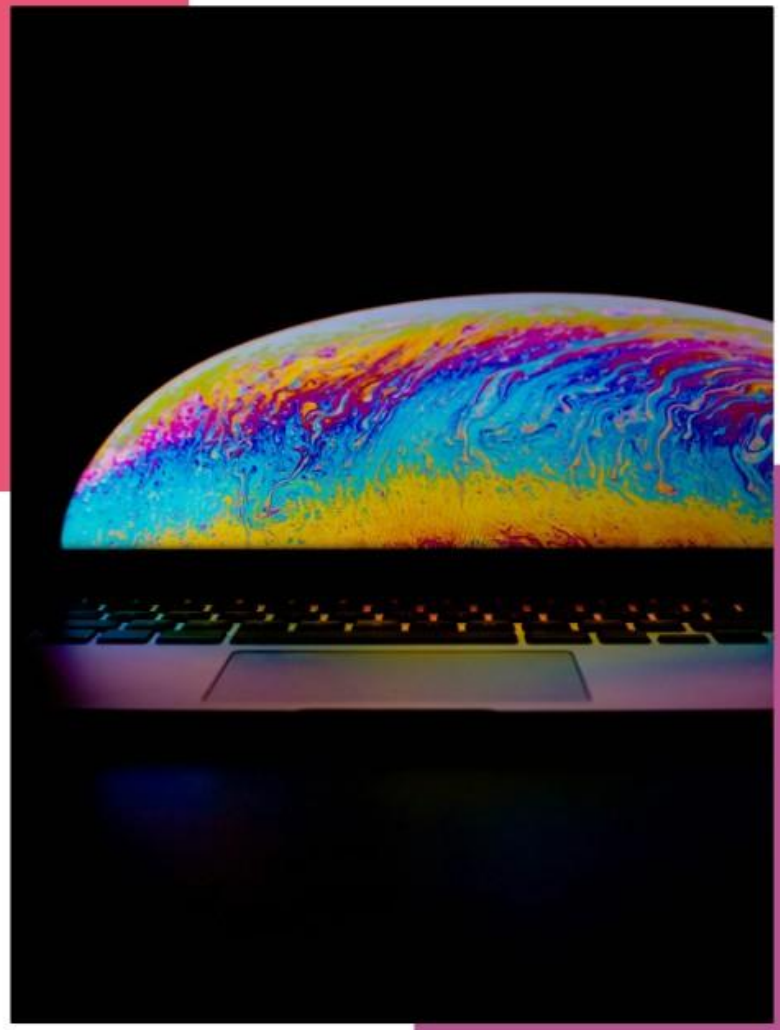


Laptop Price Prediction: A Data-Driven Approach

Developing Predictive Models to Estimate Laptop Prices Through Features Analysis and Machine Learning

Problem Definition

Defining the Problem

Understanding the Factors Influencing Laptop Pricing

Primary Goal

Develop a machine learning model to predict laptop prices based on specifications.

Key Influencing Factors

Price is affected by brand, specifications, and additional features.

Exploration of Relationships

Analyze how various features correlate with laptop prices for better predictions.

Generalization to Unseen Data

Aim to create a model that accurately predicts prices for new, unseen laptops.

Dataset Insights

Dataset Overview

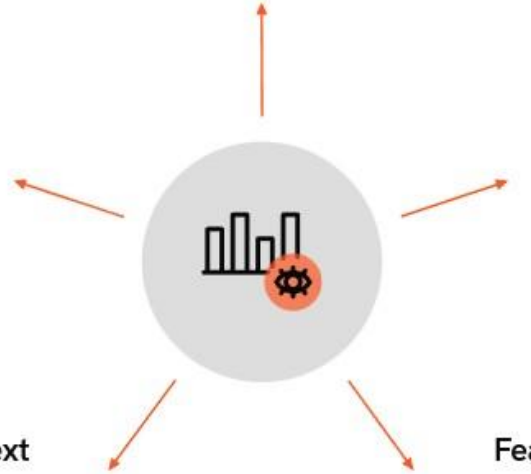
Exploring the Key Features of Our Laptop Dataset for Price Prediction

Comprehensive Dataset

The dataset consists of 1302 rows and 12 features, including brand and specifications.

Opportunity for Improvement

Future enhancement can be achieved by integrating real-time market data for better predictions.



Benchmarking Algorithms

Utilizes models like Random Forest and Decision Trees for effective price predictions.

Historical Context

Dataset from 2022 allows comparison with past studies but requires updates for accuracy.

Feature Relevance

Key features include CPU, RAM, and screen size, crucial for accurate pricing.

Exploratory Analysis

Exploratory Data Analysis (EDA)

Utilizing Visualization Techniques to Enhance Feature Understanding

Heatmaps for Correlation Analysis

Heatmaps revealed critical correlations between features and target variable `Price`, guiding further analysis.

Focus on Impactful Features

The exploratory analysis helped refine focus on features most relevant to predicting laptop prices.

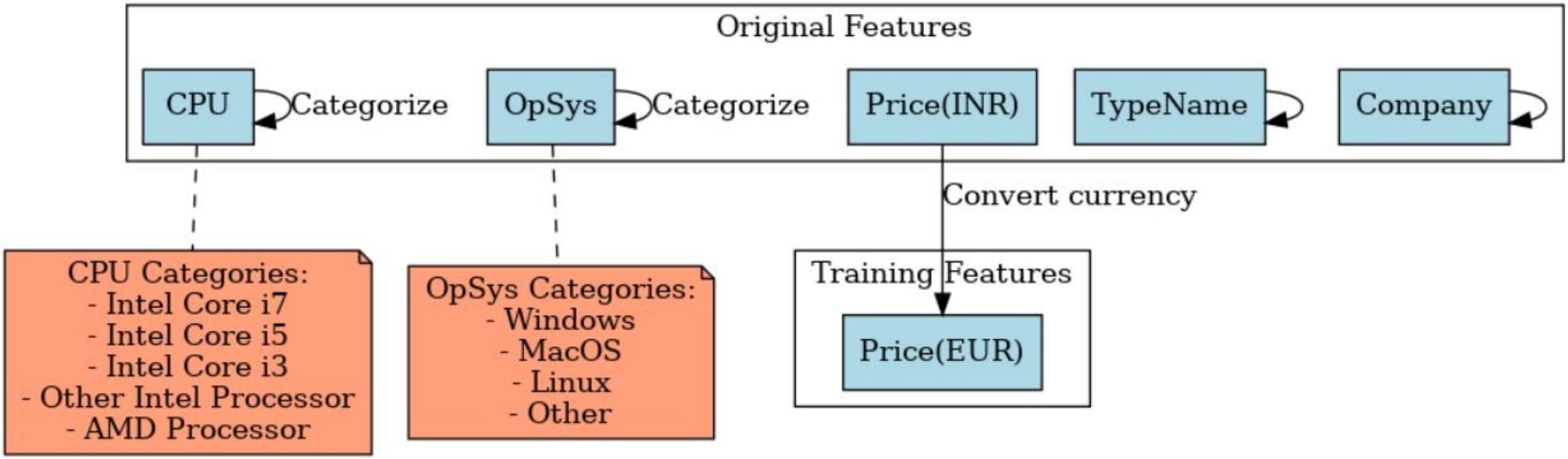
Feature Importance in Display

The analysis highlighted the significance of display features, aiding in effective price prediction.

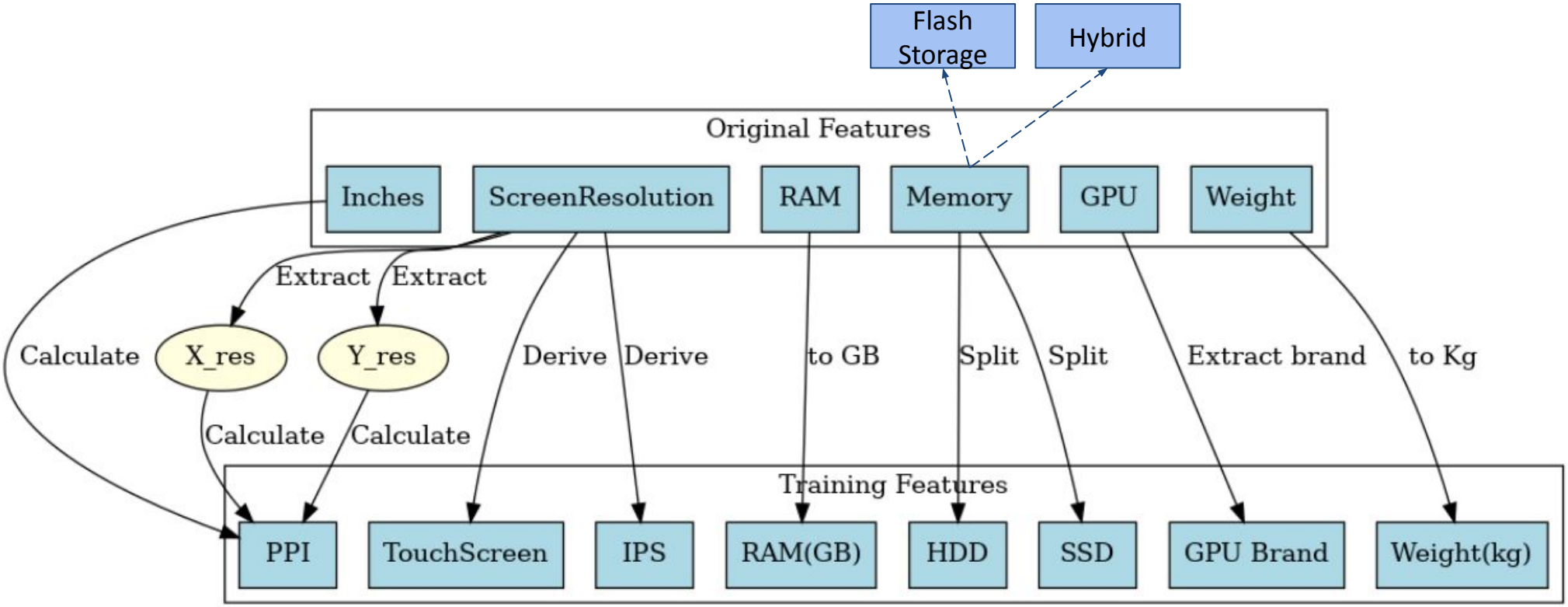
Visualization Techniques Utilized

Seaborn's countplots, barplots and distplots were employed to comprehensively explore the dataset and its attributes.

Dataset Transformations (1)

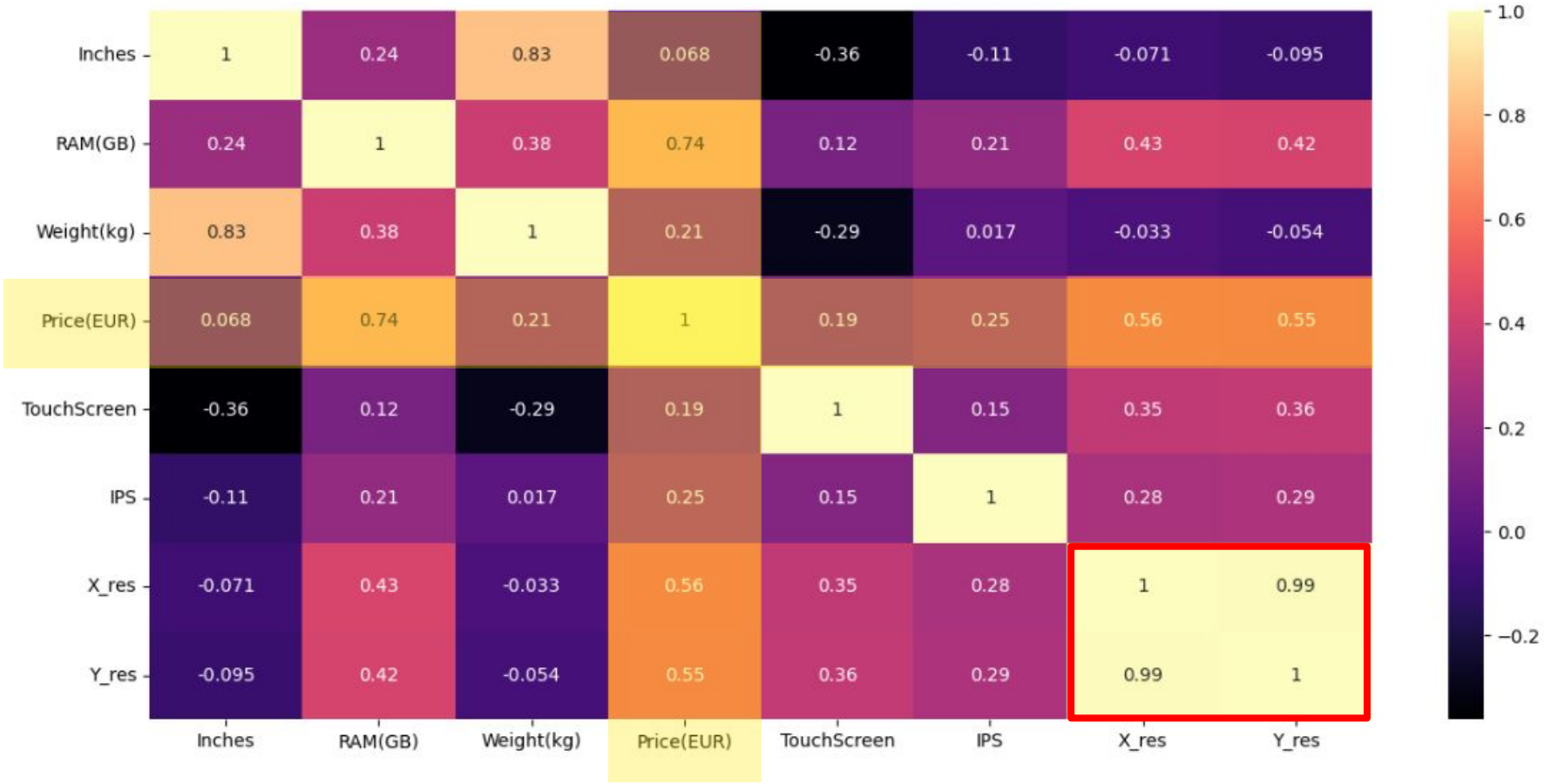


Dataset Transformations (2)

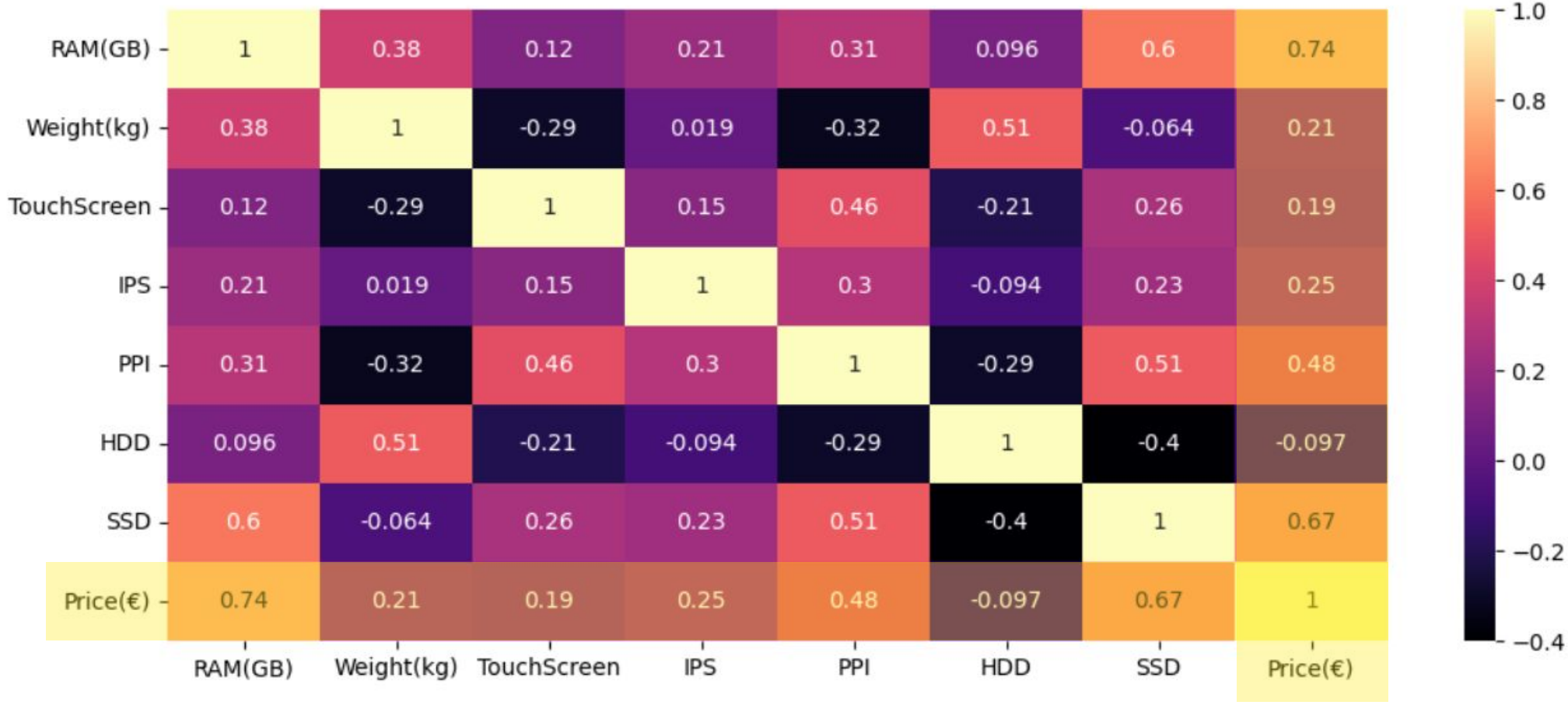


$$pixelsperinch : PPI = \frac{\sqrt{X_{resolution}^2 + Y_{resolution}^2}}{inches}$$

Heatmap (1)


Inches
RAM(GB)
Weight(kg)
Price(EUR)
TouchScreen
IPS
X_res
Y_res
1 0.24 0.83 0.068 -0.36 -0.11 -0.071 -0.095
0.24 1 0.38 0.74 0.12 0.21 0.43 0.42
0.83 0.38 1 0.21 -0.29 0.017 -0.033 -0.054
0.068 0.74 0.21 1 0.19 0.25 0.56 0.55
-0.36 0.12 -0.29 0.19 1 0.15 0.35 0.36
-0.11 0.21 0.017 0.25 0.15 1 0.28 0.29
-0.071 0.43 -0.033 0.56 0.35 0.28 1 0.99
-0.095 0.42 -0.054 0.55 0.36 0.29 0.99 1
Inches
RAM(GB)
Weight(kg)
Price(EUR)
TouchScreen
IPS
X_res
Y_res
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

Heatmap (2)


RAM(GB) 1 0.38 0.12 0.21 0.31 0.096 0.6 0.74
Weight(kg) 0.38 1 -0.29 0.019 -0.32 0.51 -0.064 0.21
TouchScreen 0.12 -0.29 1 0.15 0.46 -0.21 0.26 0.19
IPS 0.21 0.019 0.15 1 0.3 -0.094 0.23 0.25
PPI 0.31 -0.32 0.46 0.3 1 -0.29 0.51 0.48
HDD 0.096 0.51 -0.21 -0.094 -0.29 1 -0.4 -0.097
SSD 0.6 -0.064 0.26 0.23 0.51 -0.4 1 0.67
Price(€) 0.74 0.21 0.19 0.25 0.48 -0.097 0.67 1
RAM(GB) Weight(kg) TouchScreen IPS PPI HDD SSD Price(€)
1.0 0.8 0.6 0.4 0.2 0.0 −0.2 −0.4

Data Preparation

Data Preprocessing Steps

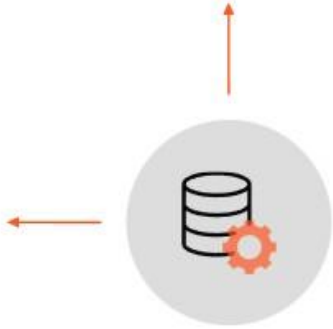
Essential Steps for Preparing Data in Price Prediction Models

Target Transformation

Applied logarithmic transformation to normalize price distribution, enhancing model performance.

Handling Missing Data

Addressed anomalies and missing values to maintain model integrity and accuracy.



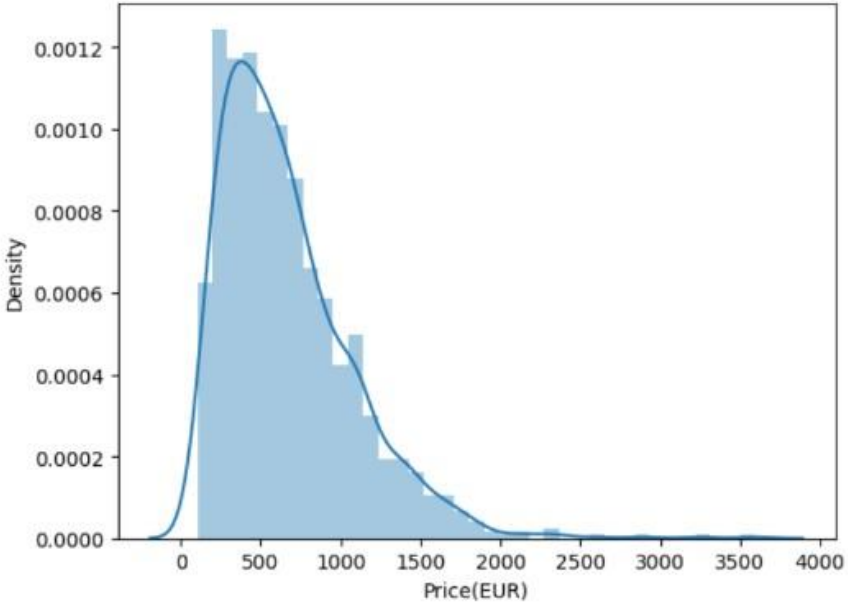
Feature Engineering

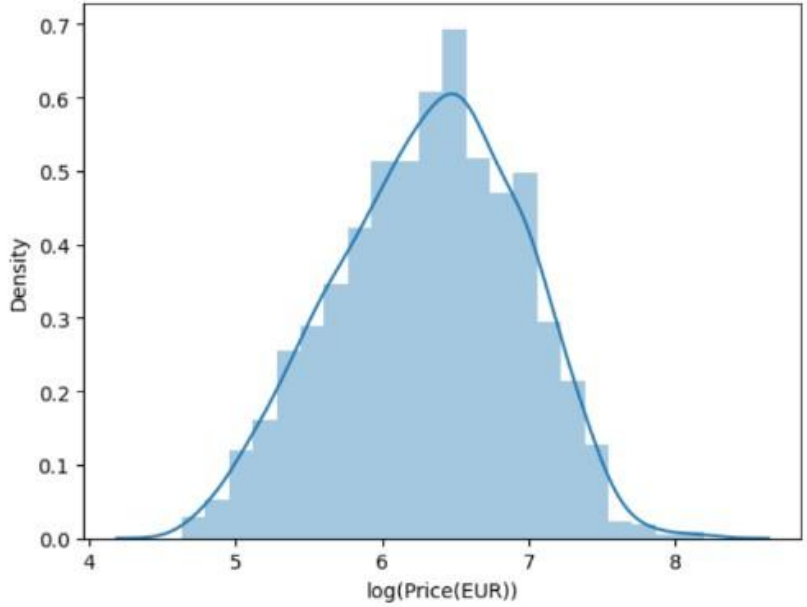
Introduced new features like PPI and refined existing variables, ensuring data quality.

Encoding Categorical Variables

Converted categorical variables such as Company and OpSys into numeric values for model compatibility.

Target Variable Transformation





Model Evaluation

Model Development

Exploring Regression Models for Accurate Price Predictions

- **Linear Regression**

  A fundamental approach assuming a linear relationship between features and price.

- **Lasso and Ridge Regression**

  Regularized models designed to prevent overfitting by penalizing large coefficients.

- **Decision Tree Regression**

  A non-linear model that segments data into branches based on feature values.

- **Random Forest Regression**

  An ensemble method using multiple decision trees to enhance accuracy and robustness.

- **K-Nearest Neighbors (KNN)**

  Predicts prices by averaging the prices of the closest laptops in feature space.

- **Hyperparameter Tuning with GridSearchCV**

  Optimizes model performance by fine-tuning hyperparameters across multiple models.

Model Performance

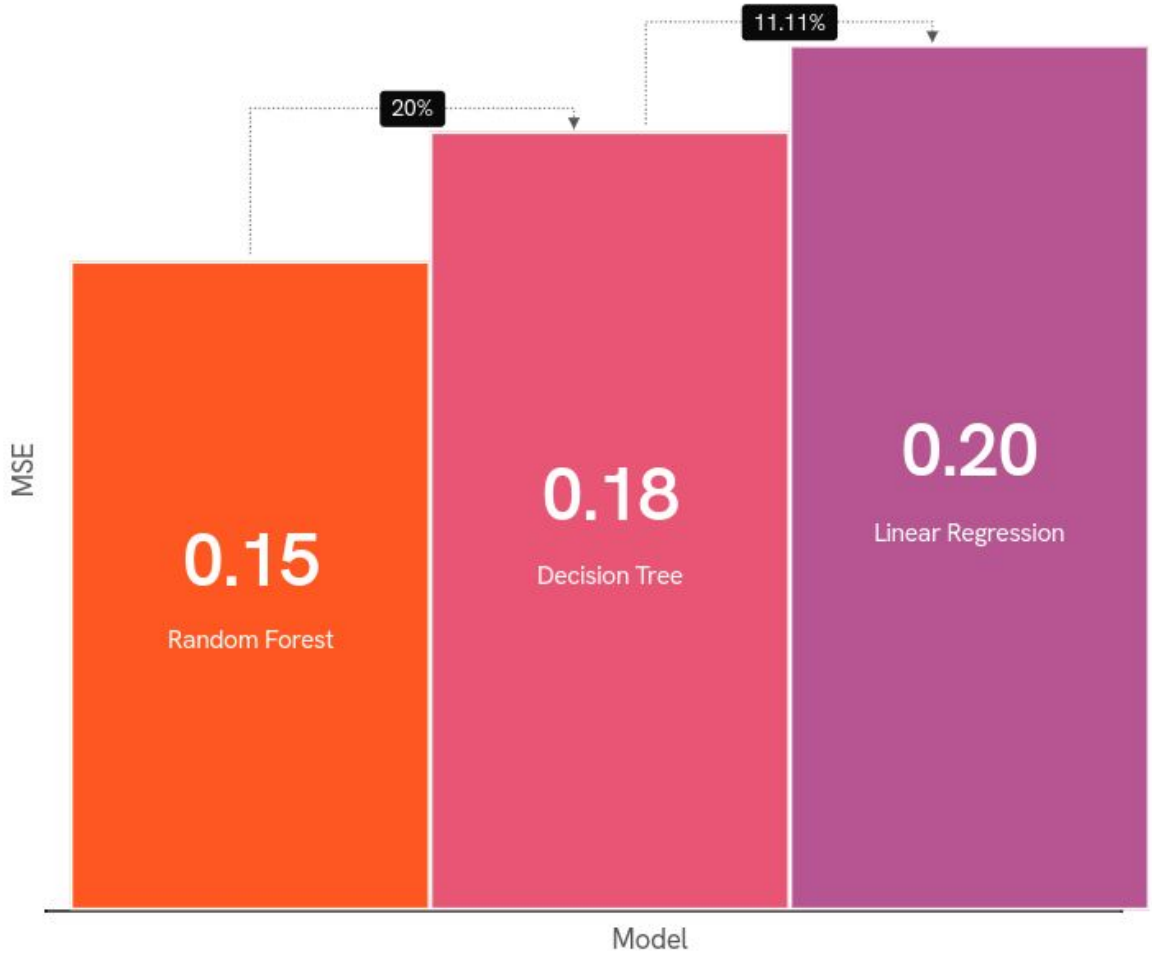
Model Performance Comparison : R^2 score

Comparative analysis of R^2 scores for various predictive models.

Model Performance

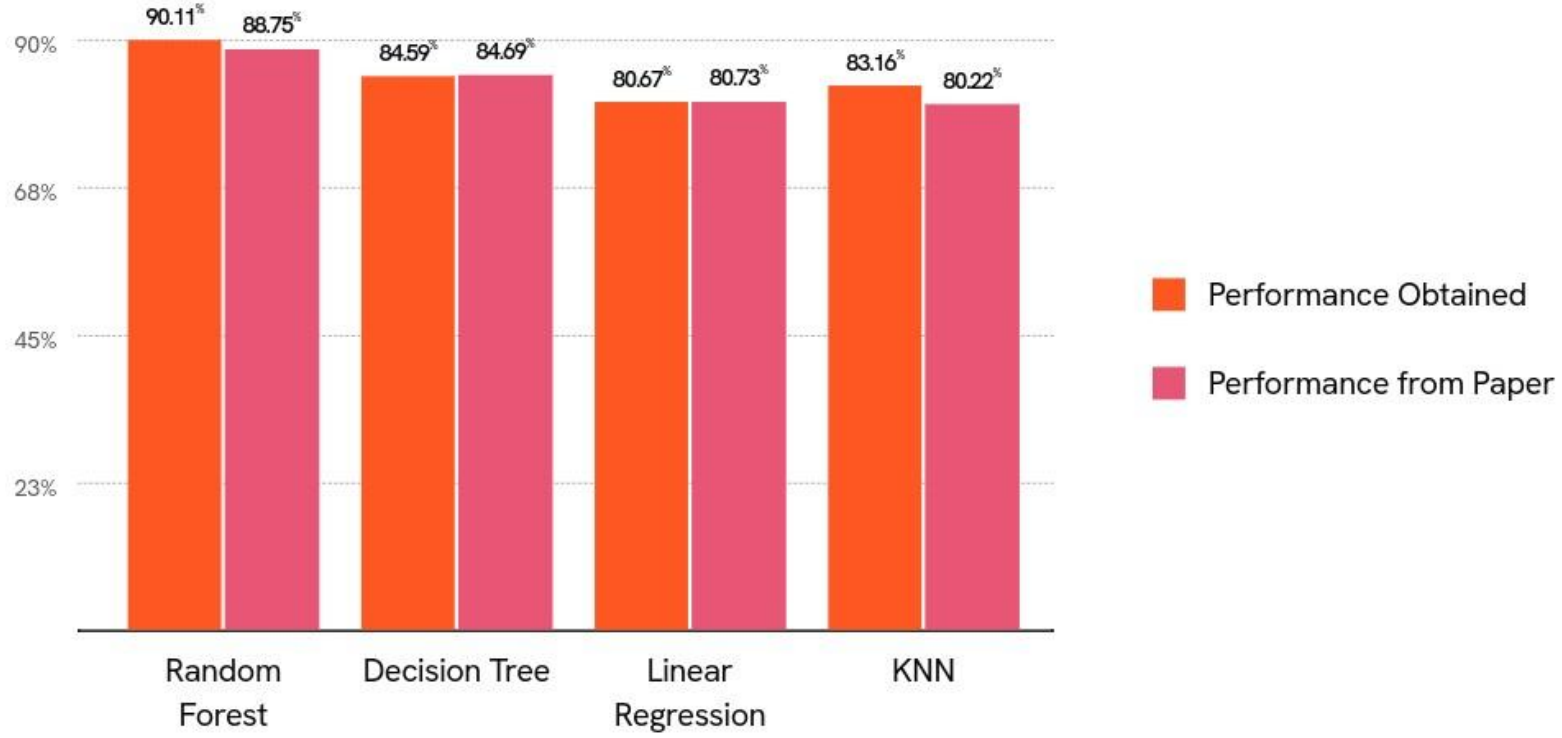
Model Performance Comparison : MSE

Comparative analysis of MSEs for various predictive models.



Comparison with Paper's Results

Compare the results obtained with the ones available in scientific paper titled "Laptop Price Prediction using Machine Learning Algorithms," presented at the 2022 International Conference on Emerging Trends in Engineering and Medical Sciences



Accuracy

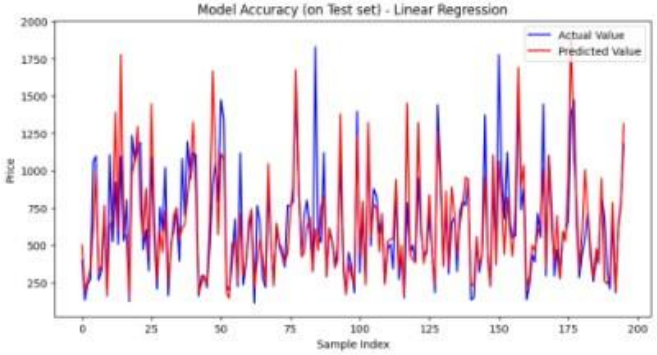


Figure 4.2: Linear Regression

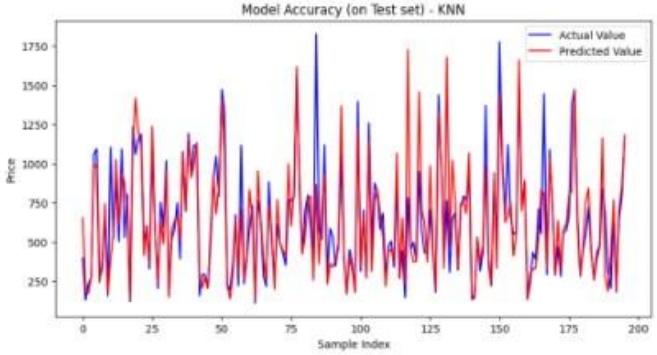


Figure 4.3: K-Nearest Neighbors

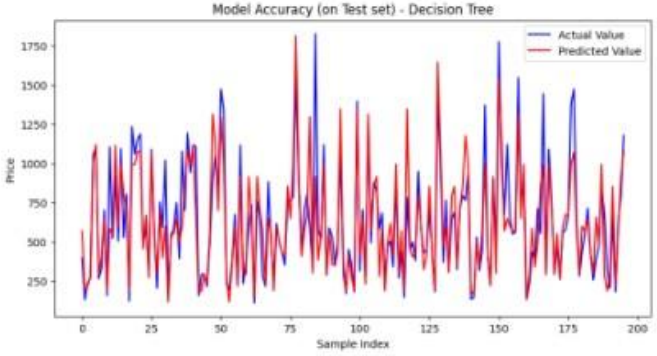


Figure 4.4: Decision Tree

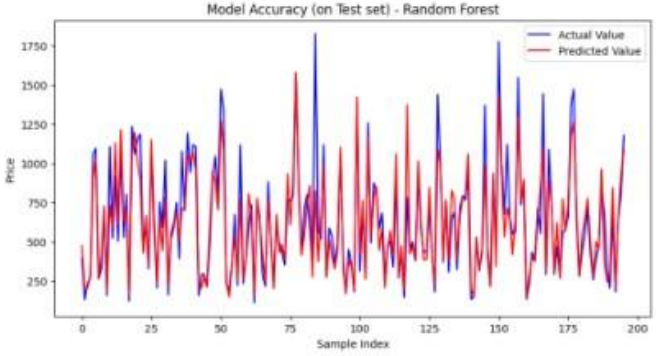


Figure 4.5: Random Forest

Paper Models' Accuracy

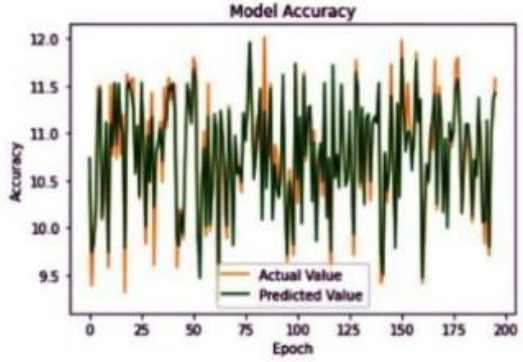


Fig.5. Multiple linear regression Accuracy

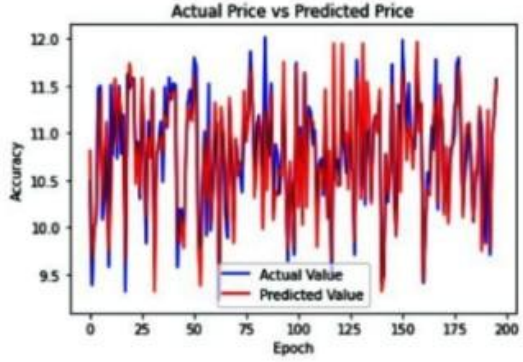


Fig.6. KNN Algorithm Accuracy

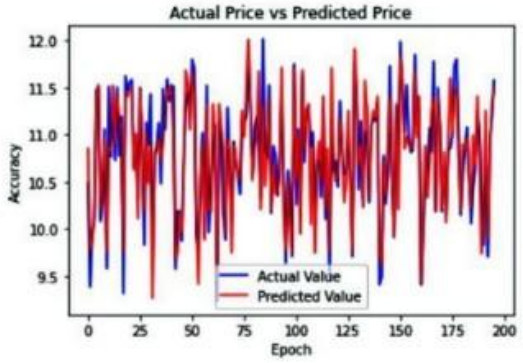


Fig.7. Decision Tree Accuracy

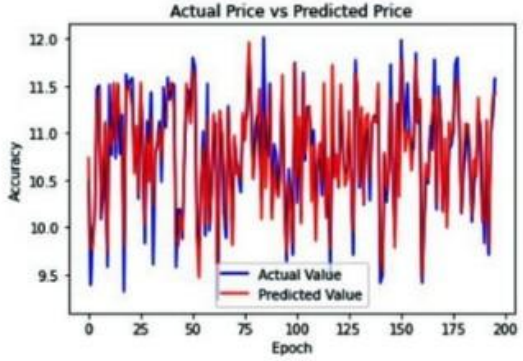
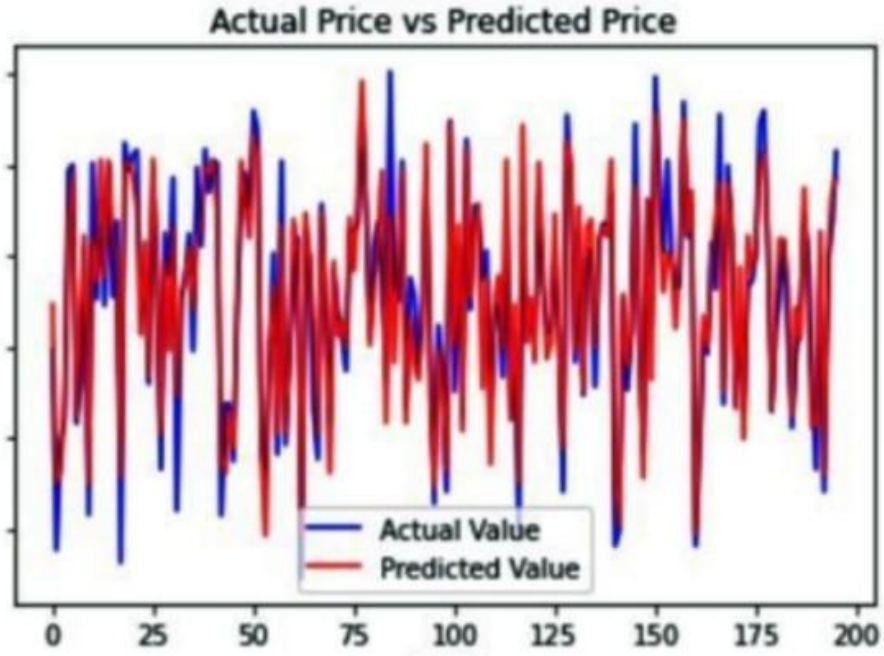
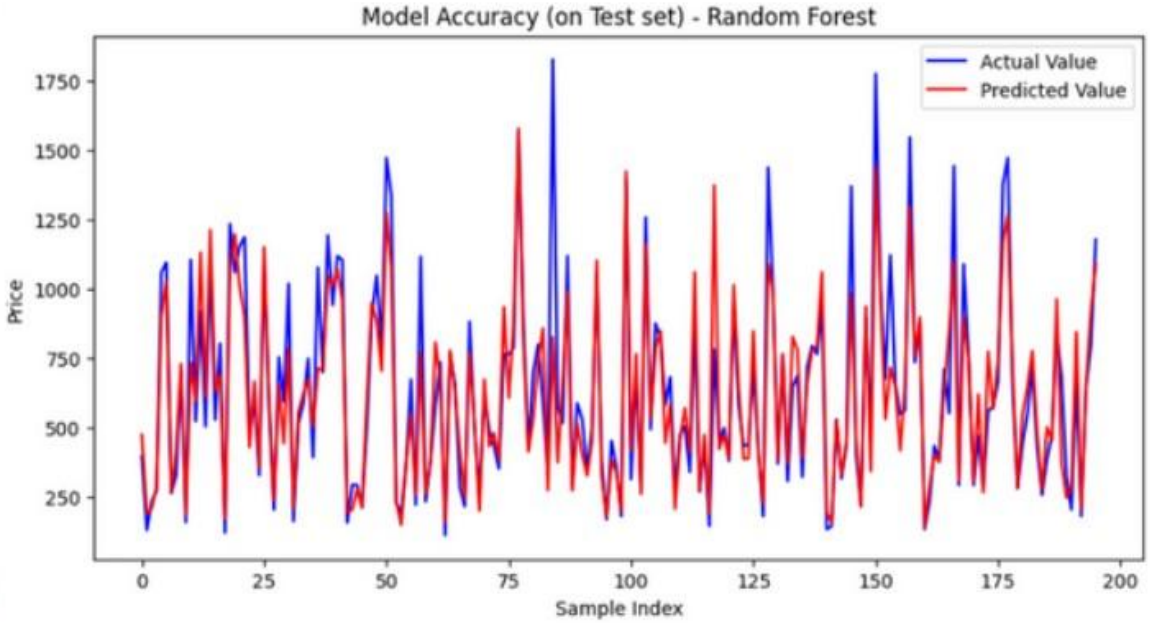


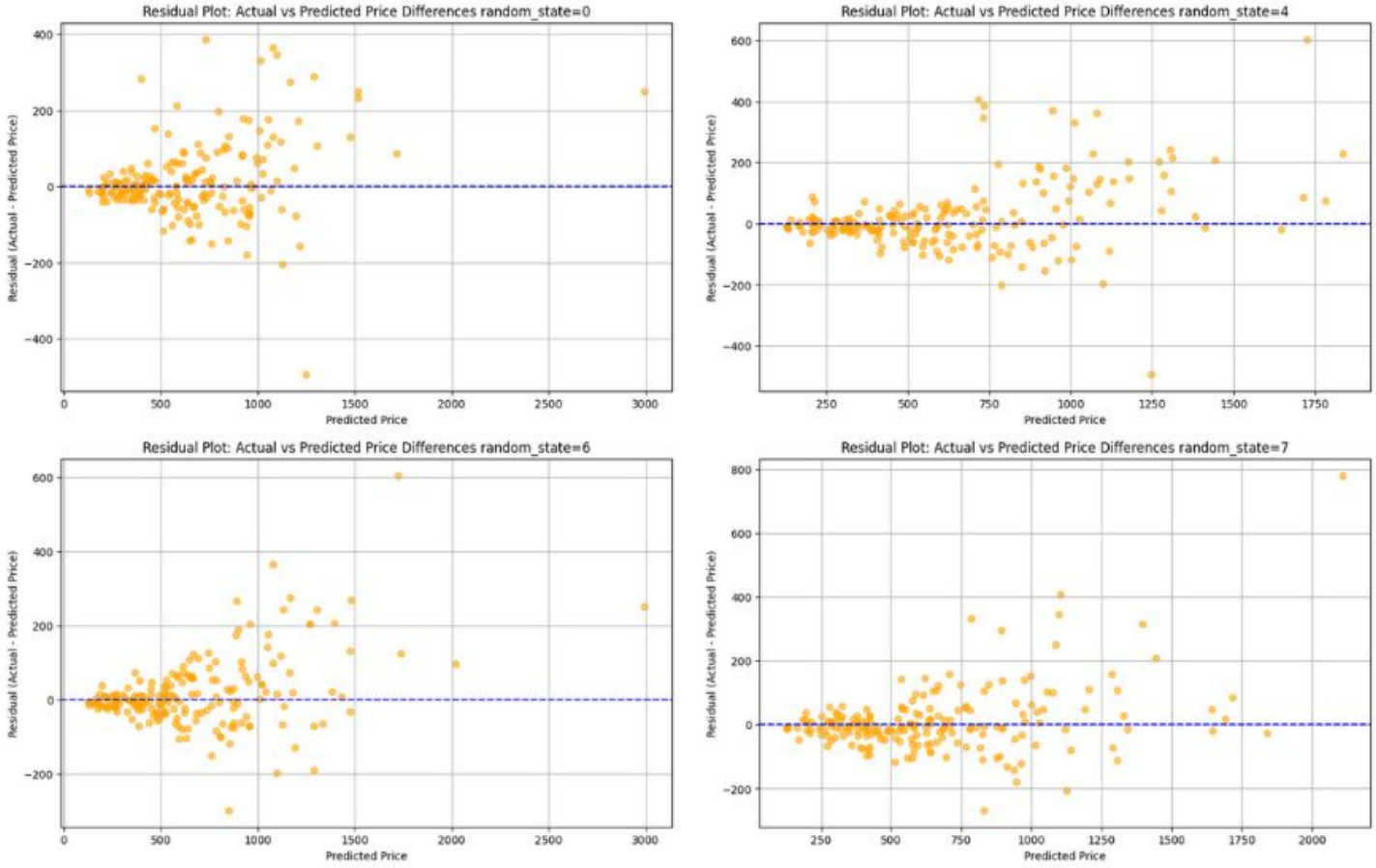
Fig.8. Random Forest Accuracy

Comparison


Actual Price vs Predicted Price
Actual Value
Predicted Value
0
25
50
75
100
125
150
175
200


Model Accuracy (on Test set) - Random Forest
Actual Value
Predicted Value
Price
1750
1500
1250
1000
750
500
250
0
25
50
75
100
125
150
175
200
Sample Index

Residual Plot: Actual vs Predicted Price



Streamlit Cloud Features

Deployment Using Streamlit Cloud

Making Machine Learning Accessible for Laptop Price Predictions

User-Friendly Interface

The application allows easy input of laptop specifications for predictions.

Model Selection

Users can choose from different models to obtain price predictions tailored to their needs.

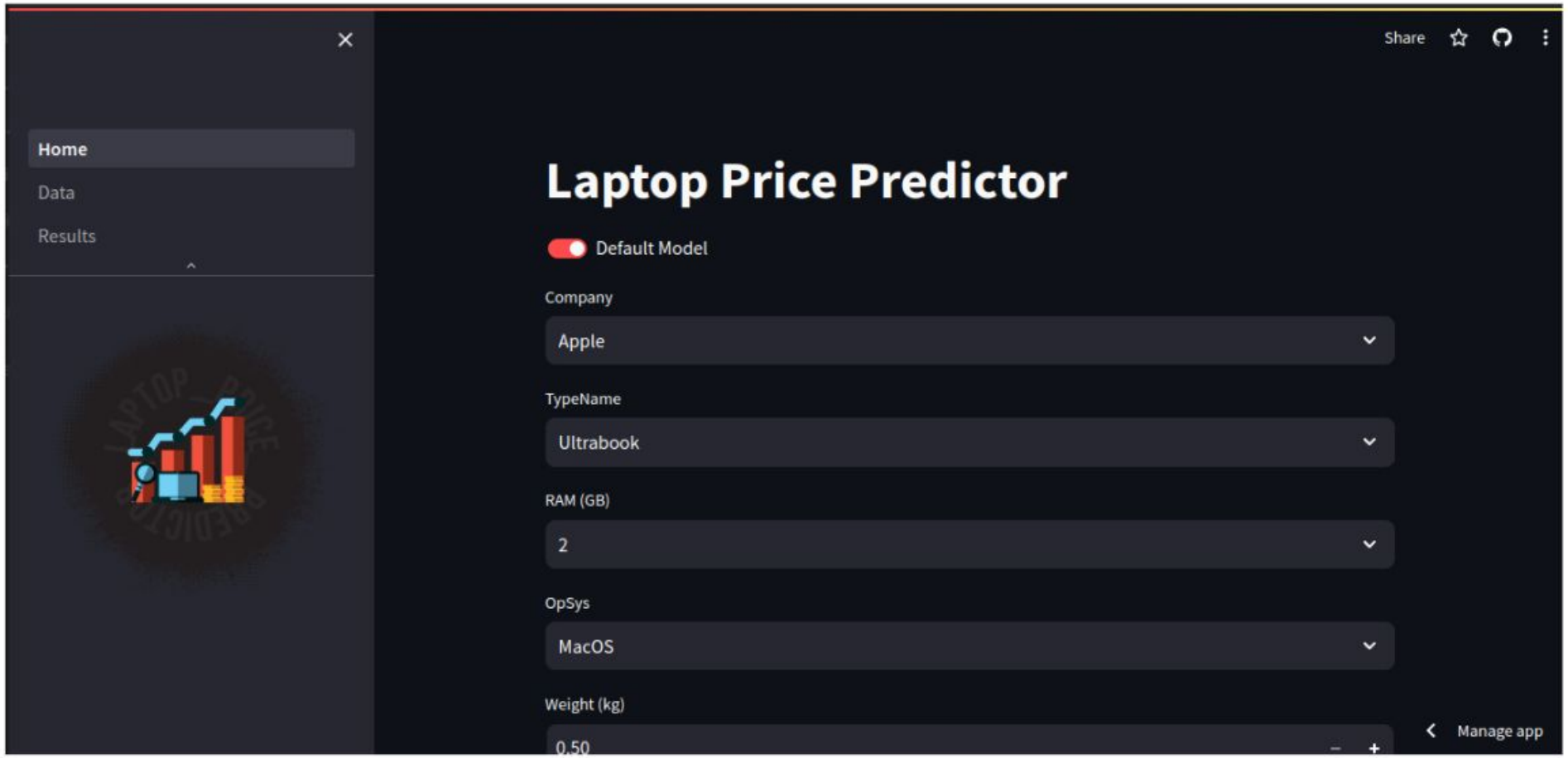
Prediction Intervals

Offers prediction intervals, giving users a range of possible prices for better decision-making.

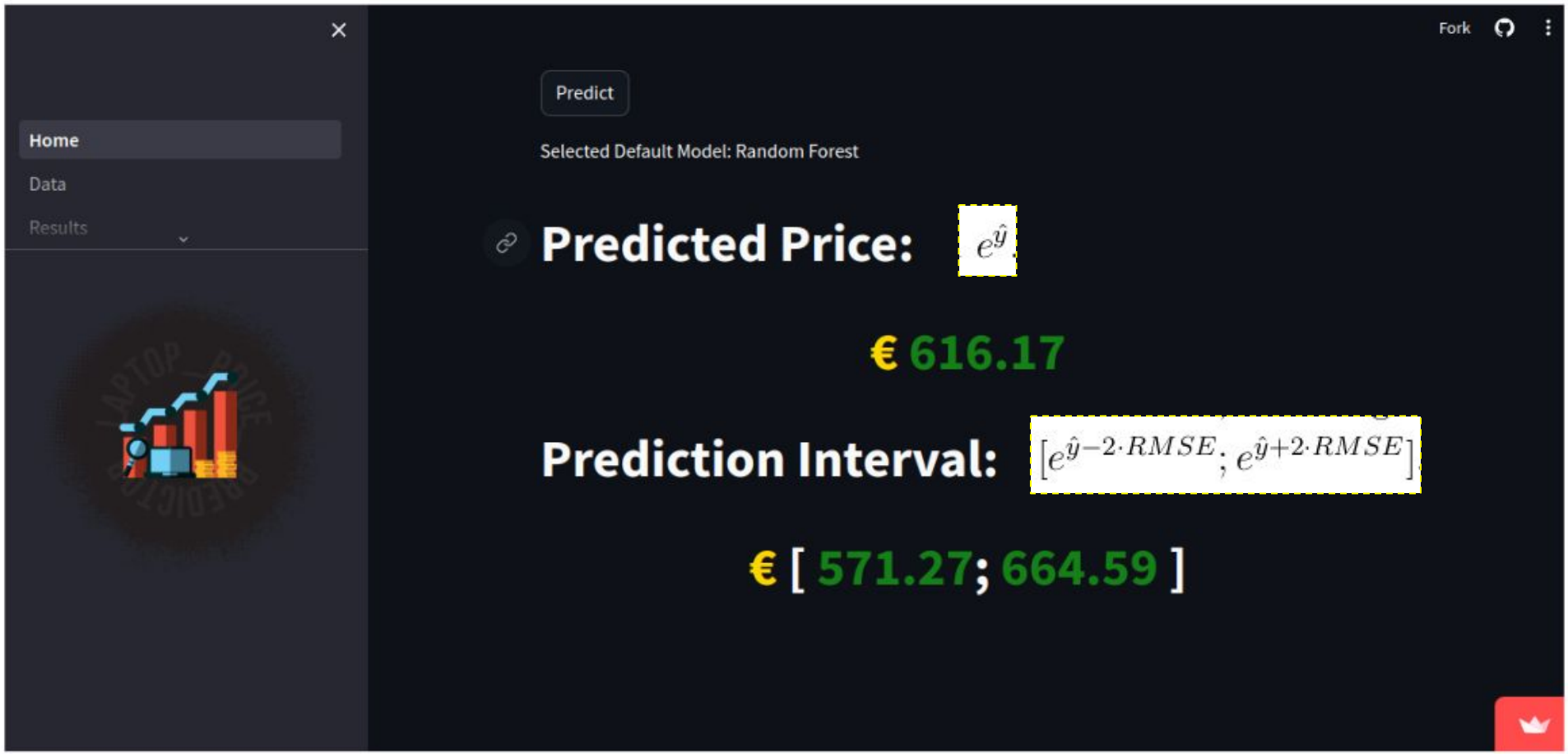
Accessibility

Deployment on Streamlit Cloud ensures that the model is accessible from anywhere with an internet connection.

Deployment on Streamlit Cloud



Target Variable Transformation - Prediction Intervals



Model Enhancement

Future Enhancements

Approaches to Improve Laptop Price Prediction Models

1

Integration of Recent Data

Updating the dataset with current pricing trends will enhance accuracy.

2

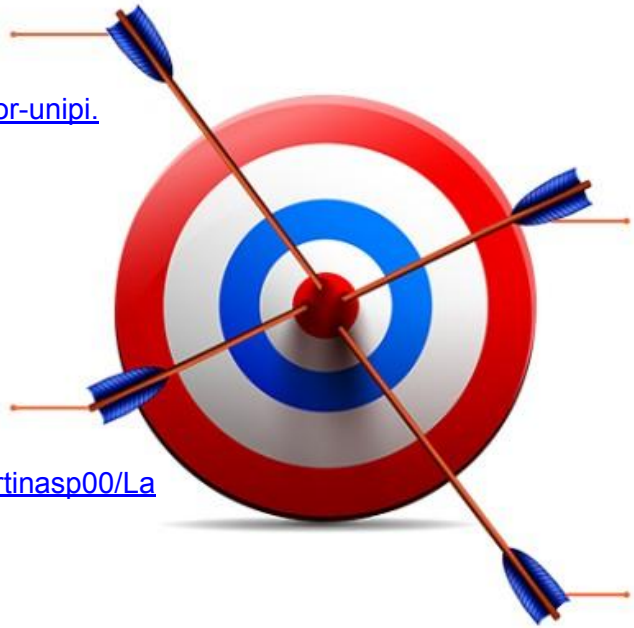
Incremental Learning Algorithms

To ensure our model stays relevant, we could use incremental learning algorithms to update the model as new data becomes available

Useful Links

Streamlit

Streamlit Application
https://laptoppricepredictor-unipi.streamlit.app/



GitHub Repository
https://github.com/martinasp00/LaptopPricePredictor/

IEEE

Scientific Paper : "Laptop Price Prediction using Machine Learning Algorithms"
Laptop Price Prediction using Machine Learning Algorithms | IEEE Conference Publication

Google Colab Notebook
Laptop Price Predictor.ipynb


THANK YOU