

**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**



**PHÂN TÍCH GIÁ BÁN LAPTOP**

<b>Nhóm 2</b>			
Sinh viên thực hiện:			
STT	Họ tên	MSSV	Ngành
38	Nguyễn Ngọc Thanh Sang	21522544	CNCL
41	Phạm Minh Triết	21522712	CNCL
43	Lê Quang Trường	21522732	CNCL

**TP. HỒ CHÍ MINH – 10/2024**

## 1. GIỚI THIỆU

Đề tài này nhằm mục đích phân tích các dữ liệu liên quan đến giá bán của laptop dựa trên bộ dữ liệu được thu thập từ trang web Kaggle[1]. Mục tiêu của đề tài là phân tích và dự đoán giá bán của laptop có độ chính xác cao.

Để thực hiện đề tài này, nhóm đã sử dụng các công cụ và kỹ thuật cho phân tích dữ liệu. Nhóm đã sử dụng Google Colab và Jupyter Notebook để hỗ trợ cho quá trình lập trình và [Github](#) để lưu trữ mã nguồn và dữ liệu. Quá trình thao tác, xử lý và phân tích dữ liệu được thực hiện thông qua ngôn ngữ lập trình Python và thư viện Pandas...viết tiếp.

Trong đề tài này, nhóm đã sử dụng bộ dữ liệu mẫu của tác giả Pradeep Jangir đăng tải trên Kaggle[2] và đề tài mẫu Laptop Price Prediction[3]. Nhóm đã...viết tiếp.

Tóm tắt kết quả đạt được. Viết tiếp.

## 2. MÔ TẢ BỘ DỮ LIỆU (BẮT BUỘC)

Bộ dữ liệu được nhóm tham khảo tại Kaggle[2].

Bộ dữ liệu được tác giả thu thập từ trang web Smartprix[4], một nền tảng trực tuyến liệt kê và so sánh các sản phẩm điện tử khác nhau, bao gồm cả máy tính xách tay. Dữ liệu được thu thập vào ngày 3/8/2024.

Bộ dữ liệu có tổng cộng **3976 dòng** và **18 cột**, trong đó bao gồm **biến mục tiêu Price**. Trong đó gồm **4 biến số**, **13 biến phân loại** và **1 cột không có nghĩa (Unnamed)**.

Bên dưới là bảng mô tả các thuộc tính trong bộ dữ liệu.

Tên thuộc tính	Mô tả thuộc tính	Loại biến	Kiểu giá trị
Unnamed: 0	Số chỉ mục (sẽ được loại bỏ).		Int64
Brand	Thương hiệu của laptop. Vd: HP, Lenovo, Dell...	Phân loại	Object
Name	Tên model cụ thể của laptop. Vd: Lenovo Ideapad Gaming 3 15IHU6 (82K101EEIN) Laptop...	Phân loại	Object
Price	Giá bán của laptop (theo đơn vị Indian Rupee).	Số	Int64
Processor_Name	Tên bộ xử lý được sử dụng. Vd: Intel Core i5, AMD Ryzen 5...	Phân loại	Object
Processor_Brand	Thương hiệu của bộ xử lý. Vd: Intel, AMD, MediaTek...	Phân loại	Object
RAM_Expandable	Cho biết RAM có mở rộng được hay không và ở mức độ nào. Vd: 12 GB Expandable...	Phân loại	Object

RAM	Dung lượng RAM được lắp sẵn. Vd: 8 GB, 16 GB...	Phân loại	Object
RAM_TYPE	Loại RAM được sử dụng. Vd: DDR4, LPDDR4X...	Phân loại	Object
Ghz	Tốc độ xung nhịp của bộ xử lý (theo đơn vị Ghz). Vd: 2.3 Ghz Processor...	Số	Object
Display_type	Loại màn hình được sử dụng. Vd: LCD, LED...	Phân loại	Object
Display	Kích thước màn hình của laptop (theo đơn vị inches).	Phân loại	Object
GPU	Kiểu GPU (card đồ họa). Vd: UHD, Iris Xe, Geforce RTX 3050 GPU 4GB...hoặc Integrated nếu là card tích hợp.	Phân loại	Object
GPU_Brand	Thương hiệu của GPU. Vd: NVIDIA, AMD, Intel...	Phân loại	Object
SSD	Dung lượng lưu trữ SSD (Solid State Drive). Vd: 512 GB SSD Storage, 1024 GB SSD Storage...	Phân loại	Object
HDD	Dung lượng lưu trữ HDD (Hard Disk Drive). Vd: No HDD, 1024 GB HDD Storage...	Phân loại	Object
Adapter	Công suất nguồn (theo đơn vị Watts).	Số	Object
Battery_life	Thời lượng pin dự kiến của laptop. Vd: Upto 7 Hrs Battery Life, Upto 10 Hrs Battery Life...	Số	Object

Trong số các thuộc tính này, có một số thuộc tính có kiểu dữ liệu trong bộ dữ liệu gốc khác với kiểu nó nên có. Ví dụ như biến Battery\_life, vì giá trị của nó được ghi dưới dạng chuỗi văn bản như là “Upto 7 Hrs Battery Life” nên nó được nhận diện ban đầu là kiểu dữ liệu object, nhưng thực tế nó nên là biến số và có kiểu dữ liệu là int hoặc float, chẳng hạn như là 7. Nhóm sẽ chuẩn hóa lại trong quá trình tiền xử lý ở sau.

Dưới đây là bảng thống kê các giá trị khuyết của các biến.

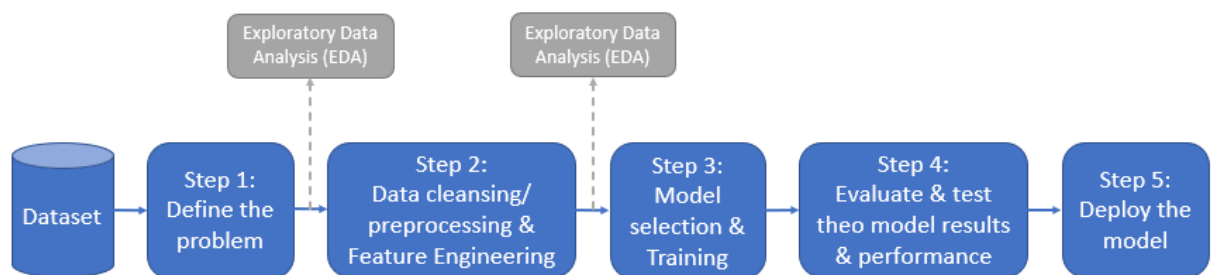
Tên thuộc tính	Số lượng bị khuyết
Brand	0
Name	0

Price	0
Processor_Name	0
Processor_Brand	0
RAM_Expandable	0
RAM	0
RAM_TYPE	0
Ghz	0
Display_type	0
Display	0
GPU	0
GPU_Brand	0
SSD	0
HDD	0
Adapter	0
Battery_life	418

### 3. PHƯƠNG PHÁP PHÂN TÍCH (Có thể đặt lại nội dung tiêu đề)

Phương pháp phân tích, có thể tiếp cận theo các cách sau

- Khuyến nghị tiếp cận theo quy trình hoặc các bước thực hiện phân tích dữ liệu đã được học. Hình:

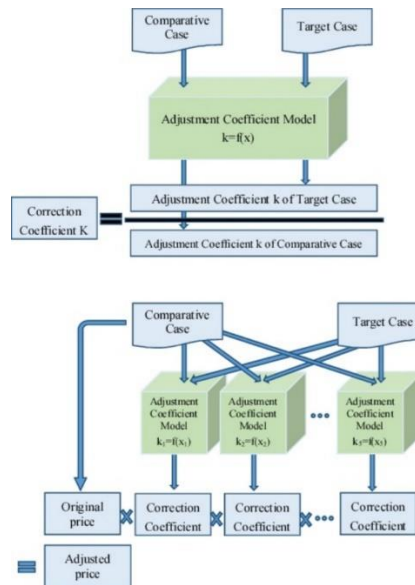


Hình 1. Quy trình PTDL.

Chú ý: Anh/chị vẽ hình đúng với quy trình các bước mà anh chị đã thực hiện đề tài. Hình trên chỉ mang tính chất tham khảo. **Không dùng lại hình trên.**

- Hoặc, theo style:
  - 1. Phương pháp (Methodology)
  - 2. Thiết kế thực nghiệm (Empirical design)
  - 3. Kết quả thực nghiệm (Empirical results)

- Hoặc, theo phương pháp, hoặc style khác.
- Nên có một hình thể hiện các bước thực hiện cả quá trình. Ví dụ:



Hình 2. Quy trình abc. Hình vừa phải không quá lớn hoặc quá nhỏ.

### 3.1. Tiêu đề Heading 2

Nội dung đoạn văn dùng Style BT

- Gạch đầu dòng này dùng Style G1.

### 3.2. Tiêu đề Heading 2

## 4. PHÂN TÍCH THĂM DÒ/SƠ BỘ (gợi ý)

Trình bày các phát hiện chính, không bắt buộc trình bày hết kết quả phân tích thăm dò các biến. Trình bày các biến quan trọng đã chọn lọc lại.



Giải thích:

## 5. KẾT QUẢ PHÂN TÍCH (gợi ý)

## 6. CHỈNH SỬA SAU BÁO CÁO

Style viết báo cáo

## 7. KẾT LUẬN

Tóm tắt lại quá trình và kết quả hiện có.

Tóm tắt kết quả đạt được?

Viết 10 dòng, hoặc nửa trang.

Viết tốt được **2 điểm**

Sinh viên trình bày không quá 10 trang cho tổng cộng 3 phần:

**Giới thiệu.**

**Nội dung.**

**Kết luận.**

(Không tính số trang tài liệu tham khảo)

**Chú ý:**

- Tại footer của báo cáo, nếu Họ tên SV dài có thể ghi tắt.  
Ví dụ: Nguyễn Thị Thu Trang => Nguyễn T.T. Trang, hoặc N.T.T. Trang
- Nếu nhóm nhiều thành viên, không đủ chỗ ghi tên tại footer thì có thể không cần ghi.

## TÀI LIỆU THAM KHẢO

- [1] Kaggle. Link: [Kaggle: Your Home for Data Science](#) (Truy cập 20/10/2024).
- [2] Kaggle. Link: [Laptop Dataset](#) (Truy cập 20/10/2024).
- [3] Kaggle. Link: [laptop\\_price\\_prediction.ipynb](#) (Truy cập 20/10/2024)
- [4] Smartprix. Link: [Smartprix - Best Online Comparison Shopping](#) (Truy cập 25/10/2024).
- [5]

*Chú ý: Đây là cách viết TLTK không đúng chuẩn. KHÔNG dùng định dạng này vào khóa luận tốt nghiệp và môn học khác.*

- [6] Họ và tên tác giả 1, Họ và tên tác giả 2, ..., Tên bài tham khảo, năm.
- [7] Tên trang web. Link: <http://abc.com.vn> (Ngày truy cập). Đối với tài liệu online **hạn chế/ko** tham khảo các blog công nghệ, wikipedia, facebook, youtube, mạng xã hội...

Trình bày không đúng style TLTK **– 1 điểm**.



## PHỤ LỤC PHÂN CÔNG NHIỆM VỤ

STT	Thành viên	Nhiệm vụ
1		
2		
3		

**(Chú ý: Ghi rõ từng nhiệm vụ chi tiết của mỗi thành viên).**

Khi làm bài thu hoạch SV phải thực hiện đúng các quy định sau:

1. Sử dụng đúng Template này. Không đúng thì **– 2 điểm**.
2. Trình bày đề tài tối thiểu là 05 trang và tối đa là 10 trang A4
  - Không tính Trang Bìa, và Phụ lục phân công công việc trong template.
  - Phụ lục code
3. Copy y chơn trên Internet thì gọi là đạo văn và **nhận 0 điểm** cho bài thu hoạch.
4. Nộp đúng hạn theo thông báo của GV.
5. **Không trình bày code** trong nội dung bài thu hoạch. Nếu muốn thì trình bày tại phụ lục.

**Sản phẩm:**

1. File word (thêm pdf) báo cáo, **ko in ra**, nộp file mềm.
2. Code Demo kết quả.
3. Slide thuyết trình.