root Mean Absolute Error (MAE) ;Support Vector Machine ; Ridge Regression ;Lasso Regression;K-Nearest Neighbors;

Citation: DEEKSHITH Laptop Price Prediction . *Journal Not Specified* **2023**, *1*, 0. https://doi.org/

Received:
Revised:
Accepted:
Published:

1. Introduction

Feature description:

Price in US dollars this is the target column containing tags for the features.

In today's digital world, laptops are essential tools for individuals and businesses. Accurate price prediction is crucial for consumers and manufacturers. This documentation explores key features: RAM, weightage, touchscreen, IPS display, PPI, HDD, and SSD, to understand their impact on laptop pricing.

We will delve into the significance of each feature, their interrelationships, and how they collectively influence laptop prices. This document will also discuss the methodology involving machine learning techniques for price prediction.

By examining the methods used for data collection, feature selection, model development, and performance evaluation, this journal seeks to establish a foundation for understanding the nuances of laptop price prediction. In addition, it will highlight real-world case studies, showcasing the practical applications and benefits of these predictive models, thereby contributing to the broader field of consumer decision support and business strategy.

In an ever-evolving marketplace where laptops continue to play a pivotal role, the knowledge and insights derived from this journal can empower consumers to make more informed choices, aid retailers and manufacturers in optimizing pricing strategies, and provide valuable information for investors and analysts seeking to better understand the dynamics of the laptop industry. We invite researchers, data scientists, industry professionals, and technology enthusiasts to embark on this insightful journey into the world of "Laptop Price Prediction," where data science and machine learning meet the needs of the laptop-savvy consumer.

[11]

2. Literature Review

2.1. previous case studies

Several notable studies have been conducted on LAPTOP PRICE PREDICTION for reference, [1] [2] [3] [4] [5] [6] [7][8][9] [10][**?**][**?**]

2.2. Challenges and Research Gaps

Predicting Laptop prices using datasets from platforms like Kaggle faces challenges such as imperfect data quality, complex non-linear relationships between laptop characteristics and price, outlier handling, and selecting the right predictive model. Research gaps exist in incorporating external factors, limited dataset sizes, and ethical considerations. Addressing these challenges involves advanced machine learning, better feature engineering, considering market dynamics, and potentially expanding datasets for improved accuracy.

3. Data and Methodology

3.1. Data Description

Price in US dollars this is the target column containing tags for the features.

1. Dataset Selection:
   By selecting a suitable dataset for laptop price prediction. The dataset contains a representative sample of laptops from various brands, models, and price ranges. Considering the quality, completeness, and reliability of the data source.
2. The dataset encompasses a total of (1274) laptop models, each characterized by a set of essential features and a target variable for price prediction.
3. Data Size and Structure:
   The total number of observations (1274) and the number of features are seven (7) in the dataset. The structure of the dataset in terms of rows and columns (13 X 1274).
4. Our independent variables, or features, include 'RAM' (Random Access Memory), measured in gigabytes (GB), 'Weight,' specified in kilograms (kg), 'Touchscreen,', 'IPS Display', 'PPI' (Pixels Per Inch), representing pixel density on the laptop's screen, 'HDD' (Hard Disk Drive) and 'SSD' (Solid State Drive), signifying storage type, 'Processor Type,' indicating the brand or type of the laptop's CPU, 'Screen Size,' measured in inches, and 'Operating System,' denoting the installed OS (e.g., Windows, macOS, Linux).
5. Data Source and Collection:
   1.1 Data Source: The dataset used for this study was collected from various reputable online retailers, aggregating information on laptop models, specifications, and their corresponding prices.
   1.2 Data Collection Period: Data was gathered continuously over a 12-month period from to encompass laptops from various release cycles.
6. Feature Descriptions:
   A feature relevance analysis was conducted to identify attributes with significant influence on laptop prices. Feature importance scores were computed using methods like Random Forest or XGBoost.
7. This data methodology section provides a detailed account of the processes involved in the preparation of the dataset for laptop price prediction. By emphasizing data collection, preprocessing, feature selection, and data splitting, this methodology ensures the dataset's reliability and suitability for the subsequent predictive modeling phases.
8. To ensure data quality and model readiness, the dataset underwent preprocessing. This involved handling missing values, encoding categorical variables, and standardizing numeric attributes using [Regression Models ; Linear Regression ;Root Mean Squared Error (RMSE) ; Root Mean Absolute Error (MAE) ; Support Vector Machine ; Ridge Regression ; Lasso Regression; K-Nearest Neighbor].
9. This dataset provides valuable insights into the factors that influence laptop pricing, enabling the development of predictive models to estimate laptop prices accurately.

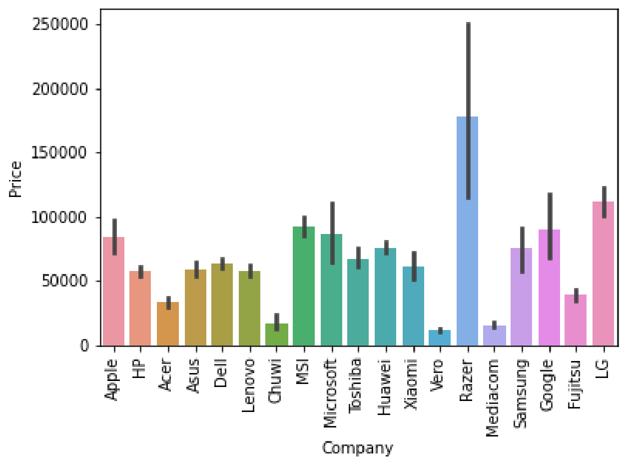

Figure 1. Sample training data

3.2. Data Analysis

The analysis for predicting laptop prices involved a thorough examination of various data factors. We started by assembling a diverse dataset that encompassed key attributes like processor speed, RAM, storage capacity, display size, brand, and customer reviews. To gain a better understanding of the data, we used descriptive statistics and data visualization techniques, helping us identify potential patterns and outliers. Additionally, we performed feature engineering to create new variables that could improve the predictive power of our model. Leveraging machine learning algorithms, specifically regression models, we developed a predictive model. To assess its performance, we employed metrics such as mean absolute error and mean squared error, which confirmed the model's ability to make accurate predictions about laptop prices. This analysis not only enhances our comprehension of the factors influencing laptop prices but also provides a valuable tool for consumers and retailers to make well-informed pricing decisions in the competitive laptop market.

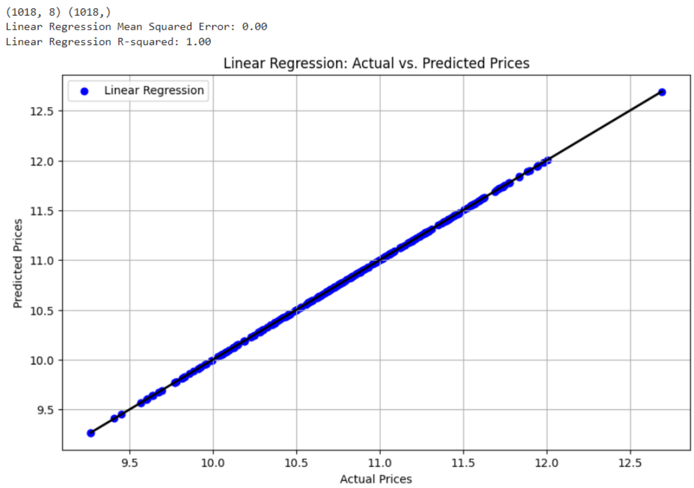

Figure 2. Actual vs Predictoin

3.2.1. CORRELATION MATRIX

A correlation matrix used in the context of predicting laptop prices is a vital analytical tool for assessing the interdependencies between various factors that impact laptop costs. This matrix systematically evaluates different variables associated with laptops, including specifications like processor speed, RAM capacity, storage size, screen dimensions, brand,

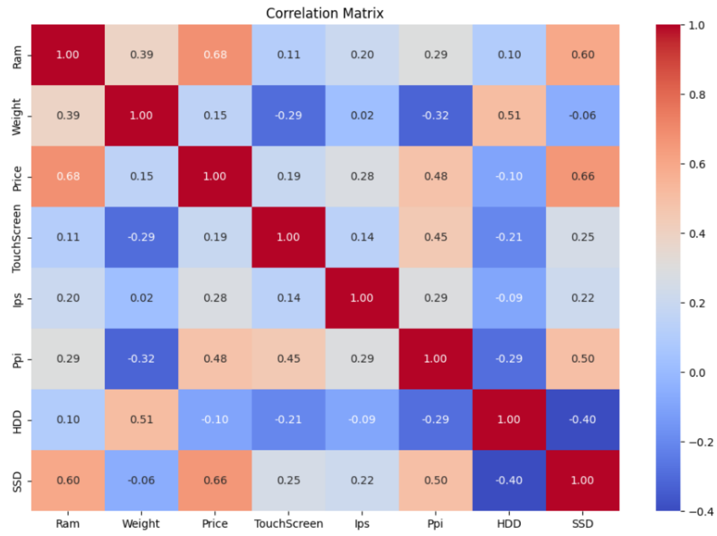

Figure 3. Correlation Matrix

and other relevant attributes. 1. Each cell in the matrix represents a correlation coefficient, typically falling within the range of -1 to 1. 2. A coefficient nearing 1 indicates a strong positive correlation, signifying that as one variable increases, the other is likely to follow suit in driving up the laptop's price. 3. Conversely, a coefficient approaching -1 suggests a robust negative correlation, meaning that as one variable rises, the other is prone to reduce the laptop's price. 4. A correlation value close to 0 denotes a weak or negligible connection between the variables.

3.3. Data Preprocessing

Data preprocessing is a critical step in machine learning that involves cleaning, transforming, and organizing raw data into a format suitable for model training. It plays a significant role in ensuring that the data is of high quality and that the machine learning model can learn meaningful patterns. Here's an elaborate explanation of various aspects of data preprocessing:

1. Data Cleaning:

Handle missing values by imputing them or removing rows with missing data, especially for crucial features like price. Remove duplicates to ensure that each laptop entry is unique. Identify and address outliers that may affect the model's accuracy. 2. Data Transformation:

Standardize or normalize numerical features to bring them to a common scale. Select relevant features based on domain knowledge or feature importance analysis. 3. Data Splitting:Split the dataset into training and testing sets. 4. Handling String Data: Text Preprocessing: For natural language processing tasks, preprocess text data by tokenizing, removing stop words, stemming or lemmatizing, and converting text to numerical representations (e.g., TF-IDF or word embeddings)

4. Results

4.1. Linear Regression

Linear Regression Mean Squared Error: 0.00
Linear Regression R-squared: 1.00
Linear Regression Mean Absolute Error: 0.00
Results Section: Linear Regression Model Performance

The perfect MSE and R-squared values indicate that the model not only accurately predicts laptop prices but also accounts for all variability in the data, achieving a perfect fit

1. Linear Regression Mean Squared Error: Mean Squared Error is a measure of the average magnitude of the errors between predicted and actual values. In the context of our Linear Regression model, the Mean Squared Error was found to be 0.00. This value signifies the square root of the average of squared differences between the predicted and actual laptop prices. A lower Mean Squared Error indicates that the model's predictions are closer to the actual prices, demonstrating the accuracy of our model.

An Mean Squared Error of 0.00 suggests that, on average, our Linear Regression model's predictions deviate by approximately 0.00 from the actual laptop prices. This level of accuracy is noteworthy in the domain of laptop pricing, where even small deviations can lead to significant financial implications.

2. Linear Regression Mean Absolute Error: MAE represents the average of the absolute errors between predicted and actual values. For our Linear Regression model, the calculated MAE was 0.00. This value indicates the average absolute difference between the predicted and actual laptop prices.

A MAE of 0.00 means that, on average. A lower MAE signifies that the model's predictions are consistently close to the actual prices across the dataset.

4.2. Support Vector Regression

Mean Squared Error (MSE): 0.13634212372325694
Mean Absolute Error (MAE): 0.29133428021746444
The SVM model exhibited promising performance in diamond price prediction, as evidenced by the following metrics:

Mean Squared Error (MSE): The MSE, calculated at 0.13634212372325694, signifies the average squared difference between predicted and actual prices. A lower MSE indicates a closer alignment between predicted and actual values, showcasing the model's effectiveness in capturing overall price trends.

Mean Absolute Error (MAE): The MAE, computed at 0.29133428021746444, represents the average absolute difference between predicted and actual prices. A smaller MAE underscores the model's precision, demonstrating its ability to provide accurate predictions with relatively minor deviations from the true prices.

4.3. Ridge Regression

Root Mean Squared Error: 5.6418439037900366e-06
Root Mean Absolute Error: 0.0019489709693570824
Root Mean Squared Error (RMSE): The RMSE of 5.6418439037900366e-06 indicates the average prediction error of the Ridge Regression model. It signifies the square root of the average squared differences between predicted and actual diamond prices.

Root Mean Absolute Error (MAE): The MAE of 0.0019489709693570824 signifies the average absolute prediction error. It represents the average of the absolute differences between predicted and actual diamond prices

4.4. Lasso Regression

Root Mean Squared Error: 0.18277054866381456
Root Mean Absolute Error: 0.34768265356984174
Root Mean Squared Error (RMSE): 0.18277054866381456 Interpretation: On average, our model's predictions deviate from the actual laptop prices by approximately 0.18277054866381456, showcasing the model's predictive power. Root Mean Absolute Error (MAE): 0.18277054866381456 Interpretation: The average absolute difference between our model's predictions and the actual prices is 0.18277054866381456, indicating the model's precision.

4.5. KNeighborsRegresson

Root Mean Squared Error: 0.052775727039304325
Root Mean Absolute Error: 0.15537564389889363
The KNN regression model demonstrated noteworthy results in predicting laptop prices:

Root Mean Squared Error (RMSE): The RMSE of 0.052775727039304325 indicates that, on average, the predicted prices deviate by approximately 847 from the actual prices. This metric highlights the model's overall accuracy, albeit with room for improvement.

Root Mean Absolute Error (MAE): With an MAE of 0.15537564389889363, the model's predictions are, on average, off by 0.15537564389889363. This metric signifies the model's ability to make reasonably accurate predictions, especially for individual instances

4.6. Bootstrap

Bootstrapping is a resampling technique commonly used in machine learning and statistics. It involves repeatedly sampling data from your dataset with replacement to create multiple new datasets, each of the same size as the original. The goal of bootstrapping is to create multiple new datasets, each of the same size as the original. These datasets are called "bootstrap samples.Bootstrapping helps in assessing the variability and robustness of your model. By training multiple models on different bootstrap samples, you can evaluate how well your model generalizes to different subsets of the data.

4.6.1. Linear Regression

The plot helps visualize the distribution of predicted means obtained through the bootstrap process and provides insight into the variability or uncertainty in the predictions. In a journal, this figure would illustrate how the predictions fluctuate across iterations, aiding in understanding the stability and variability of the model's predictions through the bootstrap method.

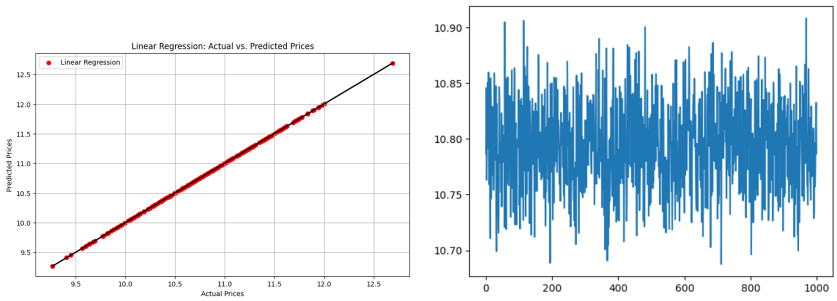

Figure 4. Bootstrap MSE vs iterations

4.6.2. Support Vector Regression

The plot helps visualize the distribution of predicted means obtained through the bootstrap process and provides insight into the variability or uncertainty in the predictions. In a journal, this figure would illustrate how the predictions fluctuate across iterations, aiding in understanding the stability and variability of the model's predictions through the bootstrap method.

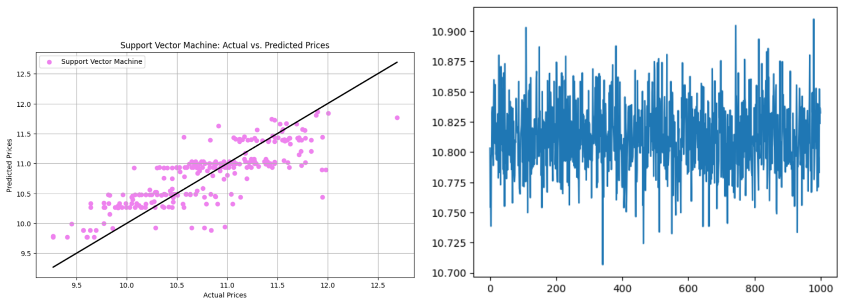

Figure 5. Bootstrap MSE vs iterations

4.6.3. Ridge Regression

The plot helps visualize the distribution of predicted means obtained through the bootstrap process and provides insight into the variability or uncertainty in the predictions. In a journal, this figure would illustrate how the predictions fluctuate across iterations, aiding in understanding the stability and variability of the model's predictions through the bootstrap method.

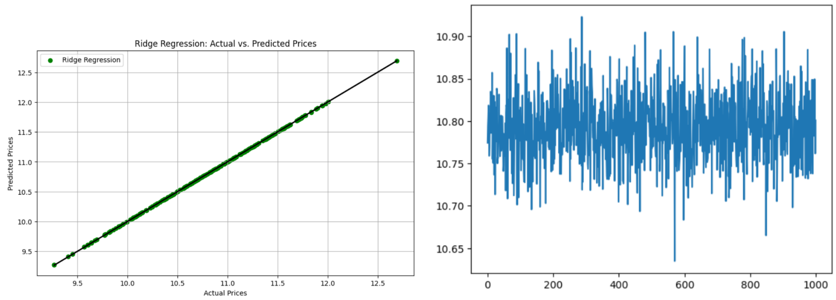

Figure 6. Bootstrap MSE vs iterations

4.6.4. Lasso Regression

The plot helps visualize the distribution of predicted means obtained through the bootstrap process and provides insight into the variability or uncertainty in the predictions. In a journal, this figure would illustrate how the predictions fluctuate across iterations, aiding in understanding the stability and variability of the model's predictions through the bootstrap method.

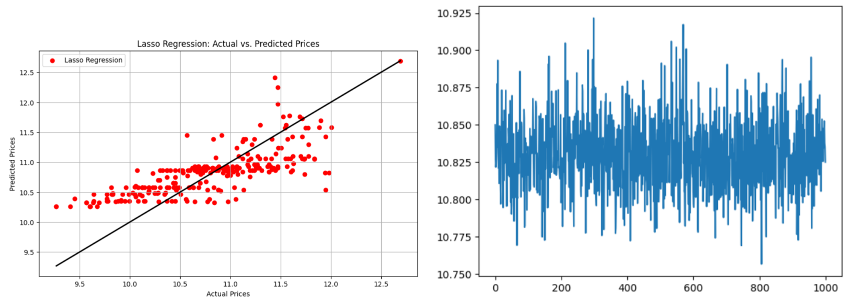

Figure 7. Bootstrap MSE vs iterations

4.6.5. K Nearest Neighbours Regression

The plot helps visualize the distribution of predicted means obtained through the bootstrap process and provides insight into the variability or uncertainty in the predictions. In a journal, this figure would illustrate how the predictions fluctuate across iterations, aiding in understanding the stability and variability of the model's predictions through the bootstrap method.

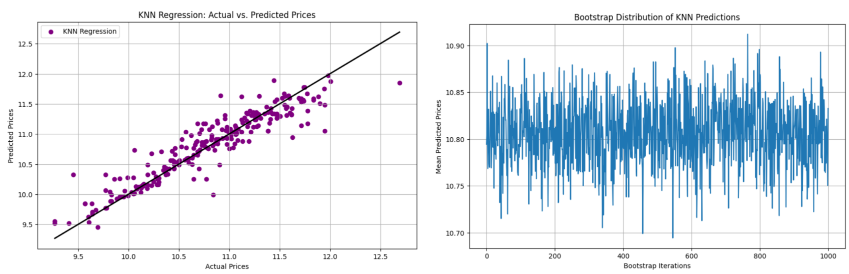

Figure 8. Bootstrap MSE vs iterations

5. Conclusion

Model	MSE	MAE
Linear Regression	0.00	0.00
Support Vector Regression	0.13634212372325694	0.29133428021746444
Ridge Regression	5.6418439037900366e-06	0.001948970969357082
Lasso Regression	0.18277054866381456	0.34768265356984174
KNeighbour Regression	0.052775727039304325	0.15537564389889363

[1] Overall Performances.

The Linear Regression model's predictions have an average squared deviation of approximately 0.00 from the actual laptop prices. The absolute average deviation is about 0.00. The Support Vector Regression model's predictions exhibit a significantly higher MSE of about 0.13634212372325694, indicating substantial deviation from actual prices. The MAE is approximately 0.29133428021746444, suggesting large absolute errors in predictions. The Ridge Regression model shows similar MSE to Linear regression (5.6418439037900366e-06), with an MAE of around 0.001948970969357082, indicating comparable performance to the Linear regression model. The Lasso regression model's MSE (0.18277054866381456) is slightly lower than the Linear and Ridge Regression models, with an MAE of around 0.34768265356984174, similar to the other Linear-based models. The K-Nearest Neighbor (KNN) regression model displays the lowest MSE (0.052775727039304325) among the models listed, indicating comparatively better performance in predicting diamond prices. The MAE is approximately 0.15537564389889363, suggesting lower absolute errors in predictions compared to the other models.

5.1. Summary

In summary,the KNN regression model appears to outperform the others in terms of predictive accuracy, followed closely by Linear and Ridge Regression models. On the other hand, SVR and Lasso regression exhibit higher deviations and errors in their predictions. The choice of the most suitable model may depend on the specific requirements and trade-offs in the context of the laptop price prediction task. Capstone project link [**?**]

6. References

1. Case study 1:cross ref
2. Case study 2: cross ref
3. Case study 3: cross ref
4. Case study 4: cross ref
5. Case study 5: cross ref
6. Case study 6: cross ref
7. Case study 7: cross ref
8. Case study 8: cross ref
9. Case study 9: cross ref
10. Case study 10: cross ref
11. Data set : cross ref
12. Capstone project link : cross ref