International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Laptop Price Estimation Using Data-Driven Predictive Analytics

Dr. Lokendra Singh[1], Mitali Gupta[2], Muskan Jain[3]

[1,2,3]dept. of computer science and engineering
Medi-Caps University
Indore, India
[1]lokendra.songare@gmail.com, [2]guptamitalii12@gmail.com, [3]muskanjain71689@gmail.com

Abstract

In today's fast-paced world, selecting a laptop that provides best performance and durability under budget is a challenge for consumers. This paper is based on predicting prices of laptops based on some inputs or specifications inputted by the user. It uses a cleaned dataset of laptops — 1,300 records — filtering out duplicates and missing variables. It has been tested for different models like XGBoost, Random Forest, Stacking model, Linear Regression, SVM, Ridge Regression, Gradient Boost, KNN, Decision tree, etc. These models investigated the cost implications of laptop specifications like brand, type, RAM, weight, touchscreen, screen size, screen resolution, CPU, HDD, SSD, GPU and operating system. Performance was being evaluated using R2 score and Mean Absolute Error (MAE). According to the results, Random Forest outperformed the others, obtaining the highest R2 score and the lowest MAE. This study assists consumers in making informed selections, and allows retailers to combine data-driven pricing strategies.

Keywords: Laptop Price Prediction , Random forest, XGBoost, Linear Regression, R2 score, RMSE, SVM, KNN.

1. Introduction

Both customers and retailers need to be able to predict laptop costs in order to make well-informed judgments and set prices strategically. The increasing variety of laptop models and the speed at which technology is developing have made it difficult to determine a fair laptop price based just on specifications. Machine learning offers powerful prediction techniques to investigate how a variety of factors, such as CPU type, size of display, windows operating system, RAM, memory type, and brand, affects price.

A broad variety of criteria, like processor type, display size, operating system, memory capacity, storage type, graphics card, weight, and brand reputation, influence the pricing of laptops, which are becoming more and more different in the market. We are using a dataset of 1,300 records of laptops and tested it against different machine learning algorithms to train the model. The model is tested with algorithms like Linear Regression, Random Forest, XGBoost, Gradient Boost algorithm and many others. The dataset is properly cleaned, analysed and the missing or duplicate values were eliminated in order to increase the accuracy.

It is highly advantageous for both buyers as well as sellers when predicting laptop prices accurately.

Machine learning models provide an effective solution to this problem by identifying patterns in historical pricing data and making precise predictions.

2. Related work

Using XGBoost on 992 laptop records, Kafabihi (2024) demonstrated that price prediction is greatly impacted by RAM, processor brand, and storage kind. The study emphasized the significance of feature selection while confirming XGBoost's higher accuracy (R2: 0.84). Competitive pricing and consumer decision-making are facilitated by machine learning. Price prediction models for customers and producers are improved by using a variety of hardware features.[1]

Predicting laptop prices is a crucial task, particularly in cases where laptops are supplied straight from the manufacturer. Using supervised machine learning approaches, this forecast examines variables like a laptop's model, RAM, and CPU, frequently using multiple linear regression (R2: 0.81). Since anticipating a price class might often be more useful than guessing a specific price, recent research has also looked into predicting laptop price ranges.[2]

Using a dataset of 1320 samples, the study examines laptop price prediction using the Linear Regression, Random Forest, and XGBoost models. It determines that the XGBoost model has the best accuracy by comparing RMSE and R2 values (R2: 0.85, RMSE: 294.11). RAM, CPU, weight, and GPU are identified via feature importance analysis as the main determinants of price prediction. This study helps consumers make well-informed purchasing decisions and dealers plan competitive pricing.[3]

Previous research highlights how technology may increase system accuracy and efficiency, highlighting methods that lead to better results. Additionally, a number of research addresses the incorporation of contemporary methods like artificial intelligence and machine learning to automate and improve procedures. This study uses Gradient Boosting Regressor to predict the price of laptops with R2 score of 0.88 and MSE was 0.06.[4]
Several laptop price prediction solutions from Kaggle contests are included in the current system, which combines conventional machine learning techniques with novel concepts including neural networks, residual regression, and logit transform. However, because of the small dataset sizes, the findings of the laptop price fluctuation forecast are not always accurate and can have high standard deviations. By applying machine learning techniques to create a more precise laptop price prediction tool, this study seeks to address these problems. The methodology entails pre-processing the data, applying the Random Forest algorithm to train a model, and evaluating the accuracy of the model.[5]

Regression-based techniques are used in the majority of current price prediction studies to estimate a certain price value. However, predicting a pricing range is far more practical for many real-world applications. There is just one study on laptop price range prediction in the literature, despite the fact that there are numerous studies on laptop price prediction. Furthermore, the laptop price range prediction problem has seen very little testing of machine learning techniques. A dataset that was originally used for laptop price prediction was modified to be used for laptop price range prediction. Preprocessing techniques like data cleaning, feature engineering, and label encoding were used to optimize the dataset for laptop price range prediction giving R2 score of 0.70.[6]

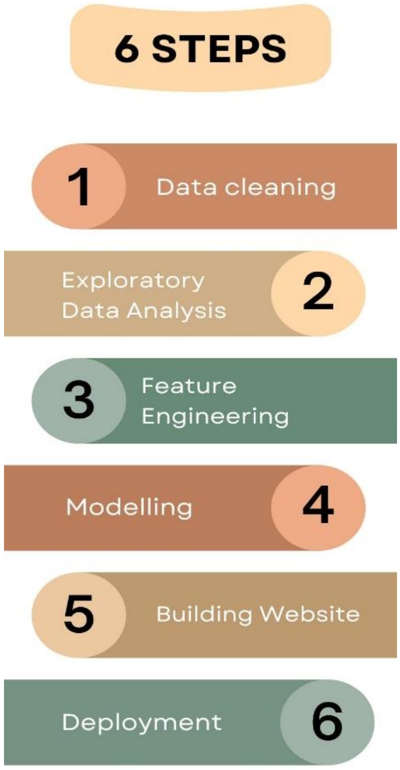


Fig. 1: Steps done

The literature study outlines many methods for predicting laptop prices through machine learning. Several regression models, such as Decision Tree, Random Forest, and Multiple Linear Regression, have been investigated in earlier research to increase prediction accuracy. According to research, Decision Tree algorithms perform well on multidimensional datasets, improving precision and minimizing overfitting with precision of 81%. Furthermore, when calculating laptop prices, features like RAM, CPU, GPU, and storage are highlighted by feature importance analysis. These results help create more trustworthy pricing models, which help customers and merchants make wise choices.[7]

With linear regression being a popular predictive analysis tool, machine learning (ML) is essential to contemporary automation. It is used to represent the interactions between variables in engineering, healthcare, and finance. By using optimization and data visualization techniques, MATLAB improves the implementation of linear regression. Regularization techniques are used by academics to overcome constraints such as sensitivity to outliers. Despite difficulties, linear regression is still essential to machine learning, and its computational power and efficiency are constantly increasing.[8]

Previous research on laptop price prediction primarily used regression-based techniques to identify particular price points. More recent research, on the other hand, recognizes the greater value of forecasting price ranges for actual situations. Only one study has been found in the literature, indicating that while laptop price prediction has been the subject of many research, laptop price range prediction has gotten less attention. Furthermore, research into machine learning methods for predicting laptop price ranges is still in its early phases.[9]

Supervised learning techniques are frequently used in laptop price prediction research, with regression being used for direct price estimation and classification being used for price range prediction. Even though a variety of techniques, such as neural networks and regression models, have been used to address this issue, it is still difficult to obtain reliable results, particularly when working with small datasets. Predicting price ranges has been investigated recently as a more useful strategy than exact numbers; nonetheless, additional research in this field using a variety of machine learning techniques is necessary.[10]

3. Proposed Methodology

Through an analysis of important hardware and software attributes, this study attempts to forecast laptop pricing using different machine learning methods. A systematic strategy is used in the study methodology, which includes data collection and cleaning, exploratory data analysis [EDA], feature engineering, model selection, building website and deployment .

1. Data Cleansing

For accurate forecasting, we cleaned 1,300 laptop records by deleting duplicates and missing values either by deletion or imputation, with outlier detection for price, RAM, and storage. Categorical variables such as brand and OS were standardized for uniformity. Proper data cleansing reduces errors and improves model performance, thus generating enhanced outcomes.

	Unnamed: 0	Company	TypeName	Inches	ScreenResolution	Cpu	Ram	Memory	Gpu	OpSys	Weight	Price
0	0	Apple	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 2.3GHz	8GB	128GB SSD	Intel Iris Plus Graphics 640	macOS	1.37kg	71378.6832
1	1	Apple	Ultrabook	13.3	1440x900	Intel Core i5 1.8GHz	8GB	128GB Flash Storage	Intel HD Graphics 6000	macOS	1.34kg	47895.5232
2	2	HP	Notebook	15.6	Full HD 1920x1080	Intel Core i5 7200U 2.5GHz	8GB	256GB SSD	Intel HD Graphics 620	No OS	1.86kg	30636.0000
3	3	Apple	Ultrabook	15.4	IPS Panel Retina Display 2880x1800	Intel Core i7 2.7GHz	16GB	512GB SSD	AMD Radeon Pro 455	macOS	1.83kg	135195.3360
4	4	Apple	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 3.1GHz	8GB	256GB SSD	Intel Iris Plus Graphics 650	macOS	1.37kg	96095.8080

Fig. 2: Dataset before Data Cleansing

	Company	TypeName	Inches	ScreenResolution	Cpu	Ram	Memory	Gpu	OpSys	Weight	Price
0	Apple	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 2.3GHz	8	128GB SSD	Intel Iris Plus Graphics 640	macOS	1.37	71378.6832
1	Apple	Ultrabook	13.3	1440x900	Intel Core i5 1.8GHz	8	128GB Flash Storage	Intel HD Graphics 6000	macOS	1.34	47895.5232
2	HP	Notebook	15.6	Full HD 1920x1080	Intel Core i5 7200U 2.5GHz	8	256GB SSD	Intel HD Graphics 620	No OS	1.86	30636.0000
3	Apple	Ultrabook	15.4	IPS Panel Retina Display 2880x1800	Intel Core i7 2.7GHz	16	512GB SSD	AMD Radeon Pro 455	macOS	1.83	135195.3360
4	Apple	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 3.1GHz	8	256GB SSD	Intel Iris Plus Graphics 650	macOS	1.37	96095.8080

Fig. 3: Dataset after Data Cleansing

2. Exploratory Data Analysis (EDA)

EDA assists in locating correlations, patterns, and trends within the dataset. We employed box plots for outliers, scatter plots for feature associations, and histograms for price distribution. Important variables like RAM, storage kind, and processor brand were highlighted using a correlation heatmap. The relationship between laptop prices and brands is depicted in the diagram below, which offers information on company-specific pricing patterns for improved forecast accuracy.

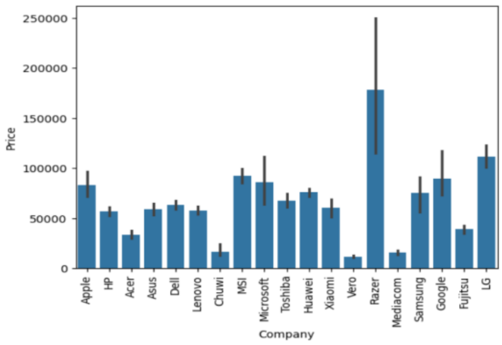
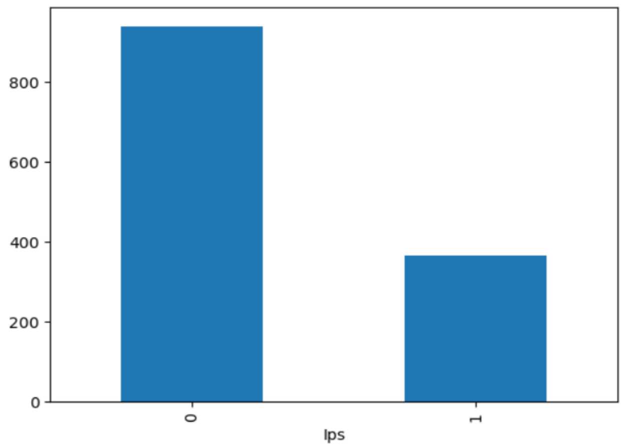


Fig. 4: Company v/s Price

3. Feature Engineering

By choosing important characteristics like brand, RAM, storage type, CPU, GPU, and screen size, we converted raw data into an appropriate format. We have created a new column for “Touchscreen” that will be either 0 or 1. If the laptop is touchscreen the value will be 1 else 0. One-hot encoding was used to encode categorical variables, and numerical features were scaled for uniformity. Secondly, we have created another new column for “IPS panel”, if there is an IPS panel present in the laptop then the value will be 1 otherwise 0.



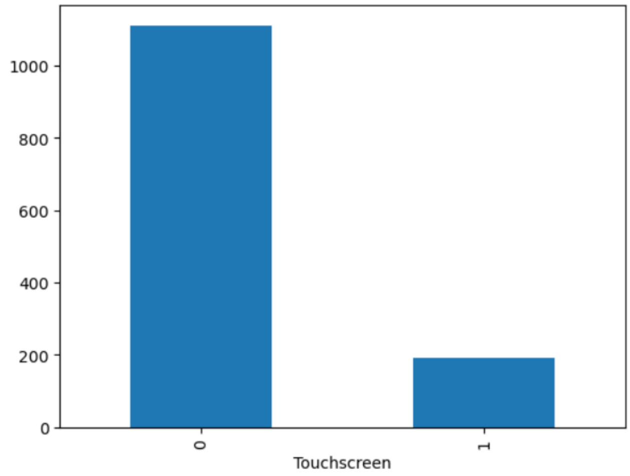


Fig. 5: Feature engineering

4. Modelling

We utilized a number of machine learning algorithms to predict laptop prices such as, but not limited to, Linear Regression, Ridge Regression, Decision Tree, Random Forest, SVM, Gradient Boost, XGBoost, KNN and Extra Trees. Historically retrained data was utilized in conjunction with R^2 score alongside Mean Absolute Error (MAE) for evaluation. Random forest proved to be the most accurate for deployment due to having the most R^2 scores and the lowest MAE.

Random Forest

```
step1 = ColumnTransformer(transformers=[
    ('col_tnf',OneHotEncoder(sparse_output=False,drop='first'),[0,1,7,10,11])
],remainder='passthrough')

step2 = RandomForestRegressor(n_estimators=290,
                              random_state=3,
                              max_samples=0.85,
                              max_features=0.35,
                              max_depth=30)

pipe = Pipeline([
    ('step1',step1),
    ('step2',step2)
])

pipe.fit(X_train,y_train)

y_pred = pipe.predict(X_test)

print('R2 score',r2_score(y_test,y_pred))
print('MAE',mean_absolute_error(y_test,y_pred))
```

```
R2 score 0.901127388722254
MAE 0.14983570578931865
```

Decision Tree

```
step1 = ColumnTransformer(transformers=[
    ('col_tnf',OneHotEncoder(sparse_output=False,drop='first'),[0,1,7,10,11])
],remainder='passthrough')

step2 = DecisionTreeRegressor(max_depth=8)

pipe = Pipeline([
    ('step1',step1),
    ('step2',step2)
])

pipe.fit(X_train,y_train)

y_pred = pipe.predict(X_test)

print('R2 score',r2_score(y_test,y_pred))
print('MAE',mean_absolute_error(y_test,y_pred))
```

```
R2 score 0.8276821607938187
MAE 0.18735161150763366
```

5. Building Website

To enhance user experience, we crafted a web application that allows consumers to input laptop details and receive estimated prices instantly. The user interface was crafted from Streamlit while backend functionalities were carried out by Python. Being able to instantly predict price aids in decision making to both retailers and consumers.

Fig. 6: Website

6. Deployment

We deployed the web application on Heroku. The website was linked directly to the trained model through an API for quick and convenient predictions. To ensure the application is dependable for price estimation and strategy optimization, routine accuracy and performance checks are executed.

```python
import streamlit as st
import pickle
import numpy as np
import pandas as pd


# import the model
pipe = pickle.load(open('pipe.pkl','rb'))
df = pickle.load(open('df.pkl','rb'))

st.title("Laptop Price Predictor")

# brand
company = st.selectbox('Brand',df['Company'].unique())

# type of laptop
type = st.selectbox('Type',df['TypeName'].unique())
```

4. Results and Conclusion

Linear Regression, Ridge Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree, Random Forest, Gradient Boosting, XGBoost were among the machine learning models we tested for laptop price prediction. With an R2 score of 0.90 and a Mean Absolute Error (MAE) of 0.14, Random Forest outperformed the others, making it the most accurate model for laptop price prediction. Multiple Linear Regression obtained an R2 score of 0.80, while XGBoost and Gradient Boosting came in second and third with R2 ratings of 0.87 and 0.88, respectively. These findings demonstrate how well ensemble models handle intricate interactions seen in the data.

Fig. 7: Prediction result

The study designed a laptop price predictor that utilized machine learning algorithms, and Random Forest had the best accuracy (R2 = 0.90, MAE = 0.14). For real-time predictions, a web application built with Streamlit using Python in its backend. The approach assists retailers in optimizing pricing strategies and consumers in making well-informed decisions. Accuracy could be improved in future research using deep learning methods and real-time data.

5. Future Scope

Demand-based pricing, time-series forecasting, and real-time market trends can all be incorporated into the laptop price prediction model to increase its accuracy. Additionally, it can provide tailored suggestions according to user preferences, assisting customers in selecting the finest laptop within their price range. The user experience and accessibility can be improved by using AI-powered chatbots and extending the model into a mobile app. The model can also be used to forecast the costs of other electronic gadgets, such as tablets and smartphones. Future enhancements could make it a useful tool for people worldwide in the consumer electronics market, such as bilingual support and region-specific pricing.

6. Acknowledgment

The authors wish to extend their genuine appreciation to prof Dr. Lokendra Singh for the guidance and continuous support which they received during the duration of this research. Also special thanks to Medicaps University for granting access to the resources and data required for the completion of this research.

References

1. R. J. Kafabihi, “Predicting Laptop Prices Based on Specifications Using Machine Learning Techniques: An Empirical Study,” IEESE International Journal of Science and Technology (IJSTE), vol. 13, no. 1, pp. 14–22, 2024.
2. B. Prajna, A. L. Sri, A. Sahiti, A. B. Teja, and A. Arzoo, “LAPTOP PRICE PREDICTION USING MACHINE LEARNING,” 2023. [Online]. Available: www.ijnrd.org
3. P. Tian, “Research On Laptop Price Predictive Model Based on Linear Regression, Random Forest and Xgboost,” 2024.
4. L. Akkouh, “Machine Learning-Based Price Prediction for Laptops.” [Online]. Available: https://www.researchgate.net/publication/377721653
5. P. Manegopale, K. Nerpagar, S. Sawant, M. Shinde, and K. Chandarkar, “Laptop Prediction & Comparison using Machine Learning,” 2023.

6. Y. Karakuş and T. T. Bilgin, "International Journal of Multidisciplinary Studies and Innovative Technologies Laptop Price Range Prediction with Machine Learning Methods," ijmsit Research Article, vol. 2022, no. 1, pp. 40–45, doi: 10.36287/ijmsit.8.1.7.
7. V. Surjuse, S. Lohakare, A. Barapatre, and A. Chapke, "Laptop Price Prediction using Machine Learning," International Journal of Computer Science and Mobile Computing, vol. 11, no. 1, pp. 164–168, Jan. 2022, doi: 10.47760/ijcsmc.2022.v11i01.021.
8. K. S. Priya, "Linear Regression Algorithm in Machine Learning through MATLAB," Int J Res Appl Sci Eng Technol, vol. 9, no. 12, pp. 989–995, Dec. 2021, doi: 10.22214/ijraset.2021.39410.
9. L. Breiman, "Random Forests," 2001.
10. Y. C. Chen, H. C. Su, S. M. Huang, C. H. Yu, J. H. Chang, and Y. L. Chiu, "Immune cell profiles and predictive modeling in osteoporotic vertebral fractures using XGBoost machine learning algorithms," BioData Min, vol. 18, no. 1, Dec. 2025, doi: 10.1186/s13040-025-00427-y.