

# Assignment 2: Model Documentation

Assignment 2: Laptop Price Prediction Model

This documentation describes the steps taken to preprocess the data, train a machine and evaluate performance.

Author: (not specified)

# Data Preprocessing Summary

## Preprocessing Steps:

- Converted 'Ram' from string format (e.g., '8GB') to numeric.
- Cleaned 'Weight' column to remove 'kg'.
- Extracted 'Cpu Brand' from the 'Cpu' column (first two words).
- Standardized 'Memory' values and converted to GB.
- Dropped irrelevant columns such as 'ScreenResolution', 'Gpu', etc.
- Applied one-hot encoding on categorical columns.
- Saved processed data for reuse in model training.

# Model Training & Evaluation

Model Training: RandomForestRegressor (100 trees, random\_state=42)

Performance Metrics:

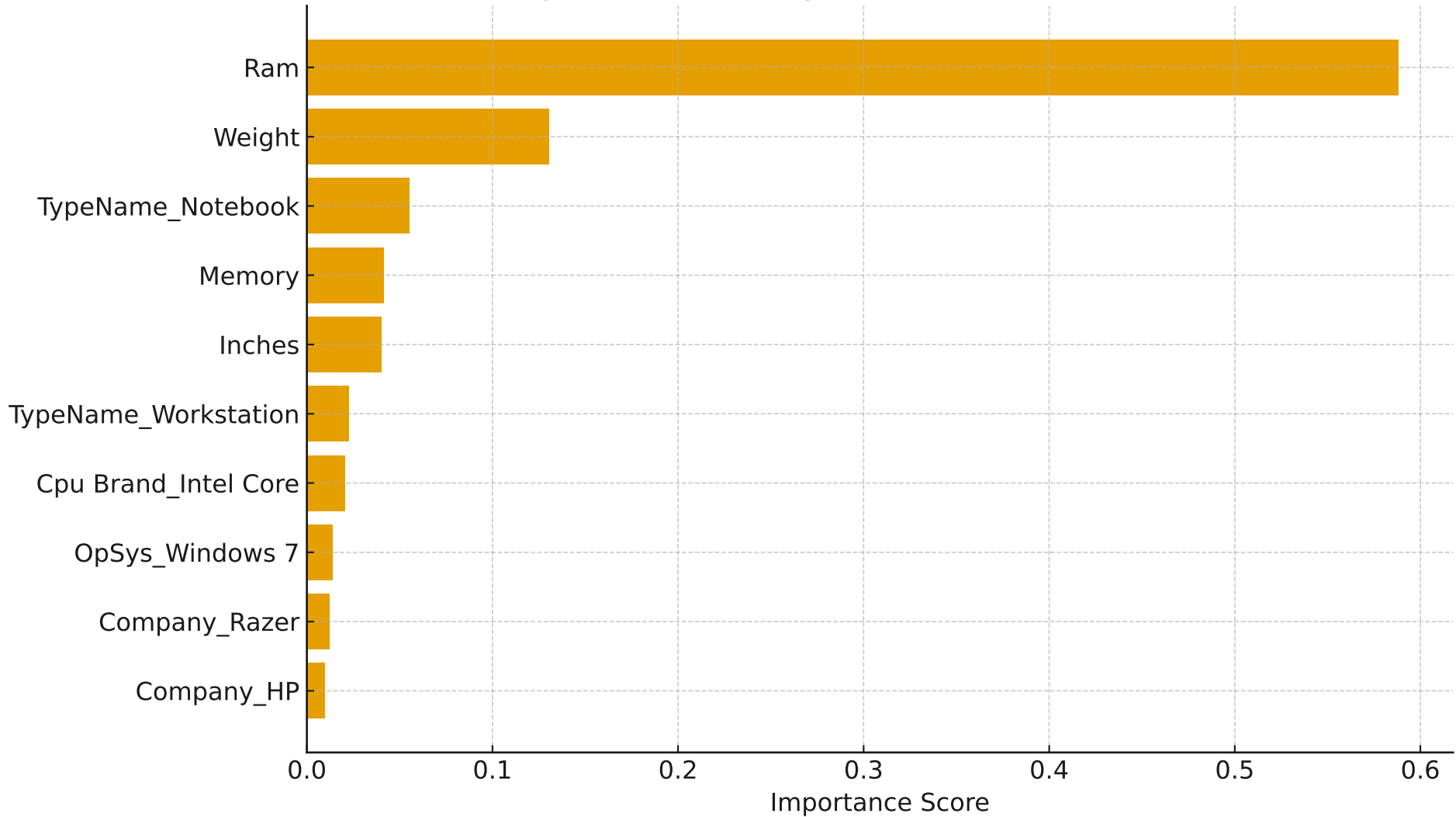
R<sup>2</sup> Score: 0.7845

Mean Absolute Error (MAE): 10862.11

Interpretation:

- R<sup>2</sup> indicates how well the model explains price variance.
- MAE shows the average prediction error in price units.

Top 10 Feature Importances (Random Forest)



# Key Code Snippet (assignment2\_model.py)

```
# Key part of assignment2_model.py

from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split

m = LaptopPriceModel('laptop_data.csv')
m.preprocess()
m.train_model()
m.evaluate_model()
m.save_model('laptop_price_model.pkl')
```

# Conclusion & Next Steps

## Conclusions and Recommendations:

- The Random Forest model performed well with strong predictive power.
- Further tuning (e.g., hyperparameter optimization) could improve accuracy.
- Feature importance analysis helps identify which laptop attributes most affect price.
- Consider saving and loading the trained model using pickle for deployment.