



Assessing the Impact of Technical Features on Laptop Pricing: A Comprehensive Evaluation.

Final Project Presentation

Patchara Suwanbordin (6636058)





Outline

- Introduction
- Objectives
- Material and Methods
- Conclusion
- Q&A



Introduction





Pantip

<https://pantip.com> › topic :

โดนโกงโดยร้านขายโน้ตบุ๊กในไอจี

Jan 4, 2565 BE — คือเราหาเงินได้จำนวนหนึ่ง แล้วเราอยากมีโน้ตบุ๊กไว้ศึกษาต่อเราก็เลยหาซื้อในไอจี มันมีสองร้าน อีกร้านราคาดีน่าเชื่อถือแต่สินค้าหมดทุกรายการ เราเป็นคนคิดน้อย ...



NotebookSPEC.COM

<https://notebookspec.com> › web › 4... :

เตือนภัย!! ร้านขายโน้ตบุ๊กราคาถูกผิดปกติที่ไม่มีทางเป็นไปได้บน ...

Mar 20, 2562 BE — เตือนภัย!! ร้านขายโน้ตบุ๊กราคาถูกผิดปกติที่ไม่มีทางเป็นไปได้บน Lazada · ราคาถูกเกินไป ต่อให้เป็นราคามือ 2 ก็ยังถูกเกินไป · ชื่อร้านแปลก และเพิ่งเปิดใหม่ไม่กี่วัน · บางร้านบ้านเรา...



Facebook · โน้ตบุ๊ก ราคาปลีก-ส่ง

50+ reactions · 1 year ago :

โน้ตบุ๊ก ราคาปลีก-ส่ง | อย่าลงเชื่อคนแบบนี้ นะคะ เราโดนโกง ...

อย่าลงเชื่อคนแบบนี้ นะคะ เราโดนโกง ใหม่ๆเลยคะ ใครตามให้ได้ช่วยตามให้ทีนะคะ ยอด4500 เลยคะ จำเป็นต้องจ่ายเงินคะ.



Thai PBS

<https://www.thaipbs.or.th> › content :

ร้อง ศปอส.ตร.ถูกหลอกขายโทรศัพท์มือถือ และโน้ตบุ๊ก

Nov 3, 2563 BE — ... ราคาถูกพร้อมส่ง หลังผู้เสียหายโอนเงินไปแล้ว กลับไม่ได้รับสินค้าและติดต่อเจ้าของเพจไม่ได้ ผู้เสียหายจึงรวมตัวกันในเพจโดนโกง จึงพบว่าผู้ก่อ ...



Objectives





Objectives

- Understand the data mining processing
- Analyze the laptop price base on laptop features
- Utilize data mining and machine learning techniques to analyze laptop data and predict sensible price by laptop spec



Material and Methods





Material - Laptop dataset

- Amazon.com (Oct 2023)
- 4500 records
- 14 features



COLUMN NAME	brand	model	screen_size	color	harddisk	cpu	ram
MISSING VALUE	0%	26%	1%	13%	13%	2%	1%
DATA TYPE	string	string	string	string	string	string	string
DESCRIPTION	ยี่ห้อ	รุ่น	ขนาดของหน้าจอ (นิ้ว)	สี	Hard disk ที่ติดตั้งมา	หน่วยประมวลผล	หน่วยความจำ

COLUMN NAME	OS	special_features	graphics	graphics_processor	cpu_speed	rating	price
MISSING VALUE	1%	54%	1%	42%	66%	51%	0%
DATA TYPE	string	string	string	string	string	float64	string
DESCRIPTION	ระบบประมวลผล	ส่วนเสริมอื่นๆ	ประเภทการ์ดจอ	การ์ดจอ	ความเร็วหน่วยประมวลผล	คะแนนจากผู้ใช้งาน 0 ถึง 5	ราคา

Methods - Data cleansing

- Transform graphics feature (Dedicated, Integrated)
- Reformat of numerical features
 - screen_size: The size of the laptop screen in inches.
 - hard disk: The size of the hard disk in GB.
 - ram: The amount of RAM in GB.
 - price: The price of the laptop in dollars.
- Reformat the OS in general (Windows, macOS, Chrome OS, Linux)



Methods - Data cleansing

brand	model	screen_size	color	hard disk	cpu	ram	OS	special_features	graphics	graphics_coprocessor	cpu_speed	rating	price
acer	Chromeb ook 511 C741LT	11.6	Shale Black	32	AMD Kabini A6-5200 M Quad Core	4	Chrome OS	Backlit	Dedicated	NaN	NaN	NaN	529.53
Dell	Inspiron 7415 2-in-1	14	Grey	512	AMD Ryzen 7	8	Windows	Wifi & Bluetooth	Integrated	NaN	NaN	NaN	633.99
Dell	Inspiron 7415 2-in-1	14	Grey	256	AMD Ryzen 7	8	Windows	Wifi & Bluetooth	Integrated	NaN	NaN	NaN	577.49
Dell	Inspiron 7415 2-in-1	14	Grey	256	AMD Ryzen 7	12	Windows	Wifi & Bluetooth	Integrated	NaN	NaN	NaN	602.47
Dell	Inspiron 7415 2-in-1	14	Grey	256	AMD Ryzen 7	12	Windows	Wifi & Bluetooth	Integrated	NaN	NaN	NaN	602.47

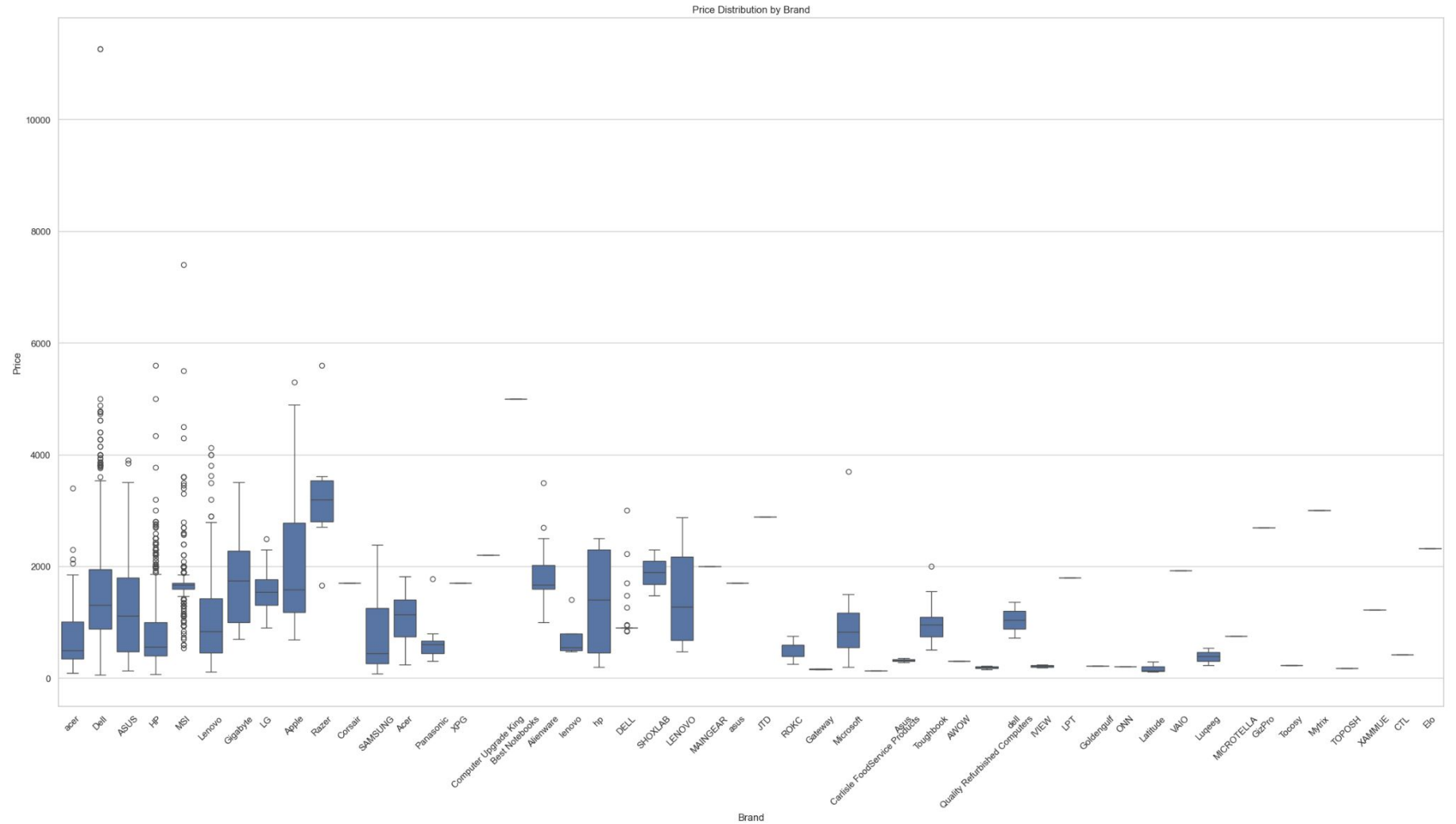
Methods - Data cleansing

feature	missing (%)
cpu_speed	65.811966
special_features	53.80117
rating	51.102114
graphics_coprocessor	42.105263
model	26.180837
color	13.022942
harddisk	13.022942
cpu	2.249213
graphics	1.461988
ram	1.37202
screen_size	0.652272
OS	0.584795
price	0.112461
brand	0

Methods - Data cleansing

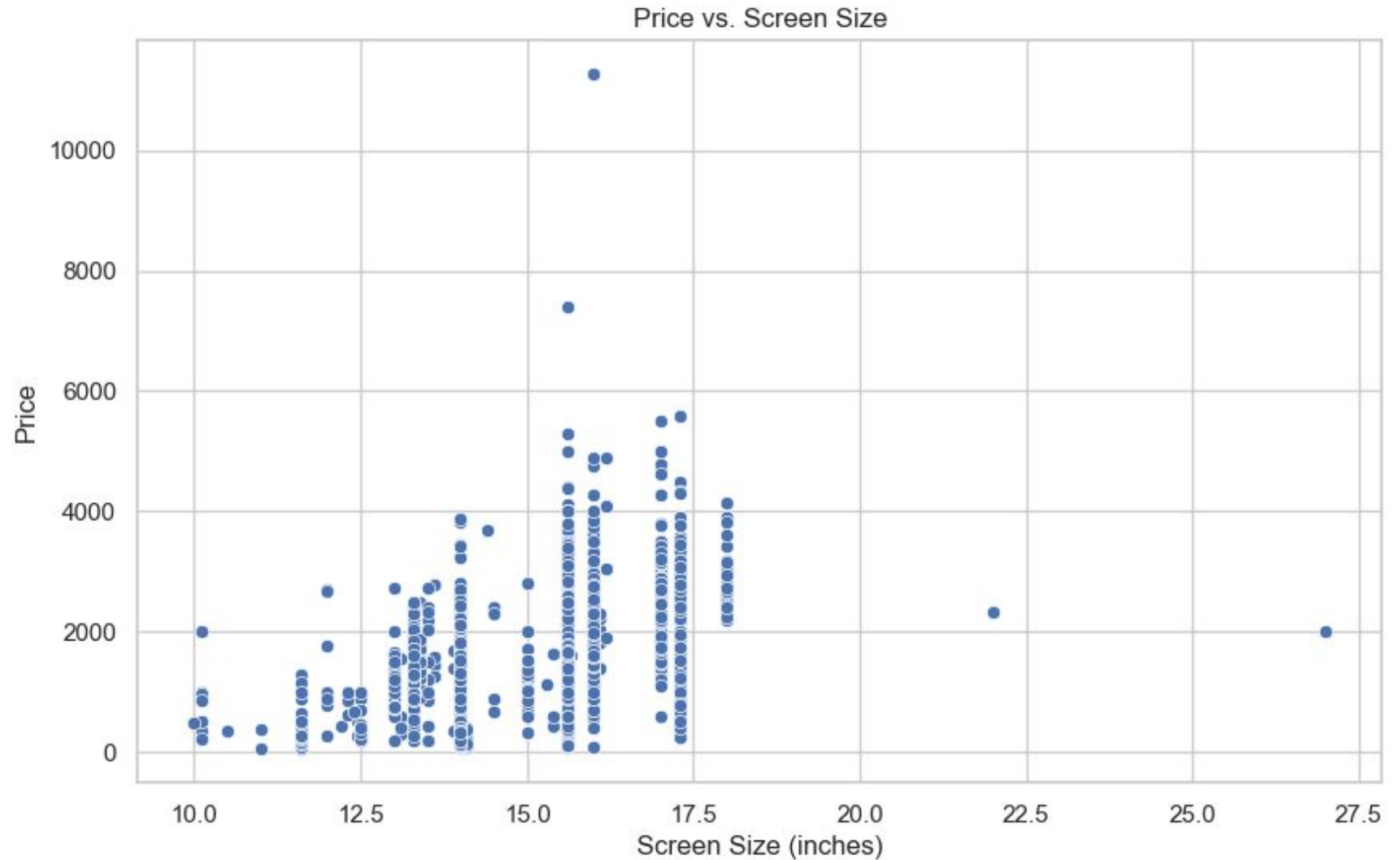
- Remove high missing value (model, color, special_features, graphics_coprocessor, cpu_speed, rating) and missing price
- Replace missing value of hard disk (numerical) with median
- Replace missing value of cpu, graphics, ram, screen_size, OS (categorical) with mode

Methods - Visualize



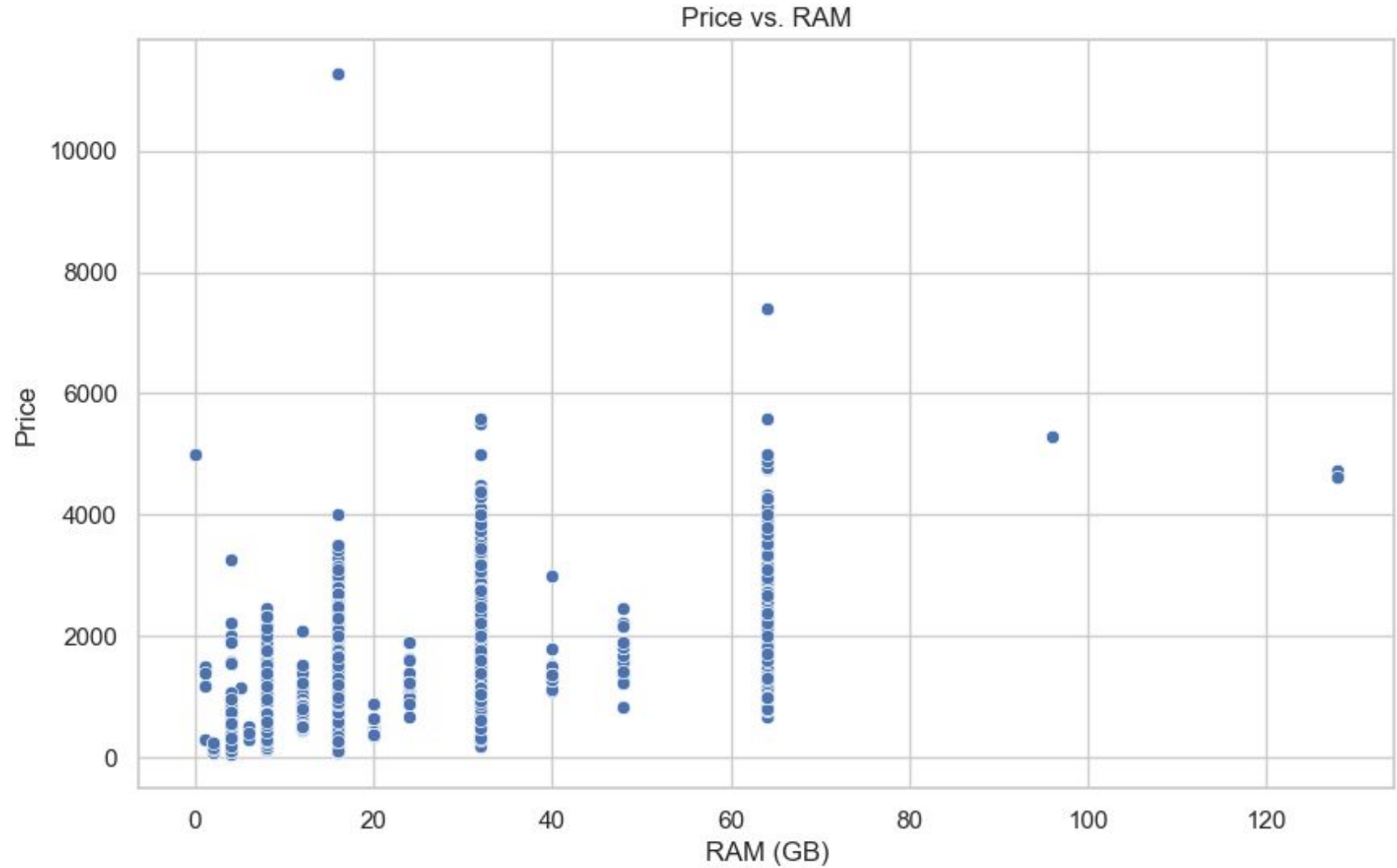


Methods - Visualize





Methods - Visualize



Methods - Column Transformer and Train/Test split

- Standardize numerical feature
- One-hot encoding categorical feature
- Train / Test split (80/20) (3552, 889)

Methods - Model selection

- **Linear Regression:**
 - RMSE = 478.557
 - R2 Score = 0.697
- **Random Forest:**
 - RMSE = 471.966
 - R2 Score = 0.706
- **Gradient Boosting**
 - RMSE = 443.282
 - R2 Score = 0.740

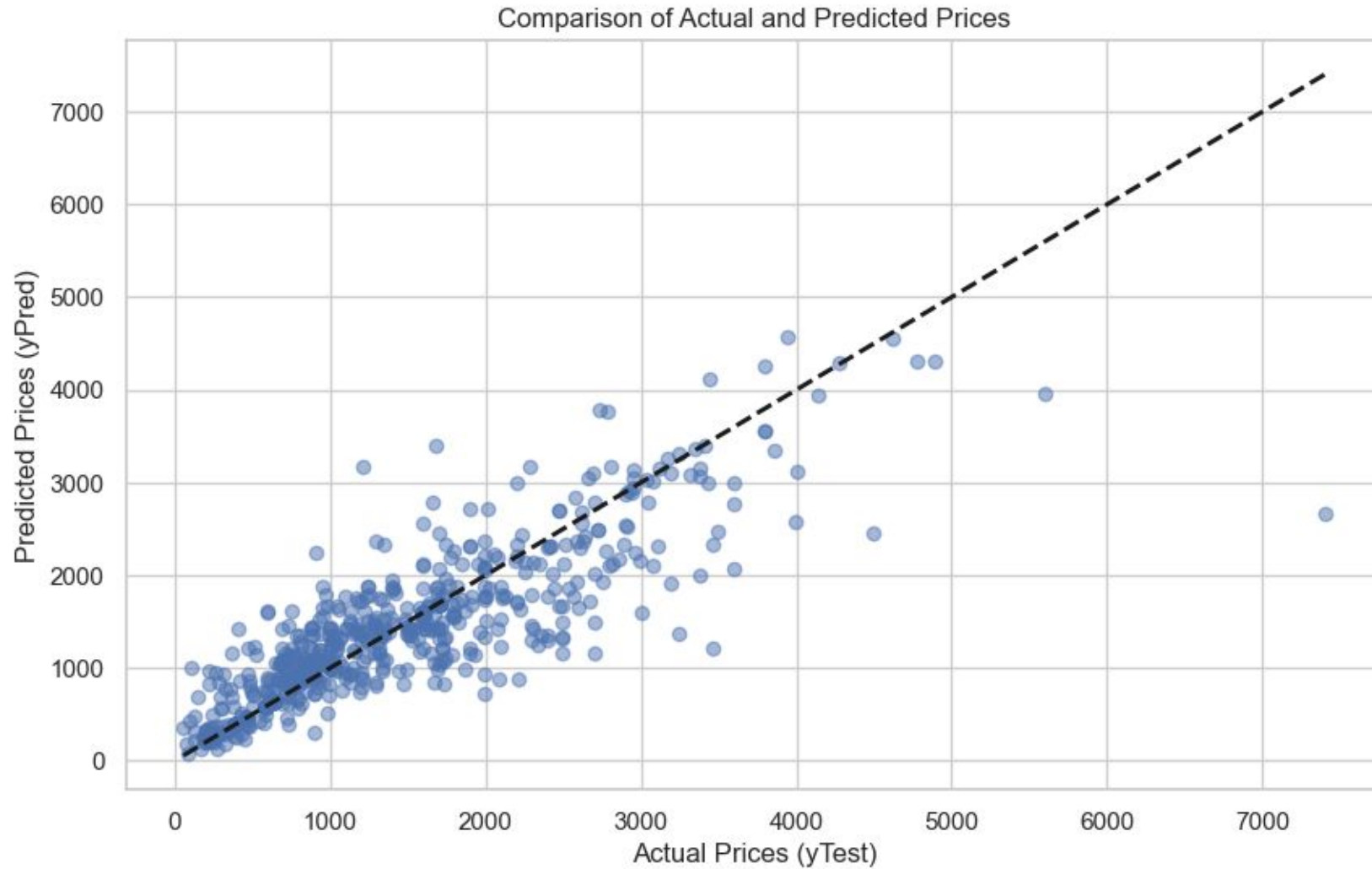


Methods - Hyperparameter Tuning

- **Randomized search on hyper parameters (Gradient Boosting for regression):**
 - RMSE = 422.760
 - R2 Score = 0.764



Methods - Model Result





Methods - Model deployment

```
# Save the model
joblib.dump(rfRegressor, 'model/rf_regressor.joblib')

# Save the preprocessor
joblib.dump(preprocessor, 'model/preprocessor.joblib')
```

POST ▼ http://localhost:1780/predict

Params Auth Headers (9) **Body** ● Pre-req. Tests Se

raw ▼ **JSON** ▼

```
1 {
2   · "brand": "Dell",
3   · "screen_size": 15.6, // Inch
4   · "harddisk": 1000, // GB
5   · "cpu": "Intel Core i7",
6   · "ram": 8, // GB
7   · "OS": "Windows",
8   · "graphics": "Integrated",
9   · "price": 900 // USD
10 }
```

Body ▼

Pretty Raw Preview Visualize **JSON** ▼

```
1 {
2   "isReasonable": true,
3   "predictedPrice": 854.018532111265
4 }
```



Conclusion





Conclusion

From conducting experiments, we start by managing a subset of data, followed by visualizing the data to see the relationships. Then, we prepare the data to train a model using Gradient Boosting techniques. The results are satisfactory, so we further develop the experimental results into a rest API to make it more user-friendly.

Thank you.



Mahidol University
Wisdom of the Land

999 มหาวิทยาลัยมหิดล ถนน พุทรมณฑลสาย 4
ตำบล ศาลายา อำเภอฟุทรมณฑล นครปฐม 73170
โทร : 02 849 6000

