LAPTOP PRICE PREDICTION

Problem Statement: The goal of this project is to predict the price of laptops based on various attributes such as brand, type, specifications (RAM, screen size, resolution, etc.), and more. Accurate price prediction can help consumers make informed decisions and assist manufacturers in pricing their products competitively. It can also be valuable for e-commerce platforms, providing users with real-time price estimates based on product features.

Objectives:

- To develop a regression model that can accurately predict laptop prices based on available features.
- To explore the relationship between different laptop specifications (e.g., RAM, screen size, processor brand, etc.) and their price.
- To evaluate different machine learning algorithms and select the one that offers the best performance in terms of prediction accuracy.
- To prepare the data for modelling by addressing issues such as missing values, data normalization, and feature selection

Questions:

- What are the most important features affecting laptop prices?
- How does RAM, processor type, GPU brand, screen size, and other features influence the price of a laptop?
- Are there any non-linear relationships between the features and the target variable (price)?
- Which machine learning model (Random Forest, Linear Regression, etc.) provides the best prediction accuracy for this problem?
- Can feature engineering (such as calculating PPI from screen resolution and size) improve model performance?

Relevance: These questions aim to identify key features impacting laptop prices, the best modelling techniques, and how to fine-tune the model for the most accurate predictions. They are central to understanding the behaviour of the dataset, performing exploratory data analysis (EDA), and selecting the optimal machine learning model.

Data:

“The dataset contains details about the specifications and pricing of various laptop models. It is used to predict the target variable, which is the laptop price. The data is diverse, covering a range of laptop types, brands, and configurations, making it a good candidate for machine learning modelling”

Data Wrangling and Cleaning Steps:

- **Data Type Conversion**: Certain features like Price, RAM, and Weight need to be converted to numerical data types for modelling. Categorical variables like Company, Type, etc., need encoding.
- **Feature Transformation:** Non-numeric columns (e.g., Company, CPU Brand) are transformed using encoding techniques such as One-Hot Encoding
- **Feature Engineering:** Derived new features such as PPI (pixels per inch) from screen resolution and screen size to help improve the model's predictive power

Insights from Exploratory Data Analysis (EDA):

- **Feature Correlations:** There is a strong correlation between RAM and Price, and between Storage (HDD/SSD) and Price. This indicates that higher RAM and larger storage capacities are likely to increase the price of a laptop
- **Price Distribution:** The price distribution is highly skewed, with most laptops being priced between a lower and middle range, with fewer high-end models.
- **Impact of Touchscreen and IPS:** Laptops with a touchscreen and an IPS display tend to be more expensive. This indicates that these features add significant value to the laptop price
- **Company/Brand Influence:** Certain brands, like Apple, tend to have higher-priced laptops across various configurations.
- **Screen Resolution Impact:** Higher screen resolution and the inclusion of features like IPS or Touchscreen also appear to contribute to higher laptop prices.
- **Histograms and Boxplots** of features like Price, RAM, and Weight to observe their distributions.
- **Correlation Matrix** to visualize which features correlate most with Price.

Model Choice, Performance Metrics, and Interpretation:

Model Choice: After testing multiple regression models, including Linear Regression, K-Nearest Neighbors, and Decision Trees, **Random Forest Regressor** was chosen as the final model. It provides strong performance due to its ability to capture complex relationships in data and handle non-linearities effectively.

Performance Metrics:

- **R^2 (R-Squared):** The Random Forest model achieves an R^2 score of 0.88, indicating that 88% of the variance in laptop prices is explained by the model. This is a strong result for a regression task.
- **MAE (Mean Absolute Error):** The MAE value of 0.15 shows that, on average, the model's price predictions are within 15% of the true prices.
- **Interpretation:** These results suggest that the Random Forest model is accurate and generalizes well on unseen data. However, the model could be improved further by tuning hyperparameters or using ensemble techniques like boosting.