Laptop Price Prediction Using Machine Learning

Problem Statement:

Laptop prices vary significantly based on hardware specifications, brand, and other features. Choosing a laptop within budget that meets user requirements can be confusing. This project aims to predict laptop prices based on various attributes—RAM, CPU, GPU, brand, screen size, resolution, and more—using supervised machine learning techniques.

Objective:

To build a full-stack machine learning solution that:

1. Accurately predicts the price of a laptop based on user input features.

2. Emphasizes **data preprocessing and EDA**, making it ideal for data science aspirants to strengthen their core concepts.

3. Deploys a web interface for real-time price prediction.

Domain:

E-commerce / Retail / Machine Learning / Web Development

Dataset:

- Laptop specifications dataset scraped from e-commerce websites like **Amazon** and **Flipkart**.

- Features include:

- **Brand**
- **Type (Notebook, Gaming, etc.)**
- **RAM**
- **Operating System**
- **Weight**
- **Touchscreen (Yes/No)**
- **IPS Display (Yes/No)**
- **Screen Size**
- **Screen Resolution**
- **CPU**
- **HDD / SSD**
- **GPU**
- **Price** (Target variable)

Tools & Technologies Used:

- **Languages**: Python, HTML, CSS

- **Libraries**: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Pickle
- **ML Model**: Gradient Boosting / Random Forest (based on experimentation)
- **Web Framework**: Streamlit or Flask (if simple), or Django
- **Deployment**: Heroku + GitHub

Project Workflow:

1. Data Loading & Preprocessing

- Convert categorical variables using label encoding or one-hot encoding
- Extract numerical features from strings (e.g., screen resolution)
- Convert TB/GB into unified numeric values
- Clean and standardize weight, RAM, and storage fields
- Deal with outliers and null values

2. Exploratory Data Analysis (EDA)

- Distribution analysis for each feature
- Correlation matrix to understand relationships
- Visualization of price trends by brand, specs, and other features

- Feature importance analysis

💡 **Learning Highlight**: The project teaches how **good preprocessing** can significantly improve model performance—even with relatively simple models.

3. Model Building

- Model selection through experimentation (Random Forest, XGBoost, etc.)
- Hyperparameter tuning using GridSearchCV
- Performance evaluation using R^2 Score, MAE, RMSE

4. Model Export

- Save trained model using Pickle
- Prepare a clean pipeline for input processing and prediction

5. Web App Development

- Build an interactive UI for users to input laptop specs
- Display predicted price in real-time
- Integrate the ML model with the UI backend

6. Deployment on Heroku

- Push project to GitHub

- Configure `Procfile`, `requirements.txt`, and necessary buildpacks

- Deploy to Heroku for public access

Key Takeaways:

- Emphasis on **data cleaning, transformation, and feature engineering**

- Hands-on with **EDA techniques** that form the foundation of data science

- End-to-end full-stack ML project from model training to deployment

- Real-world applications mimicking actual laptop price estimators on e-commerce platforms

Conclusion:

This complete supervised machine learning project provides great exposure to data preprocessing, model tuning, and full-stack deployment. It mimics real-world use cases and is highly recommended for beginners aiming to build their data science foundation, especially in understanding the power of EDA and clean data.