Laptop Price Prediction

Presented by: Yash Sahu

Table of Content

Introduction & Problem Statement

Introduction

- Laptop prices vary based on multiple factors like brand, specifications, and market trends.
- Predicting laptop prices helps buyers make informed decisions and assists businesses in pricing strategies.

Problem Statement

- Manually determining a laptop's fair price is complex due to diverse specifications.
- The goal is to build a machine learning model that predicts laptop prices accurately based on key features.

Dataset Description

Overview

- The dataset contains various features influencing laptop prices.
- Key attributes include Brand, Processor, RAM, Storage, Screen Size, GPU, Operating System, and Price.

Data Insights

- There are total 1275 rows and 23 columns.
- There is no any null values in dataset.

Dataset Description

```
df.head()
```

	Company	Product	TypeName	Inches	Ram	OS	Weight	Price_euros	Screen	ScreenW	...	RetinaDisplay	CPU_company	CPU_freq	CPU_model	PrimaryStorage	SecondaryStorage	PrimaryStorageType	SecondaryStorageType	GPU_company	GPU_model
0	Apple	MacBook Pro	Ultrabook	13.3	8	macOS	1.37	1339.69	Standard	2560	...	Yes	Intel	2.3	Core i5	128	0	SSD	No	Intel	Iris Plus Graphics 640
1	Apple	Macbook Air	Ultrabook	13.3	8	macOS	1.34	898.94	Standard	1440	...	No	Intel	1.8	Core i5	128	0	Flash Storage	No	Intel	HD Graphics 6000
2	HP	250 G6	Notebook	15.6	8	No OS	1.86	575.00	Full HD	1920	...	No	Intel	2.5	Core i5 7200U	256	0	SSD	No	Intel	HD Graphics 620
3	Apple	MacBook Pro	Ultrabook	15.4	16	macOS	1.83	2537.45	Standard	2880	...	Yes	Intel	2.7	Core i7	512	0	SSD	No	AMD	Radeon Pro 455
4	Apple	MacBook Pro	Ultrabook	13.3	8	macOS	1.37	1803.60	Standard	2560	...	Yes	Intel	3.1	Core i5	256	0	SSD	No	Intel	Iris Plus Graphics 650

5 rows × 23 columns

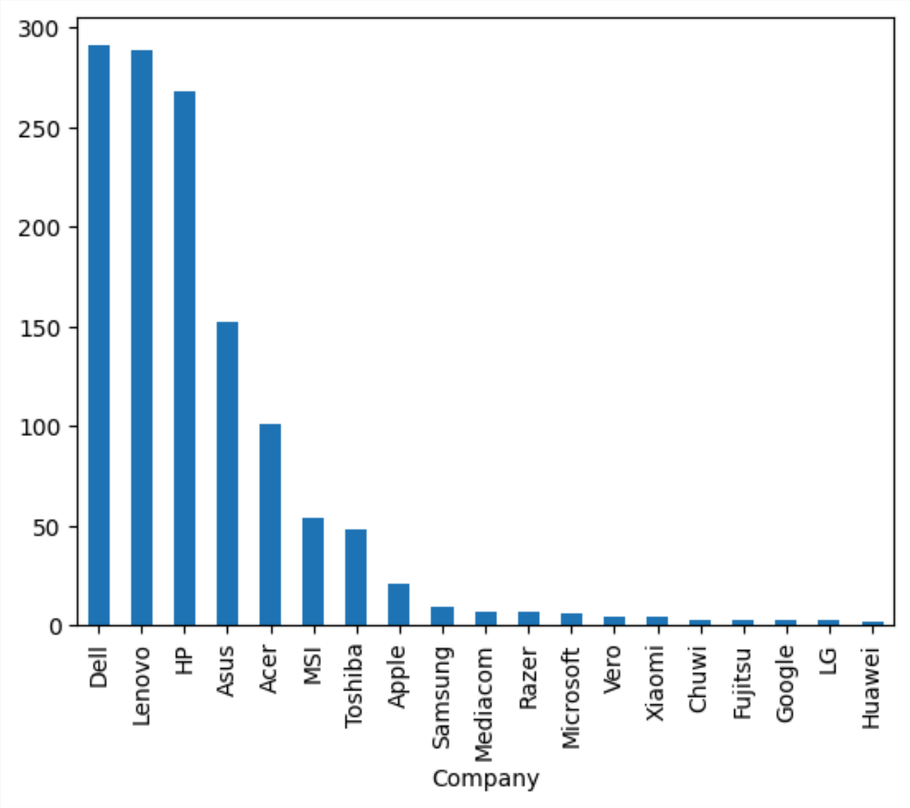
This is the output of df.head() first five rows of dataset.

```
[ ] df.columns
```

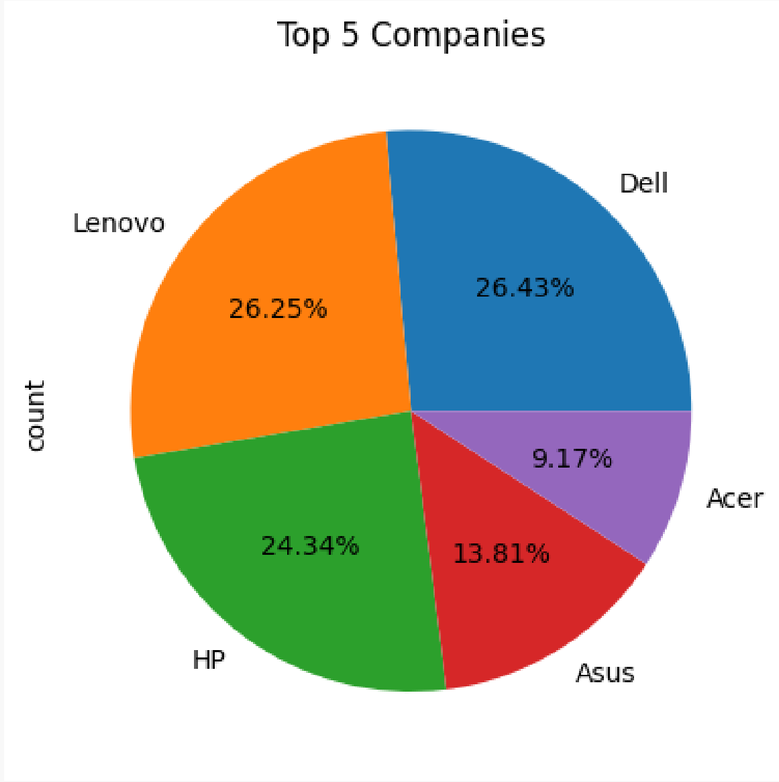
```
Index(['Company', 'Product', 'TypeName', 'Inches', 'Ram', 'OS', 'Weight',
       'Price_euros', 'Screen', 'ScreenW', 'ScreenH', 'Touchscreen',
       'IPSpanel', 'RetinaDisplay', 'CPU_company', 'CPU_freq', 'CPU_model',
       'PrimaryStorage', 'SecondaryStorage', 'PrimaryStorageType',
       'SecondaryStorageType', 'GPU_company', 'GPU_model'],
      dtype='object')
```

List of columns in dataset.

Data Visualization

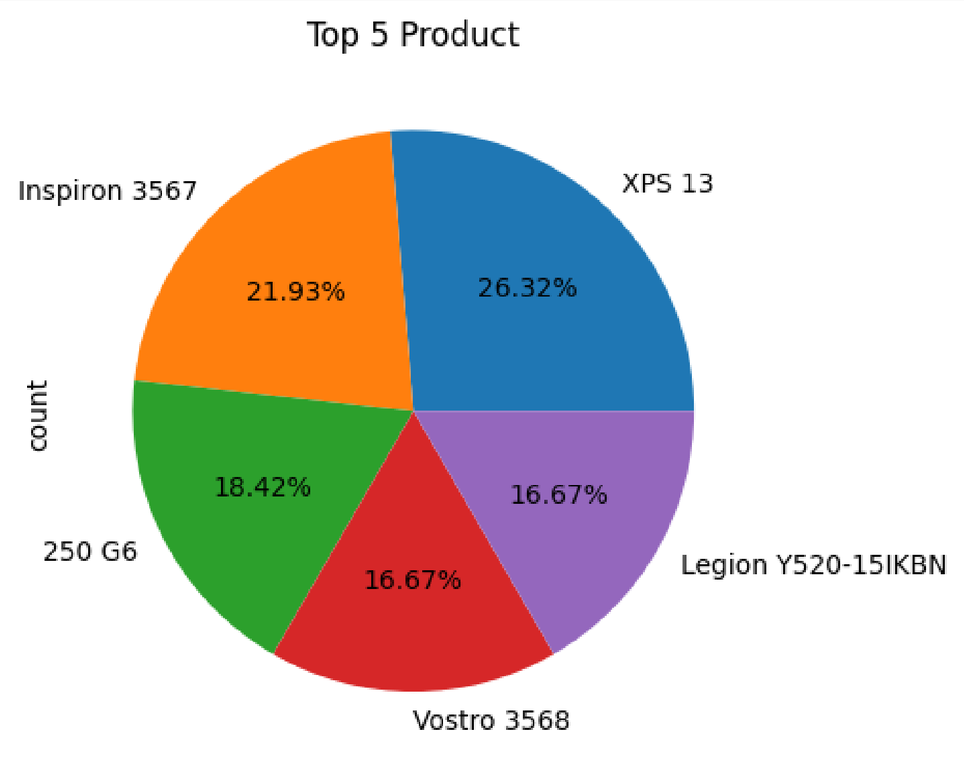


Laptop Company Distribution

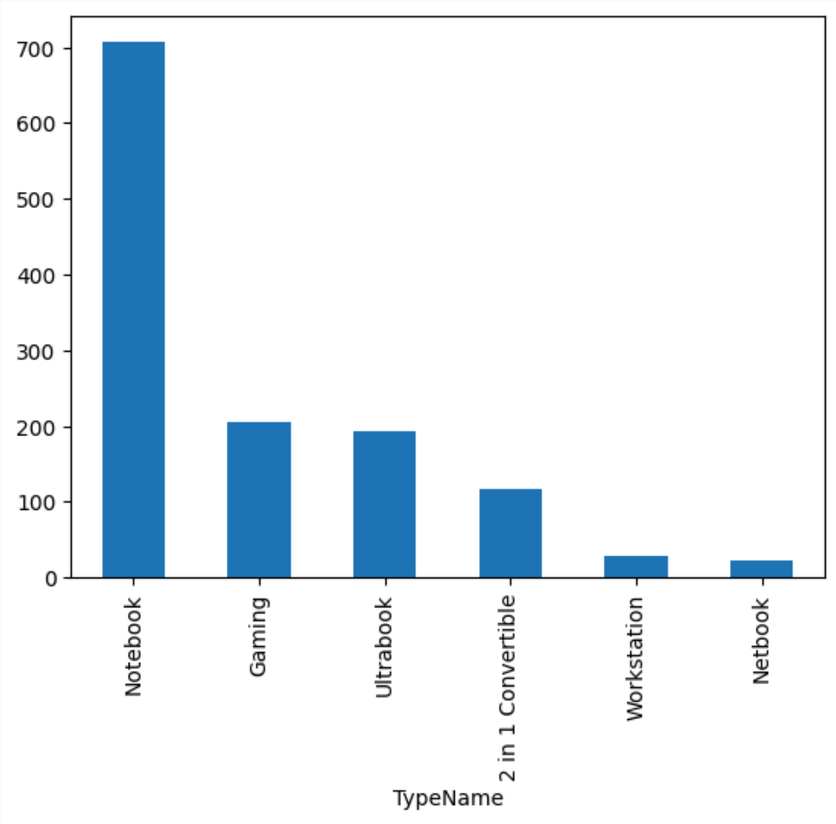


Top 5 Companies

Data Visualization

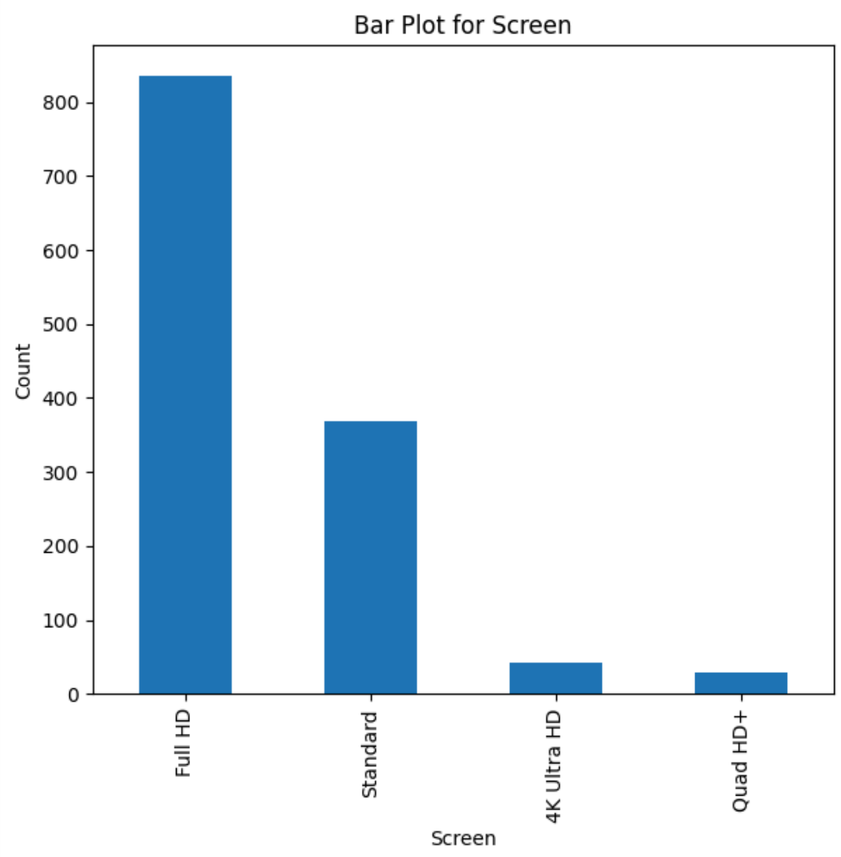


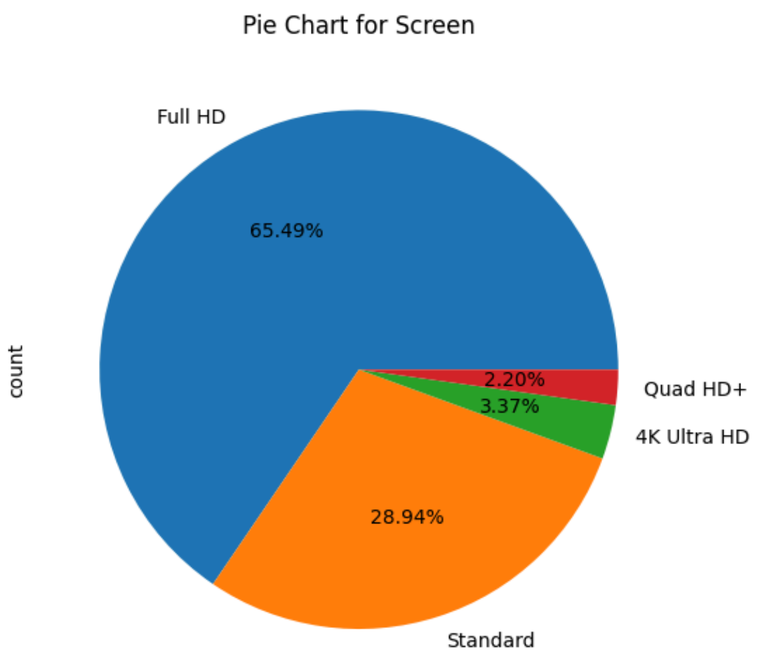
Top 5 Products



Barchart for Laptop Type

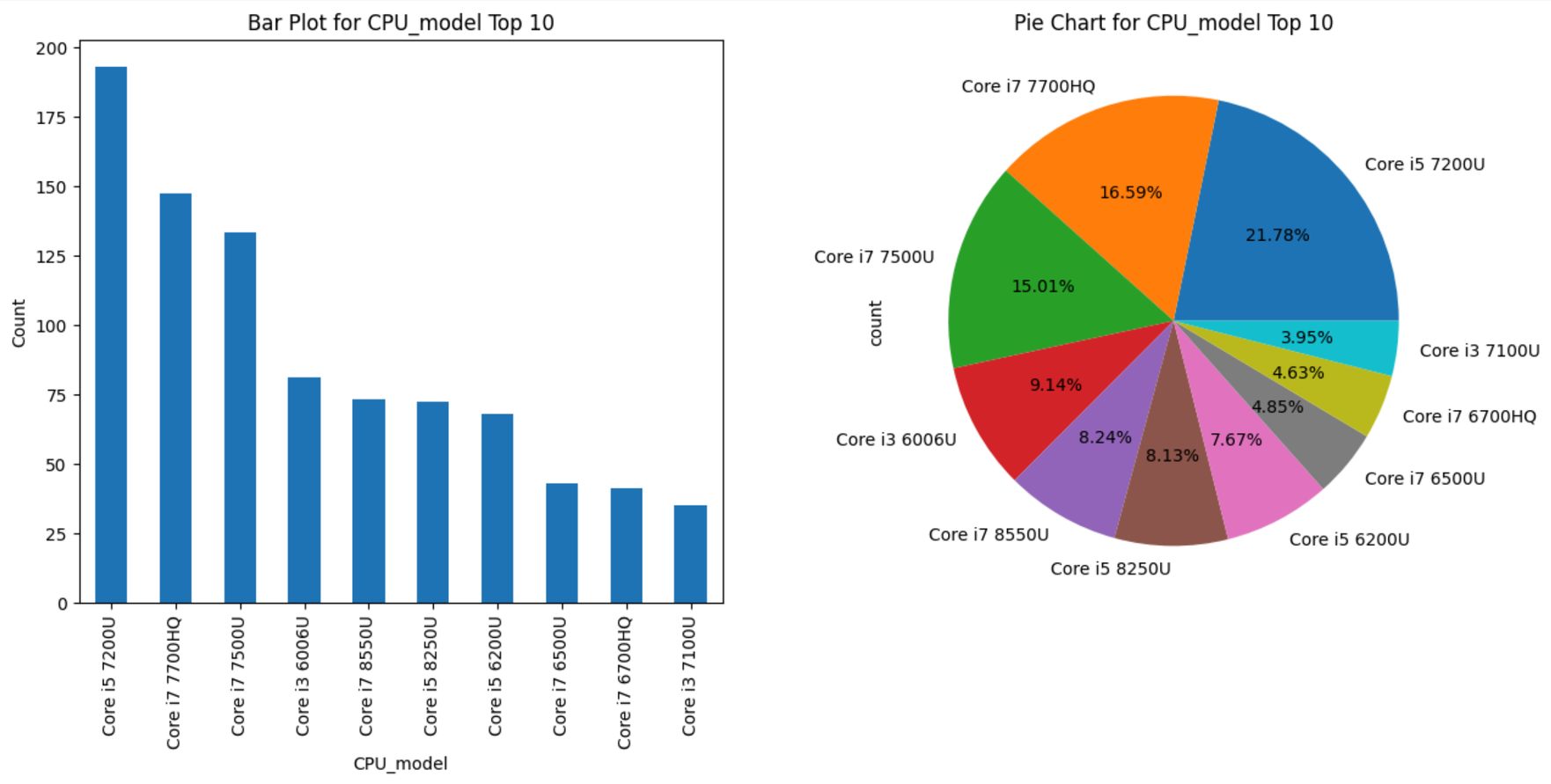
Data Visualization





Plots for Screen feature

Data Visualization



Plots for CPU Model feature

Techniques Used

Introduction

experimented with multiple regression algorithms to predict laptop prices and evaluated their performance. The models tested include:

- Linear Regression – Basic model, but it struggled with complex relationships.
- Decision Tree Regressor – Performed better but prone to overfitting.
- Random Forest Regressor – Achieved the best performance.
- XGBoost Regressor – Showed good results but slightly lower than Random Forest.

Best Model Performance:

- After testing, Random Forest Regressor provided the highest accuracy:
- **R^2 Score on Training Data: 0.98**
- **R^2 Score on Test Data: 0.87**

Techniques Used

Model Evaluation Metrics

- R^2 Score – Measures model accuracy.
- Mean Squared Error (MSE) – Determines prediction errors.

Model Training & Evaluation

Model Training Process:

1. **Data Splitting:**

- Dataset was split into 80% training data and 20% test data

2. **Feature Engineering & Preprocessing:**

- One-Hot Encoding for categorical variables (e.g., Brand, OS).
- Feature Scaling applied where necessary.

3. **Model Training:**

- Tested multiple algorithms (Linear Regression, Decision Tree, XGBoost, etc.).
- Random Forest Regressor performed the best.

Performance Analysis

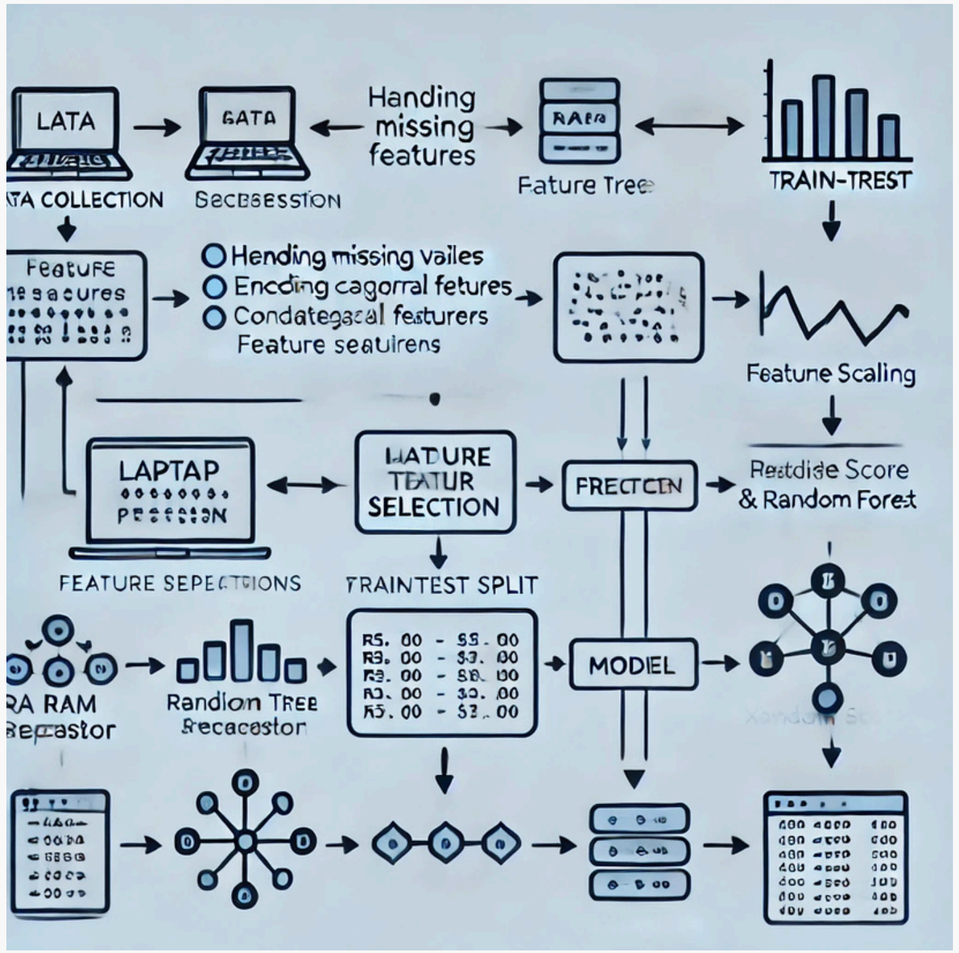
- **High training accuracy (0.98)** suggests a well-fitted model.
- **Test accuracy (0.87)** indicates good generalization.

Model Training & Evaluation

Workflow Steps

1. Data Collection – Gather laptop specifications & price dataset.
2. Data Preprocessing – Handle missing values, apply encoding, and feature scaling.
3. Feature Selection – Choose key features affecting price (RAM, Processor, Storage, etc.).
4. Train-Test Split – Split data into training (80%) and testing (20%).
5. Model Selection & Training – Train multiple models (Linear Regression, Decision Tree, Random Forest, XGBoost).
6. Model Evaluation – Compare R^2 score & MSE, select the best model.
7. Prediction & Output – Use the trained model to predict laptop prices.
8. Deployment (if applicable) – Integrate model into a web app (Flask/Streamlit).

Model Training & Evaluation



Workflow Diagram

Web Application

Web Application

Framework Used: Streamlit

- Developed an interactive web application for laptop price prediction using Streamlit.
- Simple and lightweight UI for real-time predictions.

Workflow of Web App

1. User enters laptop details in the input form.
2. Model processes the input and makes a prediction.
3. Predicted price is displayed instantly.

Output & Results

Web Application

- The **Random Forest Regressor** was the best-performing model.
- **R^2 Score:**
- Training Data: 0.98
- Test Data: 0.87

Observations

- Model predicts prices with high accuracy.
- Some slight variations due to feature importance & dataset limitations.
- Further tuning could improve generalization.

Thank You!