LAPTOP PRICE PREDICTION

MACHINE LEARNING

(UML501)

Submitted by:

NAME	ROLLNO
Dushar Khatri	102217166

GROUP: – 3CS5

SUBMITTED TO: - Dr. Sachin Kansal

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY, PATIALA

NOV 2024

Introduction

In a rapidly evolving tech market, the prices of laptops can be confusing for various factors that influence those costs. Brand, processor, RAM, and screen resolution are all very influential factors and are very technical in navigation, making it tough for the average buyer to differentiate between these aspects. Students especially require laptops which balance between performance and cost because they have to utilize them for coursework, programming, and even graphic design.

This project fulfils this gap by providing a machine learning-based price prediction model that streamlines the buying process. In addition to saving consumers' time and effort, it enables them to have data-driven insights in negotiating a better deal or comparing what is available in the market.

This project focuses on predicting laptop prices using machine learning techniques, with the goal of providing an interactive and reliable tool for consumers. It caters to a variety of use cases, such as students purchasing laptops that fit their academic and technical needs, professionals seeking high-performance machines, and general users aiming to make informed decisions. By inputting specific configurations like RAM, CPU type, and storage, users can estimate a fair price range for their desired laptop model.

Literature Review

Key Research -

Machine learning models like Linear Regression and Random Forest have proven effective in price prediction tasks by handling both categorical and numerical features. These models are widely used in industries such as e-commerce and real estate.

Gap Addressed -

Existing tools for price prediction are often generic and lack the ability to cater to specific needs like laptop configurations. They fail to provide real-time interactivity and practical insights for students, freelancers, or consumers without technical knowledge.

Role of This Project -

This project bridges the gap by creating an interactive tool tailored for laptop price prediction. It uses advanced feature engineering and robust models like Random Forest to deliver accurate predictions, making it highly relevant for students and professionals seeking personalized recommendations.

Methodology

Dataset: -

The dataset used for this project, laptop_data.csv, contains detailed information about various laptop models, including their technical specifications and price. It serves as the foundation for building the machine learning model by providing diverse and relevant features.

Key Features in the Dataset

1. **Numerical Attributes**:
    - **RAM**: The amount of memory (in GB), a crucial factor affecting performance and price.
    - **Weight**: The laptop’s weight (in kg), which can indicate portability and build quality.
    - **Storage (HDD and SSD)**: Separate attributes for different storage types, capturing capacity and the technology used.
2. **Categorical Attributes**:
    - **Brand (Company)**: Identifies the manufacturer, influencing price due to brand reputation and warranty.
    - **TypeName**: The laptop's category (e.g., Ultrabook, Gaming), directly linked to intended use cases.
    - **CPU and GPU**: Information about processors and graphics cards, critical for performance and pricing.
3. **Screen Features**:
    - **ScreenResolution**: A combination of horizontal (X_res) and vertical (Y_res) pixels, which impacts display quality.
    - **Inches**: Screen size, important for usability and pricing.
4. **Operating System**:
    - Captures the laptop's OS, such as Windows, Mac, or Linux, influencing cost and user preference.
5. **Price**:
    - The target variable, representing the laptop’s market price.

Workflow: -

The methodology involved the following key steps:

1. **Preprocessing**: Removing duplicates, handling null values, transforming features (e.g., ppi for screen quality).
2. **Feature Engineering**: Extracted categorical features like Gpu Brand and numeric features like SSD size.
3. **Model Evaluation**: Compared models (e.g., Ridge, Random Forest) using R^2 and MAE.
4. **Deployment**: Built a Streamlit app for end-user interaction.

Results and Analysis

Model Insights

Several machine learning models were trained and evaluated to identify the best-performing algorithm for laptop price prediction.

- Random Forest Regressor emerged as the most accurate model with an R^2 score of approximately 0.85, indicating that 85% of the variability in laptop prices could be explained by the features used in the model.
- The Mean Absolute Error (MAE) was around 0.11 (after applying a logarithmic transformation to the target variable), reflecting minimal deviation between predicted and actual prices.
- Other models like Ridge and Lasso Regression showed competitive performance but lacked the flexibility of Random Forest in handling feature interactions and non-linearity.

Key Features Impacting Price

1. **RAM:**

Strongest positive correlation with price (e.g., higher RAM laptops are significantly costlier).

RAM upgrades enhance performance, especially for tasks like gaming or programming, justifying the price difference.

2. **SSD:**

Laptops with SSDs, particularly larger capacities, showed higher prices compared to those with HDDs.

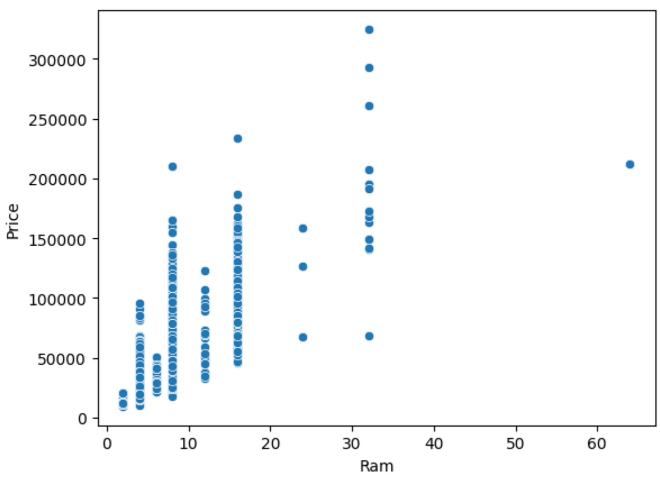
SSDs offer better speed and efficiency, which is reflected in their influence on pricing.

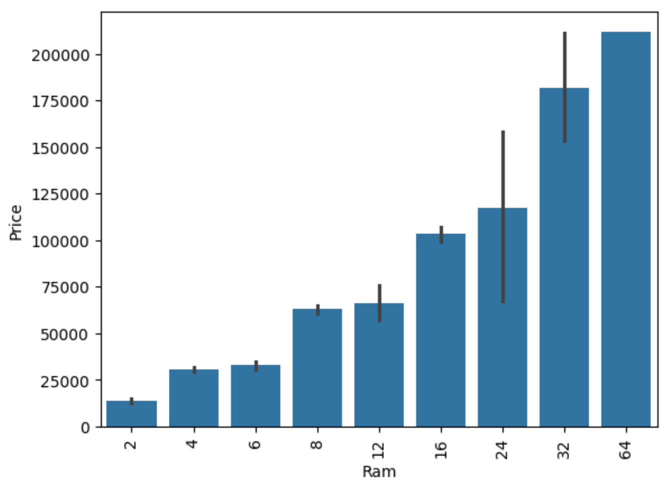
3. **ppi (Pixels Per Inch):**

Derived from screen resolution and size, ppi represents display quality.

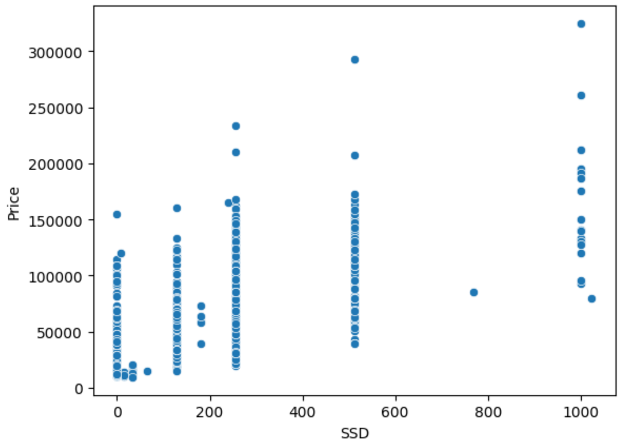
Laptops with higher ppi are priced higher, likely due to better screen clarity and premium design.

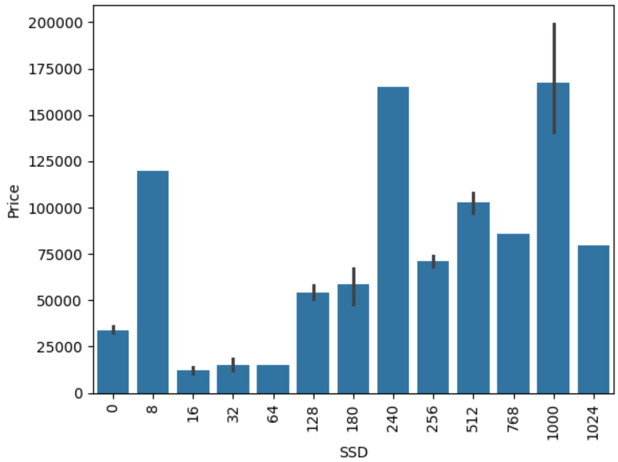
RAM vs Price



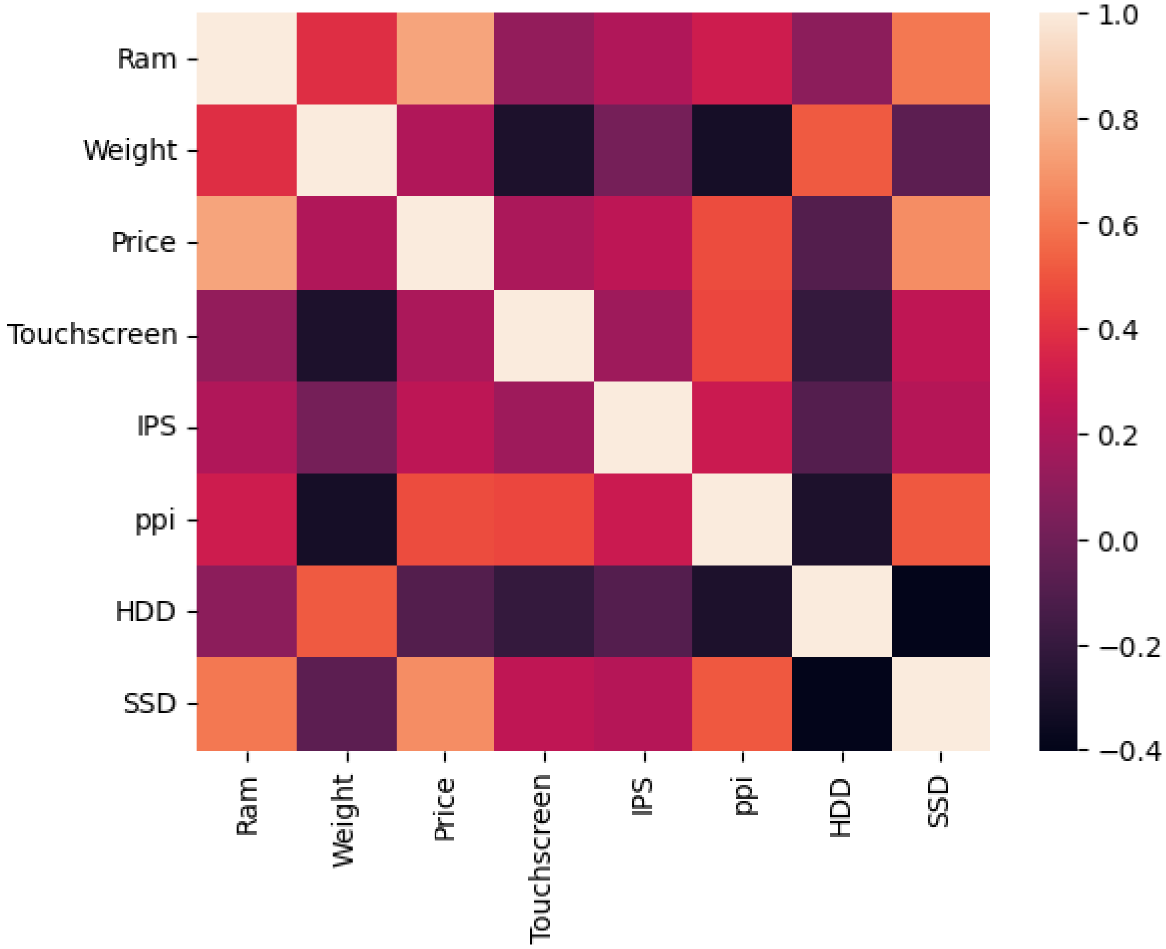


SSD Capacity vs Price





Correlation Heatmap



Conclusion and Recommendations

Conclusion

The project successfully demonstrates the application of machine learning in predicting laptop prices based on technical specifications. Key findings include:

- **Significant Influencers:** Features like RAM, SSD, and ppi (display quality) have the strongest impact on pricing.
- **Model Performance:** Random Forest Regressor outperformed other models, achieving an R^2 score of 0.85, making it reliable for predictions.
- **Practical Utility:** The interactive Streamlit app allows users to input specific configurations and receive accurate price estimates, catering to students, professionals, and general buyers.

Recommendations for Future Work: -

- **Enhance Dataset:** Incorporate more data from diverse sources to improve model generalization and accuracy for niche laptop configurations.
- **Explore Advanced Models:** Experiment with neural networks or ensemble methods to capture more intricate relationships.
- **Expand Features:** Include additional attributes such as battery life, warranty, and customer reviews for more comprehensive predictions.
- **Broader Application:** Adapt the tool for other electronic devices like smartphones or tablets, leveraging similar methodologies.