Laptop Price Prediction Using XGBoost + Optuna

INTRODUCTION

In today's world, laptops are essential tools for both personal and professional activities. This project aims to predict laptop prices using machine learning techniques. The model will help manufacturers, retailers, and consumers make informed decisions about laptop pricing based on various specifications or features.

We are using dataset 'Laptop Prices' by Omar Wagih from Kaggle. Dataset: 1275 laptops, 23 features (CPU, RAM, storage, etc.)..

Authors
Ghazi Athari – 2703343483
Pierce Amadeo – 270226161
Luqman Hakim – 2702233570
Ananda Dwi Putra – 2702239624

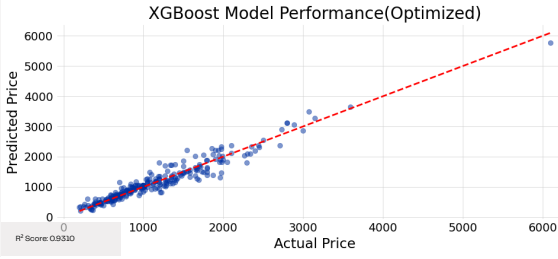
Affiliations
Data Science, Binus University – Group 8 Machine Learning

OBJECTIVE

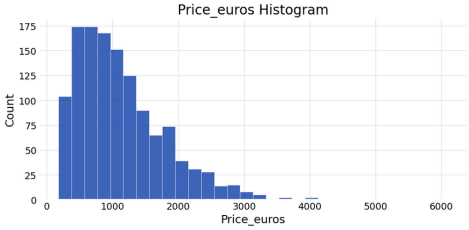
- Predict laptop prices using machine learning.
- Provide insights for competitive pricing strategies.

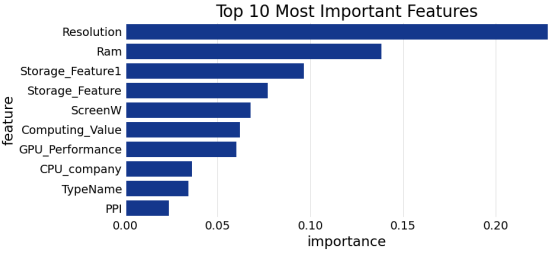
APPROACH

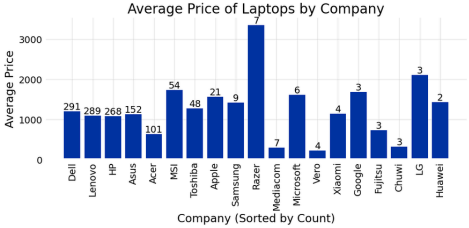
- Data preprocessing: Handled categorical data, normalized numerical features.
- EDA: Analyzed correlations and key trends.
- Modeling: XGBoost Regression
- Evaluation Metrics: R-squared (R^2), Mean Squared Error (MSE), Root Mean Squared Error (RMSE).

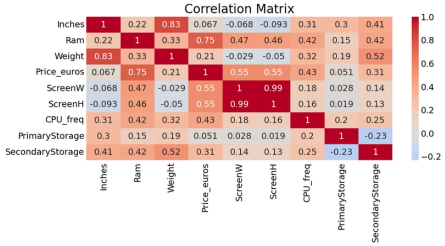


DATA UNDERSTANDING + RESULTS/FINDINGS









ANALYSIS

The optimized XGBoost model achieved superior performance with an R^2 of 93%, significantly surpassing both the baseline XGBoost (R^2 ~ 92%) and Linear Regression (R^2 ~ 81%). Key predictive features included resolution, RAM, and storage, alongside other influential factors like screen width and GPU performance. These results align with domain knowledge, reflecting the importance of performance and display aspect in determining laptop prices. The optimization process, which fine-tuned parameters such as the number of estimators, learning rate, and tree depth, enhanced the model's accuracy and reliability.

CONCLUSION

The Optuna-optimized XGBoost model effectively captures complex relationships in the data, providing a robust tool for predicting laptop prices. This approach offers valuable insights for manufacturers, retailers, and consumers by enabling data-driven pricing strategies. Future work could explore advanced feature engineering, the incorporation of additional data sources, and further model optimization to refine predictions and improve overall performance.

Model	R^2	MSE	RMSE	MAE	Key Features
Linear Regression	0.81	95,560.99	309.13	214.56	RAM, Weight, Storage
Base XGBoost	-0.92	-41,172.96	-202.91	-142.41	RAM, ScreenW, GPU
Optimized XGBoost	0.93	34,681.72	186.23	134.05	RAM, Resolution, Storage