

Laptop
Price
Prediction

Martina
Speciale

Problem
Definition

Dataset
Description

EDA

Data Pre-
processing

Models

Best Model

Results

Deployment

Future Im-
provements

Laptop Price Prediction

Martina Speciale

Data Mining & Machine Learning Project
University of Pisa, Academic Year 2023/2024

Problem Definition

Laptop
Price
Prediction

Martina
Speciale

Problem
Definition

Dataset
Description

EDA

Data Pre-
processing

Models

Best Model

Results

Deployment

Future Im-
provements

- **Objective:** Build a system that predicts laptop prices based on their specifications.
- **Purpose:** Helps consumers make informed purchasing decisions and assists retailers in pricing strategies.

► [Github Repository](#)

► [Streamlit App](#)

Dataset Description

► Data

- **Source:** Dataset available on Kaggle (`laptop_data.csv`)

- **Features:**

- *Company* : Company that manufactured the laptop.
- *TypeName* : type of laptop (e.g. Notebook, Gaming).
- *Inches* : size of the laptop screen measured diagonally.
- *Screen Resolution*: Additional information about the display.
- *Cpu*: Type and speed of the CPU.
- *Ram* : Size of memory (GB).
- *Memory* : Type and capacity of storage (HDD/SSD).
- *Gpu* : specifics about the GPU.
- *OpSys* : Operating System.
- *Weight* : weight in kg.

- **Target Variable:**

- *Price*: laptop price, to be predicted

Exploratory Data Analysis

EDA Highlights:

- Initial analysis of the dataset revealed the distribution of laptop prices and the relationships between different features and prices.
- Transformation techniques were applied to normalize the price data, ensuring better model performance.

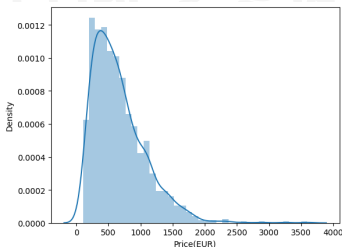


Figure: (before)

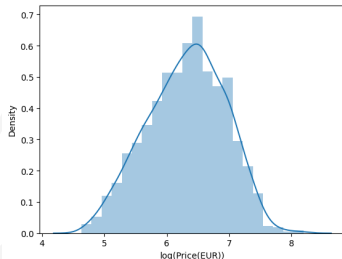
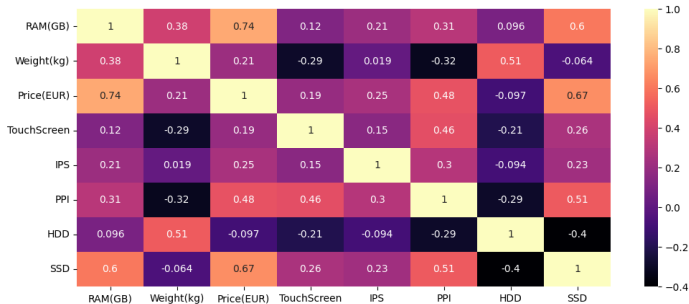


Figure: (after)

Data Preprocessing

Preprocessing Steps:

- Data cleaning
- Feature engineering included creating new features from existing data and encoding categorical variables.
- The final preprocessed dataset provided a robust foundation for model building.



Building the Models

- Several regression models were tested: *Linear Regression, Lasso Regression, Ridge Regression, Decision Tree, Random Forest, and K-Nearest Neighbors*.
- Evaluation Metrics:
 - The models were evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared.
 - These metrics helped in comparing model performances and selecting the best one.

Model	R2 score	MAE	RMSE
Random Forest (with Hyperparameter Tuning)	0.9011	0.1482	0.0378
KNN (with Hyperparameter Tuning)	0.8316	0.1783	0.2538
Linear Regression	0.8068	0.2101	0.0739
Ridge Regression	0.8127	0.2095	0.0716
Lasso Regression	0.8071	0.2116	0.0738
Decision Tree Regressor	0.846	0.1802	0.0589
Random Forest Regressor (Base)	0.8827	0.1622	0.2119

Best Model

Laptop
Price
Prediction

Martina
Speciale

Problem
Definition

Dataset
Description

EDA

Data Pre-
processing

Models

Best Model

Results

Deployment

Future Im-
provements

Random Forest with *Hyperparameter Tuning*:

- **Random Forest** was identified as the best model due to its ability to capture non-linear relationships and reduce overfitting. Key hyperparameters tuned included the number of trees, maximum depth, and minimum samples split, which significantly improved performance.

```
# Using parameters found in grid_search.best_params_ after GridSearchCV
rf_best = RandomForestRegressor(
    bootstrap=False,
    max_depth= 15,
    max_features='sqrt',
    min_samples_leaf=1,
    min_samples_split=5,
    n_estimators=300
)
```

Results

Laptop
Price
Prediction

Martina
Speciale

► Results

Problem
Definition

Dataset
Description

EDA

Data Pre-
processing

Models

Best Model

Results

Deployment

Future Im-
provements

Model Performance:

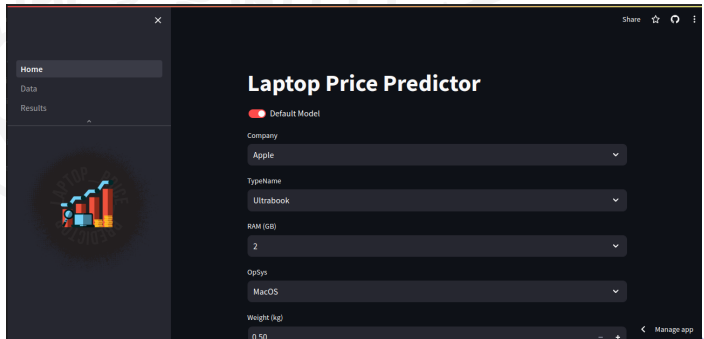
- The **Random Forest** model showed the highest accuracy among all tested models, making it the ideal choice for predicting laptop prices.
- A comparative study between our models and the ones developed in the paper highlighted similar results and confirmed the superior performance of Random Forest.

Deployment on Streamlit Cloud

► App

Deployment Process:

- The model was prepared and deployed using Streamlit Cloud, making it accessible for real-time predictions.
- Example outputs illustrate the prediction capabilities of the deployed model.



The screenshot shows a web application titled "Laptop Price Predictor" with a dark theme. On the left is a sidebar menu with options: Home (selected), Data, and Results. Below the menu is a circular logo featuring a laptop and a bar chart. The main content area has a toggle for "Default Model" (which is turned on). Below this are several input fields with dropdown menus: "Company" (set to "Apple"), "TypeName" (set to "Ultrabook"), "RAM (GB)" (set to "2"), "OpSys" (set to "MacOS"), and "Weight (kg)" (set to "0.50"). At the bottom right, there are icons for "Share", "Manage app", and a settings menu.

Laptop
Price
Prediction

Martina
Speciale

Problem
Definition

Dataset
Description

EDA

Data Pre-
processing

Models

Best Model

Results

Deployment

Future Im-
provements

Future Improvements

Laptop
Price
Prediction

Martina
Speciale

Problem
Definition

Dataset
Description

EDA

Data Pre-
processing

Models

Best Model

Results

Deployment

Future Im-
provements

Future updates may include :

- integration with more recent data
- additional features for prediction and optimization of model performance

Laptop
Price
Prediction

Martina
Speciale

Problem
Definition

Dataset
Description

EDA

Data Pre-
processing

Models

Best Model

Results

Deployment

Future Im-
provements

Thank You!