Laptop Price Analysis

Problem Statement:

Laptop prices depend on multiple specifications such as **brand, processor, RAM, GPU, storage, and screen quality.**
For a buyer, it is hard to estimate the fair price of a laptop.
Our goal is to predict laptop prices based on their specifications using Machine Learning Regression models.

Dataset Overview

- The dataset contains details of laptops such as Company, Type, Inches, Screen Resolution, CPU, RAM, Storage, GPU, Operating System, Weight, and Price.
- Target Variable: Price (continuous).
- Shape: ~1300 rows × 12 columns.

Key Features of Dataset

- **Company** → Brand of the laptop (e.g., Apple, Dell, HP, Lenovo).
- **TypeName** → Category of laptop (Notebook, Ultrabook, Gaming, Netbook, etc.).
- **Inches** → Screen size (in inches).
- **ScreenResolution** → Display resolution (e.g.1920x1080) and screen type (IPS, Retina, Full HD, etc.).
- **Cpu** → Processor details (brand, series, and speed).
- **Ram** → Memory size (in GB).
- **Memory** → Storage type and size (HDD, SSD, or hybrid).
- **GPU** → Graphics card details (integrated or dedicated GPU).
- **OpSys** → Operating System (Windows, macOS, Linux, No OS, etc.).
- **Weight** → Weight of the laptop (in kilograms).
- **Price** → Target variable, laptop price (in ₹).

Preprocessing Steps:

1.**Feature Engineering**

- Extracted CPU brand, frequency, and model.
- Extracted GPU brand.
- Converted ScreenResolution into width, height, PPI, IPS flag, Touchscreen flag.
- Split Memory into HDD size, SSD size.
- Categorized OS into Windows / Mac / Others.

2.Cleaning

- Removed 28 duplicate rows.
- Converted RAM/Storage into numerical GB values.
- Converted categorical features with OneHotEncoding.
- Kept numeric columns as-is.

3.Target Transformation

- Applied log transformation on Price to handle skewness.

4. Modeling Approach:

We tested multiple machine learning models:

A. Individual Models

- **Lasso Regression** → R^2 = 0.80, MAE = 0.21
- **KNN Regressor** (k=3) → R^2 = 0.79, MAE = 0.19
- **Decision Tree** (depth=8) → R^2 = 0.83, MAE = 0.19
- **SVR (RBF kernel)** → R^2 = 0.81, MAE = 0.20
- **Random Forest (tuned)** → R^2 = 0.88, MAE = 0.16

❖ **Random Forest** gave the best performance among single models.

B. Final Model – Stacking Regressor:

We combined multiple models for better accuracy:

Base Models:

- Random Forest Regressor
- Gradient Boosting Regressor
- XGBoost Regressor

Results (Stacking):

- R^2 Score: 0.88
- MAE: 0.16

❖ Stacking was chosen because it combines strengths of all models, leading to robust and accurate predictions.

Deployment

- Saved processed dataset → `df.pkl`.
- Saved final trained pipeline (Stacking Regressor) → `pipe.pkl`.
- Built an interactive Streamlit app where users can input specifications:
  - Brand, Type, RAM, Storage, CPU, GPU, Screen Size, Resolution, OS, etc.
  - App automatically calculates features (PPI, Touchscreen, IPS).
  - Displays Predicted Laptop Price in INR

Results & Conclusion:

- Achieved $R^2 = 0.88$ and MAE ≈ 0.16, meaning predictions are quite close to actual prices.
- Stacking Regressor outperformed individual models like Lasso, KNN, Decision Tree, and SVR.
- Deployment with Streamlit makes it easy for users to predict laptop prices instantly.