Laptop Price Prediction Using Machine Learning

Tapas Mandal

july ,2025

Abstract

This project aims to predict laptop prices based on their specifications using various machine learning models. The dataset was cleaned, transformed, and enriched with engineered features such as display resolution (PPI), storage types (HDD, SSD), and operating system categories. Multiple regression models were tested, including ensemble methods, to find the best predictor of laptop prices. Results indicate that tree-based ensemble models performed the best, achieving high accuracy with low error.

1 Introduction

Laptop prices vary significantly depending on specifications such as RAM, storage type, display quality, processor, GPU, and operating system. The goal of this project is to build a machine learning model capable of predicting laptop prices based on these features. Accurate prediction can help consumers, manufacturers, and retailers in decision-making.

2 Dataset and Preprocessing

The dataset `laptop_data.csv` was used. Initial preprocessing steps included:

- Removing unnecessary columns (`Unnamed: 0`).
- Converting RAM from string format (e.g., "8GB") to integer values.
- Converting Weight from string format (e.g., "2.2kg") to float values.
- Extracting resolution (x, y) from screen specifications and calculating Pixels Per Inch (PPI).
- Cleaning and splitting memory column into HDD, SSD, Hybrid, and Flash storage.

3 Feature Engineering

Several new features were created:

- **Touchscreen**: Binary variable (1 if present).
- **IPS Panel**: Binary variable (1 if present).
- **PPI**: Derived from resolution and screen size.
- **CPU Brand**: Categorized into Intel Core i3/i5/i7, other Intel processors, and AMD processors.
- **GPU Brand**: Extracted from GPU column (e.g., Nvidia, AMD).
- **Operating System (OS)**: Categorized into Windows, Mac, and Others/Linux/No OS.

4 Exploratory Data Analysis

Several visualizations were generated using Seaborn and Matplotlib:

- Distribution of laptop prices.
- Price comparisons across companies and product types.
- Scatter plots of price vs. inches and price vs. weight.
- Correlation heatmap for numeric features.
- Log transformation of price to stabilize variance.

5 Modeling

The target variable was the natural logarithm of Price. Data was split into 85% training and 15% testing. Categorical variables were encoded using OneHotEncoder within a pipeline.

The following models were trained and evaluated:

- Linear Regression, Ridge, Lasso
- K-Nearest Neighbors (KNN)
- Decision Tree
- Support Vector Regressor (SVR)
- Random Forest
- Extra Trees
- Gradient Boosting
- XGBoost
- Ensemble Models: Voting Regressor and Stacking Regressor

6 Results

Model performance was evaluated using R^2 score and Mean Absolute Error (MAE). The results are summarized in Table 1.

7 Conclusion

This project demonstrates the effectiveness of machine learning for predicting laptop prices. Key insights:

- RAM, PPI, and SSD storage strongly influence price.
- Touchscreen and IPS panels add significant value.
- CPU and GPU brands are important predictors.
- Tree-based ensemble models (Random Forest, Extra Trees, Gradient Boosting, XGBoost) outperform linear models.
- Stacking Regressor achieved the best performance with $R^2 = 0.90$ and MAE = 0.17.

The final trained pipeline and processed dataset were saved as `pipe.pkl` and `df.pkl` for deployment.

Model	R^2 **Score**	**MAE**
Linear Regression	0.80	0.21
Ridge Regression	0.80	0.20
Lasso Regression	0.80	0.21
K-Nearest Neighbors	0.80	0.19
Decision Tree	0.84	0.18
Support Vector Regr	0.80	0.20
Random Forest	0.87	0.15
Extra Trees	0.88	0.15
Gradient Boosting	0.87	0.17
XGBoost	0.84	0.16
Voting Regressor	0.88	0.15
Stacking Regressor	0.90	0.15

Table 1: Comparison of regression models for laptop price prediction.