# Mini Project: Laptop Price Prediction

**Submitted by: Group L**

**Course: CPE221**

**Date: 09/9/2025**

**Introduction**

In today’s technology-driven world, laptops are among the most commonly used electronic devices. However, when purchasing a used laptop, it can be difficult to estimate a fair price because it depends on many factors such as brand, processor, RAM, storage, and graphics card.

The aim of this project is to use machine learning techniques to build a predictive model that estimates the price of used laptops based on their specifications. By analyzing patterns and trends in the dataset, the project provides insights into which features most strongly influence laptop prices, and selects the best model for accurate predictions.

**Data and Exploratory Data Analysis (EDA)**

The dataset consisted of laptop specifications along with their corresponding prices. Before building the model, Exploratory Data Analysis (EDA) was carried out to identify important patterns and relationships.

Key Findings from EDA:

Laptop Type: Gaming laptops and performance-oriented laptops are generally priced higher than ultrabooks or budget laptops.

Processor (CPU): Laptops with newer and more powerful processors (e.g., Intel i7/i9 or AMD Ryzen 7/9) showed significantly higher prices compared to older or entry-level processors.

Memory (RAM): Laptops with larger RAM capacity (8GB, 16GB, or 32GB) were priced higher than those with 4GB or less.

Storage: Laptops equipped with SSD storage had higher prices than those with only HDD storage, even when total storage capacity was similar.

Graphics Card (GPU): Laptops with dedicated GPUs (e.g., NVIDIA GeForce, AMD Radeon) were priced higher than laptops with integrated graphics.

Operating System and Brand: Premium brands such as Apple, Dell, and MSI had higher average prices compared to budget brands.

The overall price distribution was right-skewed, meaning that most laptops fell into a mid-range price category, while a smaller number of high-end gaming or premium laptops drove the upper range.

**Model Training and Testing**

To predict laptop prices, several machine learning models were trained and compared. The dataset was divided into training and testing sets to ensure that the models could be evaluated fairly.

Models Used:

Linear Regression: A simple baseline model assuming a straight-line relationship between features and price.

Decision Tree Regressor: A tree-based model that splits the dataset into smaller groups to capture non-linear relationships.

Random Forest Regressor: An ensemble method combining multiple decision trees to improve accuracy and reduce overfitting.

Gradient Boosting Regressor (e.g., XGBoost): A boosting method that builds models sequentially, each correcting the errors of the previous one.

**Observations During Training:**

Linear Regression: Performed poorly because laptop prices depend on non-linear interactions between specifications.

Decision Tree: Improved predictions but was prone to overfitting, performing well on training data but less effectively on test data.

Random Forest: Provided a good balance, capturing non-linear patterns while maintaining generalization to unseen data.

Gradient Boosting: Delivered strong results by reducing errors gradually and was comparable to Random Forest in accuracy.

**Results**

After training and evaluating all models, the following results were obtained:

Linear Regression: Lowest accuracy and highest error values.

Decision Tree: Better than linear regression but suffered from overfitting.

Random Forest and Gradient Boosting: Achieved the best performance with higher R² scores and lower error values.

The Random Forest Regressor was chosen as the final model because it provided stable and accurate predictions. It achieved an R² score of approximately XX% on the test set, indicating that the model could explain most of the variation in laptop prices.

**Conclusion**

This project demonstrated how machine learning can be used effectively to predict the prices of used laptops. Through data analysis and model comparison, the study highlighted that laptop specifications such as processor type, RAM, storage, and GPU are the most important factors influencing prices.

The Random Forest model proved to be the most reliable predictor, outperforming simpler models like Linear Regression and avoiding overfitting issues faced by Decision Trees.

Recommendations for Future Work:

Hyperparameter Tuning: Fine-tuning the model parameters to further improve performance.

Larger Dataset: Expanding the dataset to include more brands, models, and recent laptops for better generalization.

Advanced Models: Testing deep learning techniques (e.g., neural networks) to capture more complex relationships.

In conclusion, the project not only produced a working price prediction model but also provided useful insights into which laptop features most strongly influence price.