



# BÁO CÁO THỰC TẬP TỐT NGHIỆP



Giảng viên hướng dẫn:

ThS. NGUYỄN VĂN CHỨC

Sinh viên thực hiện:

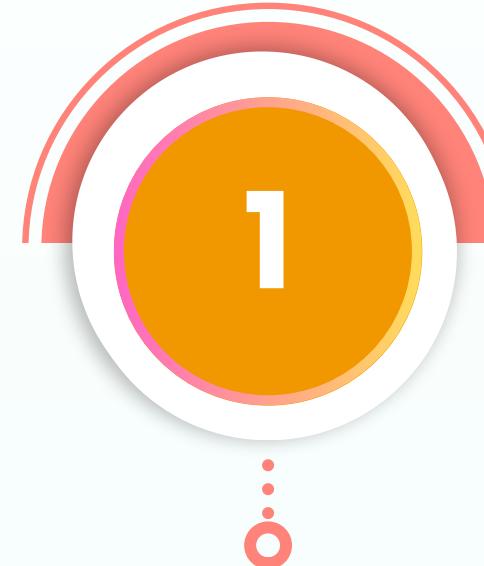
NGUYỄN THỊ DUYÊN  
NGUYỄN THỊ MÙI

**Đề tài:**

**Phân tích dữ liệu laptop của Thế Giới Di  
Động và xây dựng mô hình đề xuất laptop  
cho khách hàng**

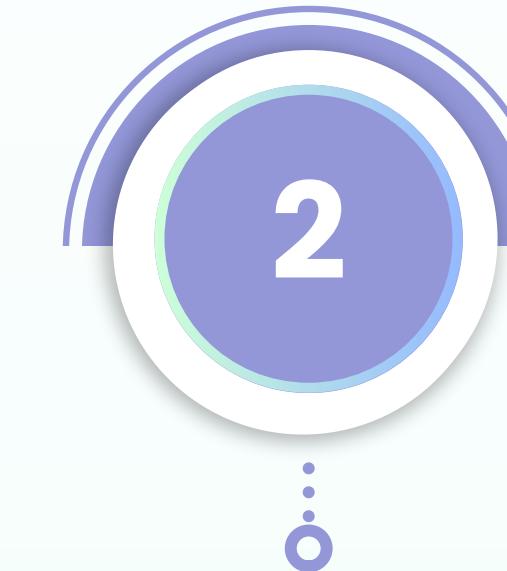
# TỔNG QUAN ĐỀ TÀI NGHIÊN CỨU

1



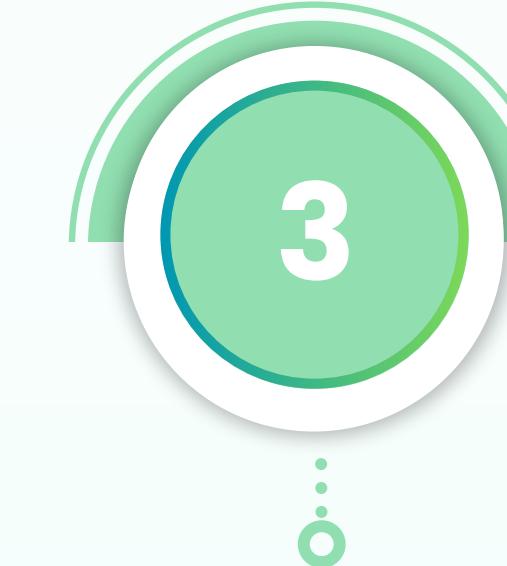
Giới thiệu

2



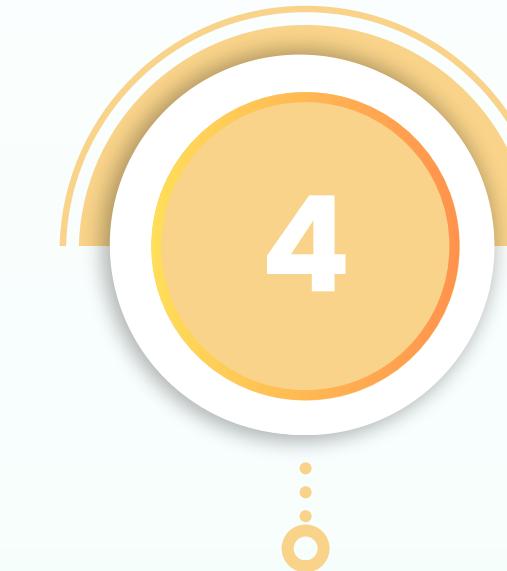
Xử lý dữ liệu

3



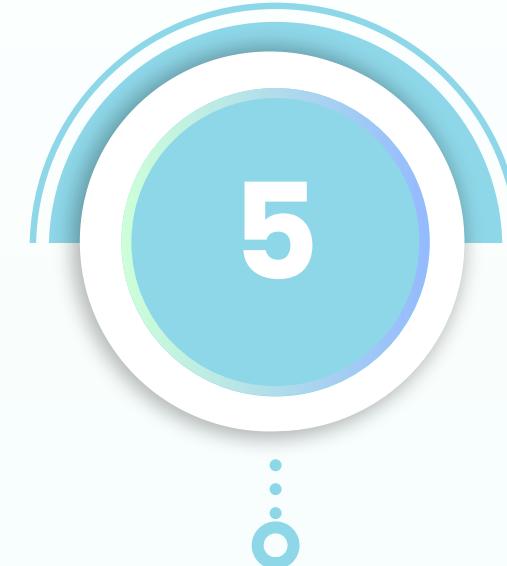
Trực quan hóa

4



Xây dựng mô hình

5



Kết luận và  
hướng phát triển





1

# GIỚI THIỆU

# GIỚI THIỆU ĐỀ TÀI

## Lý do chọn đề tài



### Mục tiêu đề tài



Phân tích khám phá, trực quan hóa dữ liệu

Nhóm các sản phẩm laptop có sự tương đồng

Dự đoán mức độ hài lòng của khách hàng

Đề xuất các mẫu laptop phù hợp

Đánh giá hiệu suất các mô hình

# GIỚI THIỆU ĐƠN VỊ THỰC TẬP

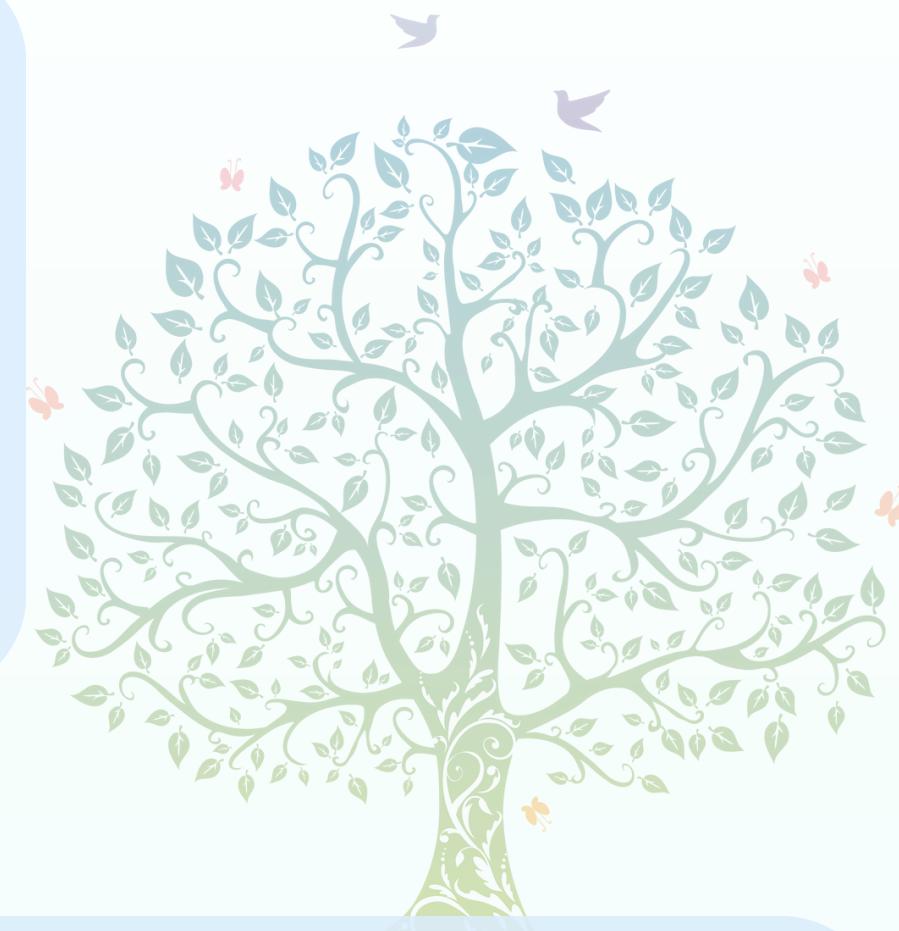
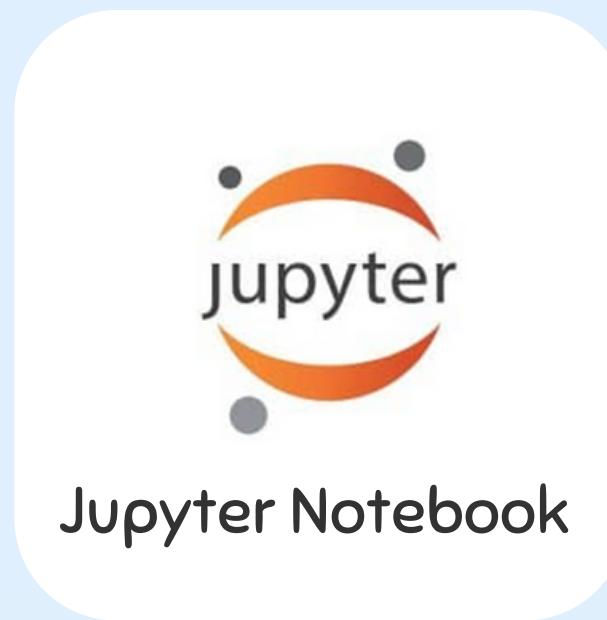
Là đơn vị chủ chốt của Tập đoàn Bưu chính Viễn thông Việt Nam, VNPT-IT luôn phấn đấu không ngừng nâng cao chất lượng sản phẩm dịch vụ về mọi mặt để trở thành thương hiệu có uy tín trong lĩnh vực công nghệ thông tin.

## VỊ TRÍ THỰC TẬP: DATA ANALYST

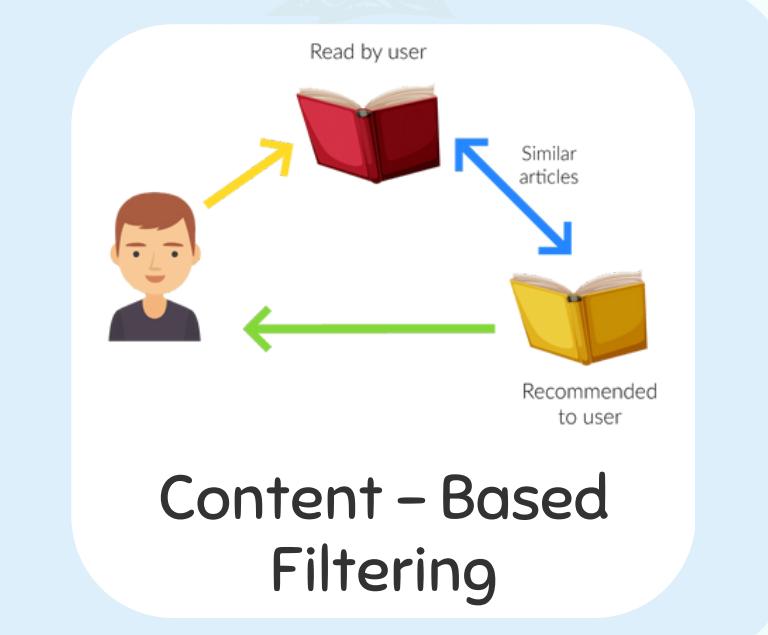
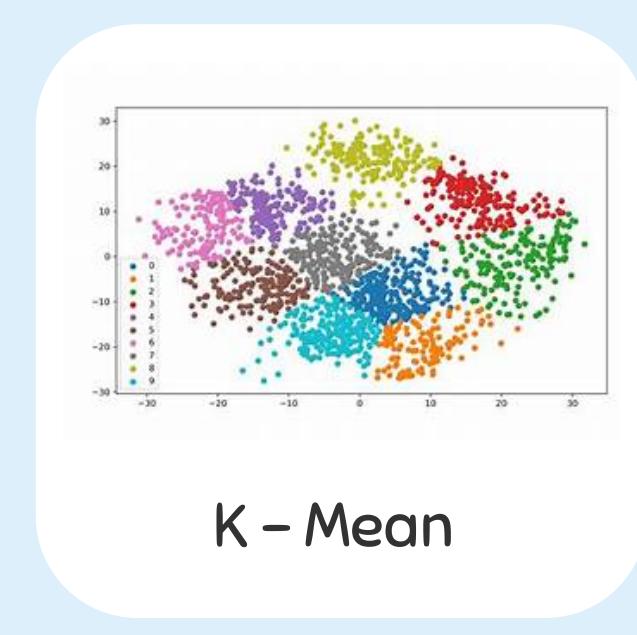
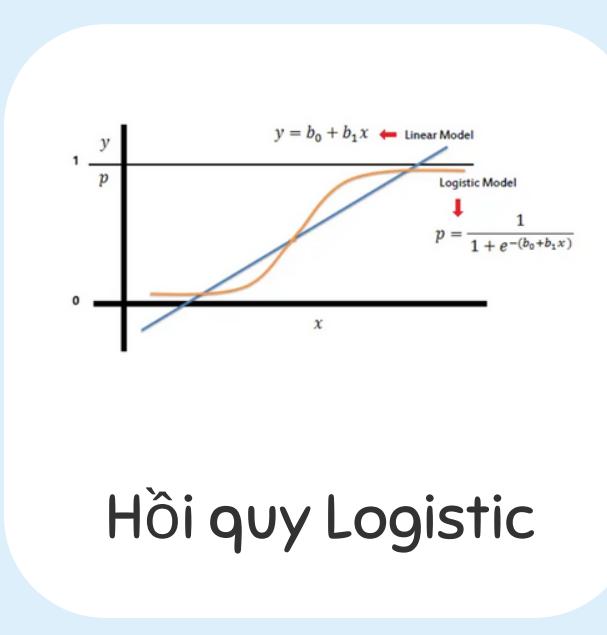
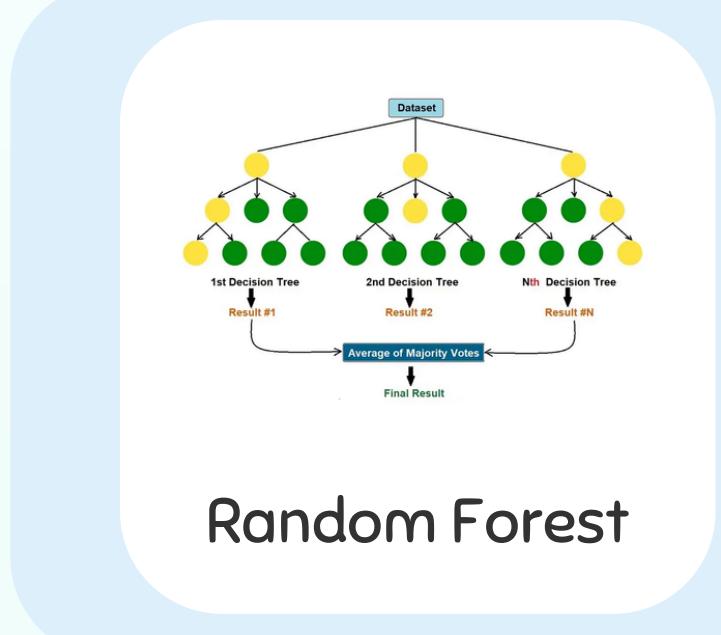


# GIỚI THIỆU

## Công cụ triển khai



## Thuật toán triển khai



# GIỚI THIỆU DỮ LIỆU

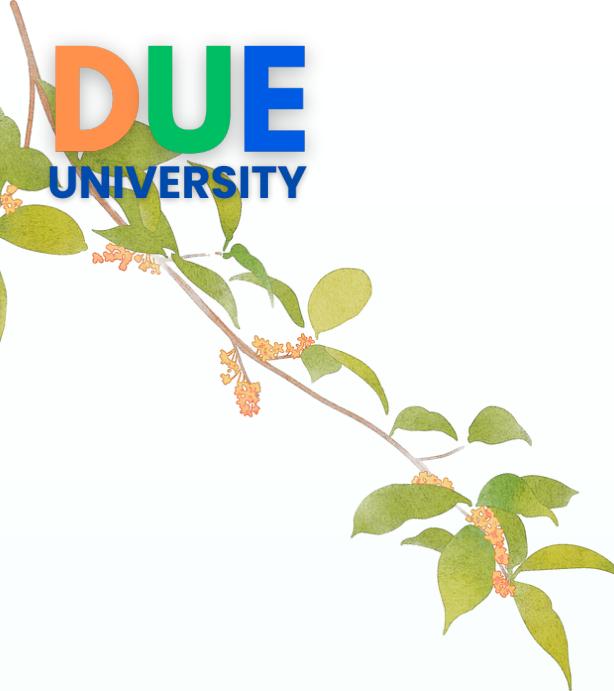
## Dữ liệu về Laptop

Nguồn dữ liệu: website Thế Giới Di Động

47 trường - 1027 bản ghi

Model	Processor	RAM	Storage	Price	Discount	Rating
HP 15 fc008AU R5 7430U (A6V8PA)	Radeon	RAM 16 GB	SSD 512 GB	13.690.000đ	16.190.000đ -9%	4.9 • Đã bán 8.2k
Asus Vivobook 15 OLED A1505ZA i5 12500H...	Intel Iris Xe	RAM 16 GB	SSD 512 GB	16.490.000đ	19.990.000đ -17%	Quà 2.490.000đ
Acer Aspire 7 A715 76 A1505ZA i5 12500H...	Intel UHD	RAM 16 GB	SSD 512 GB	13.990.000đ	16.990.000đ -12%	Quà 2.390.000đ
Lenovo IdeaPad Slim 3 15IRH10 i5 13420H...	Intel UHD	RAM 16 GB	SSD 512 GB	16.190.000đ	16.990.000đ -12%	Quà 400.000đ
Dell Inspiron 15 3520 i5 1235U (N515057W1)	Intel Iris Xe	RAM 16 GB	SSD 512 GB	16.490.000đ	16.990.000đ -12%	Quà 2.490.000đ
MSI Gaming Thin 15 B12UCX i5 12450H...	RTX 2050 4GB	RAM 16 GB	SSD 512 GB	15.690.000đ	18.990.000đ -17%	Quà 400.000đ

Link	Thương hiệu	Tên sản phẩm	Thông tin	Giá	Tên khách hàng	Rating	Hài lòng	Bình giá	RAM	Loại RAM	Tốc độ Bus	Hỗ trợ RAID	Ổ cứng	Công nghệ	Số nhân
<a href="https://www.Acer">https://www.Acer</a>	Laptop Acer	Bộ sưu tập	124900000	Phạm Dur	1	1	Máy để bàn	16 GB	DDR4X (4266 MHz)	Không hỗ trợ	512 GB SSD	AMD Ryzen	8		
<a href="https://www.Acer">https://www.Acer</a>	Laptop Acer	Bộ sưu tập	124900000	Lê Văn Đô	1	1	Mua sắm	16 GB	DDR4X (4266 MHz)	Không hỗ trợ	512 GB SSD	AMD Ryzen	8		
<a href="https://www.Acer">https://www.Acer</a>	Laptop Acer	Bộ sưu tập	124900000	Long Hoar	5	1	trong tầm	16 GB	DDR4X (4266 MHz)	Không hỗ trợ	512 GB SSD	AMD Ryzen	8		
<a href="https://www.Acer">https://www.Acer</a>	Laptop Acer	Bộ sưu tập	124900000	Hải Đăng	3	1	không có	16 GB	DDR4X (4266 MHz)	Không hỗ trợ	512 GB SSD	AMD Ryzen	8		
<a href="https://www.Acer">https://www.Acer</a>	Laptop Acer	Bộ sưu tập	124900000	Nguyễn H	4	1	Mới mua	16 GB	DDR4X (4266 MHz)	Không hỗ trợ	512 GB SSD	AMD Ryzen	8		
<a href="https://www.Acer">https://www.Acer</a>	Laptop Acer	Bộ sưu tập	124900000	Nguyễn Tr	2	0	Tôi mới m	8 GB	DDR4 2 kh	Từ 2400 M	32 GB	512 GB SSD	Intel Core	10	
<a href="https://www.Acer">https://www.Acer</a>	Laptop Acer	Bộ sưu tập	124900000	Tuyêt	5	0	máy ok	8 GB	DDR4 2 kh	Từ 2400 M	32 GB	512 GB SSD	Intel Core	10	
<a href="https://www.Acer">https://www.Acer</a>	Laptop Acer	Bộ sưu tập	124900000	Lương Thế	3	0	Máy tạm	8 GB	DDR4 2 kh	Từ 2400 M	32 GB	512 GB SSD	Intel Core	10	
<a href="https://www.Acer">https://www.Acer</a>	Laptop Acer	Bộ sưu tập	124900000	Công Ty Tr	5	1	Sản phẩm	8 GB	DDR4 2 kh	Từ 2400 M	32 GB	512 GB SSD	Intel Core	10	
<a href="https://www.Acer">https://www.Acer</a>	Laptop Acer	Bộ sưu tập	124900000	Thương	5	1	Tốt	8 GB	DDR4 2 kh	Từ 2400 M	32 GB	512 GB SSD	Intel Core	10	
<a href="https://www.Acer">https://www.Acer</a>	Laptop Acer	Bộ sưu tập	124900000	Anh Long	5	0	Rất tốt	8 GB	DDR4 2 kh	Từ 2400 M	32 GB	512 GB SSD	Intel Core	10	
<a href="https://www.Acer">https://www.Acer</a>	Laptop Acer	Bộ sưu tập	124900000	'VO PHAN'	4	0	Tốt	8 GB	DDR4 2 kh	Từ 2400 M	32 GB	512 GB SSD	Intel Core	10	
<a href="https://www.Acer">https://www.Acer</a>	Laptop Acer	Bộ sưu tập	124900000	Ngô Mỹ	4	0	Sé giới thi	8 GB	DDR4 2 kh	Từ 2400 M	32 GB	512 GB SSD	Intel Core	10	
<a href="https://www.Acer">https://www.Acer</a>	Laptop Acer	Bộ sưu tập	169900000	Dinh Võ Q	3	1	Laptop m	16 GB	DDR4 2 kh	3200 MHz	32 GB	512 GB SSD	Intel Core	8	
<a href="https://www.Acer">https://www.Acer</a>	Laptop Acer	Bộ sưu tập	169900000	Nguyễn Tr	3	1	Máy sài	16 GB	DDR4 2 kh	3200 MHz	32 GB	512 GB SSD	Intel Core	8	
<a href="https://www.Acer">https://www.Acer</a>	Laptop Acer	Bộ sưu tập	169900000	Huỳnh Vũ	3	1	Máy đẹp	16 GB	DDR4 2 kh	3200 MHz	32 GB	512 GB SSD	Intel Core	8	
<a href="https://www.Acer">https://www.Acer</a>	Laptop Acer	Bộ sưu tập	169900000	Nguyễn Tr	3	1	mình vừa	16 GB	DDR4 2 kh	3200 MHz	32 GB	512 GB SSD	Intel Core	8	
Số lượng	Tốc độ CPU	Tốc độ tối ưu NPU	Hiệu năng	Màn hình	Độ phân giải	Tần số quay	Độ phủ m	Công nghệ	Màn hình	Độ sáng	S Card màn	Công nghệ	DTS	Công giao	
16	1.8GHz	Turbo Boost 4.3 GHz	14"	WUXGA	60Hz	45%	NTSC	Tấm nền IPS						Card tích	I Acer Purified Voice
16	1.8GHz	Turbo Boost 4.3 GHz	14"	WUXGA	60Hz	45%	NTSC	Tấm nền IPS						Card tích	I Acer Purified Voice
16	1.8GHz	Turbo Boost 4.3 GHz	14"	WUXGA	60Hz	45%	NTSC	Tấm nền IPS						Card tích	I Acer Purified Voice
16	1.8GHz	Turbo Boost 4.3 GHz	14"	WUXGA	60Hz	45%	NTSC	Tấm nền IPS						Card tích	I Acer Purified Voice
16	1.8GHz	Turbo Boost 4.3 GHz	14"	WUXGA	60Hz	45%	NTSC	Tấm nền IPS						Card tích	I Acer Purified Voice
12	1.3GHz	Turbo Boost 4.4 GHz	15.6"	Full HD (160Hz)					TFT					Card tích	I Stereo speakers
12	1.3GHz	Turbo Boost 4.4 GHz	15.6"	Full HD (160Hz)					TFT					Card tích	I Stereo speakers
12	1.3GHz	Turbo Boost 4.4 GHz	15.6"	Full HD (160Hz)					TFT					Card tích	I Stereo speakers
12	1.3GHz	Turbo Boost 4.4 GHz	15.6"	Full HD (160Hz)					TFT					Card tích	I Stereo speakers
12	1.3GHz	Turbo Boost 4.4 GHz	15.6"	Full HD (160Hz)					TFT					Card tích	I Stereo speakers
12	1.3GHz	Turbo Boost 4.4 GHz	15.6"	Full HD (160Hz)					TFT					Card tích	I Stereo speakers
12	1.3GHz	Turbo Boost 4.4 GHz	15.6"	Full HD (160Hz)					TFT					Card tích	I Stereo speakers
12	2.1GHz	Turbo Boost 4.6 GHz	15.6"	Full HD (160Hz)					Tấm nền IPS					Card rời	- Acer Purified Voice
12	2.1GHz	Turbo Boost 4.6 GHz	15.6"	Full HD (160Hz)					Tấm nền IPS					Card rời	- Acer Purified Voice
12	2.1GHz	Turbo Boost 4.6 GHz	15.6"	Full HD (160Hz)					Tấm nền IPS					Card rời	- Acer Purified Voice
12	2.1GHz	Turbo Boost 4.6 GHz	15.6"	Full HD (160Hz)					Tấm nền IPS					Card rời	- Acer Purified Voice
12	2.1GHz	Turbo Boost 4.6 GHz	15.6"	Full HD (160Hz)					Tấm nền IPS					Card rời	- Acer Purified Voice
12	2.1GHz	Turbo Boost 4.6 GHz	15.6"	Full HD (160Hz)					Tấm nền IPS					Card rời	- Acer Purified Voice
Kết nối	kh Webcam	Đèn bàn	p	Tính năng	Tần suất	Tần số	Chất liệu	Thông tin	Hệ điều h	Thời điểm	Dài(mm)	Rộng(mm)	Dày(mm)	Trọng Lượ	
Wi-Fi 6 (802.11ax)	webcam	Không có đèn					Vỏ nhựa	3-cell, 50V	Windows	2024	318.2	225.5	17.8	1.4	
Wi-Fi 6 (802.11ax)	webcam	Không có đèn					Vỏ nhựa	3-cell, 50V	Windows	2024	318.2	225.5	17.8	1.4	
Wi-Fi 6 (802.11ax)	webcam	Không có đèn					Vỏ nhựa	3-cell, 50V	Windows	2024	318.2	225.5	17.8	1.4	
Wi-Fi 6 (802.11ax)	webcam	Không có đèn					Vỏ nhựa	3-cell, 50V	Windows	2024	318.2	225.5	17.8	1.4	
Bluetooth 5.2	HD webca	Không có đèn					Vỏ nhựa	3-cell, 40V	Windows	2022	362.9	241.26	19.9	1.7	
Bluetooth 5.2	HD webca	Không có đèn					Vỏ nhựa	3-cell, 40V	Windows	2022	362.9	241.26	19.9	1.7	
Bluetooth 5.2	HD webca	Không có đèn					Vỏ nhựa	3-cell, 40V	Windows	2022	362.9	241.26	19.9	1.7	
Bluetooth 5.2	HD webca	Không có đèn					Vỏ nhựa	3-cell, 40V	Windows	2022	362.9	241.26	19.9	1.7	
Bluetooth 5.2	HD webca	Không có đèn					Vỏ nhựa	3-cell, 40V	Windows	2022	362.9	241.26	19.9	1.7	



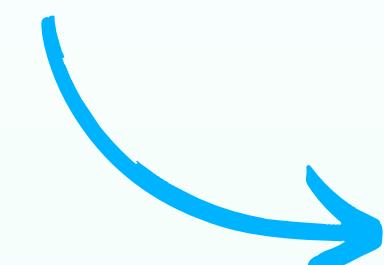
2

# XỬ LÝ DỮ LIỆU

# XỬ LÝ DỮ LIỆU

Import thư viện  
và dữ liệu

```
import numpy as np
import pandas as pd
data = pd.read_csv('data_laptop.csv')
```



Kiểm tra thông  
tin dữ liệu

data.info()

RangeIndex: 1027 entries, 0 to 1026			
	Column	Non-Null Count	Dtype
0	Link	1027 non-null	object
1	Thương hiệu	1027 non-null	object
2	Tên sp	1027 non-null	object
3	Thông tin	1005 non-null	object
4	Giá	1017 non-null	float64
5	Tên khách hàng	816 non-null	object
6	Rating	816 non-null	float64
7	Hài lòng	816 non-null	float64
8	Dánh giá	816 non-null	object
9	RAM	1027 non-null	object
10	Loại RAM	1027 non-null	object
11	Tốc độ Bus RAM	1027 non-null	object
12	Hỗ trợ RAM tối đa	1027 non-null	object
13	Ổ cứng	1027 non-null	object
14	Công nghệ CPU	1027 non-null	object
15	Số nhân	1027 non-null	int64
16	Số luồng	1027 non-null	object
17	Tốc độ CPU	1027 non-null	object
18	Tốc độ tối đa	1027 non-null	object
19	NPU	24 non-null	object
20	Hiệu năng xử lý AI (TOPS)	129 non-null	object
21	Màn hình	1027 non-null	object
22	Độ phân giải	1027 non-null	object
23	Tần số quét	1027 non-null	object
24	Độ phủ màu	557 non-null	object
25	Công nghệ màn hình	1026 non-null	object
26	Màn hình cảm ứng	40 non-null	object
27	Độ sáng SDR	11 non-null	object
28	Card màn hình	827 non-null	object
29	Công nghệ âm thanh	806 non-null	object
30	DTS	59 non-null	object
31	Cổng giao tiếp	827 non-null	object
32	Kết nối không dây	827 non-null	object
33	Webcam	821 non-null	object
34	Đèn bàn phím	827 non-null	object
35	Tính năng khác	503 non-null	object
36	Tàn nhiệt	171 non-null	object
37	Khe đọc thẻ nhớ	224 non-null	object
38	Kích thước	827 non-null	object
39	Chất liệu	827 non-null	object
40	Thông tin Pin	827 non-null	object
41	Hệ điều hành	827 non-null	object
42	Thời điểm ra mắt	824 non-null	float64
43	Dài(mm)	827 non-null	float64
44	Rộng(mm)	827 non-null	float64
45	Dày(mm)	827 non-null	float64
46	Trọng Lượng(kg)	827 non-null	float64
	dtypes:	float64(8), int64(1), object(38)	

Kiểm tra các cột  
có giá trị duy nhất

data.unique()

Link	340
Thương hiệu	11
Tên sp	341
Thông tin	155
Giá	171
Tên khách hàng	746
Rating	5
Hài lòng	2
Dánh giá	736
RAM	11
Loại RAM	28
Tốc độ Bus RAM	17
Hỗ trợ RAM tối đa	9
Ổ cứng	28
Công nghệ CPU	83
Số nhân	10
Số luồng	13
Tốc độ CPU	33
Tốc độ tối đa	20
NPU	2
Hiệu năng xử lý AI (TOPS)	6
Màn hình	14
Độ phân giải	27
Tần số quét	10
Độ phủ màu	5
Công nghệ màn hình	28
Màn hình cảm ứng	1
Độ sáng SDR	2
Card màn hình	26
Công nghệ âm thanh	24
DTS	1
Cổng giao tiếp	8
Kết nối không dây	10
Webcam	8
Đèn bàn phím	12
Tính năng khác	12
Tàn nhiệt	10
Khe đọc thẻ nhớ	3
Kích thước	83
Chất liệu	11
Thông tin Pin	42
Hệ điều hành	8
Thời điểm ra mắt	6
Dài(mm)	50
Rộng(mm)	60
Dày(mm)	40
Trọng Lượng(kg)	55

Kiểm tra giá trị  
null của mỗi cột

data.isnull().sum()

Link	0
Thương hiệu	0
Tên sp	0
Thông tin	22
Giá	10
Tên khách hàng	211
Rating	211
Hài lòng	211
Dánh giá	211
RAM	0
Loại RAM	0
Tốc độ Bus RAM	0
Hỗ trợ RAM tối đa	0
Ổ cứng	0
Công nghệ CPU	0
Số nhân	0
Số luồng	0
Tốc độ CPU	0
Tốc độ tối đa	0
NPU	1003
Hiệu năng xử lý AI (TOPS)	898
Màn hình	0
Độ phân giải	0
Tần số quét	0
Độ phủ màu	470
Công nghệ màn hình	1
Màn hình cảm ứng	987
Độ sáng SDR	1016
Card màn hình	200
Công nghệ âm thanh	221
DTS	968
Cổng giao tiếp	200
Kết nối không dây	200
Webcam	206
Đèn bàn phím	200
Tính năng khác	524
Tàn nhiệt	856
Khe đọc thẻ nhớ	803
Kích thước	200
Chất liệu	200
Thông tin Pin	200
Hệ điều hành	200
Thời điểm ra mắt	203
Dài(mm)	200
Rộng(mm)	200
Dày(mm)	200
Trọng Lượng(kg)	200

# XỬ LÝ DỮ LIỆU

```
data.duplicated().sum()  
0
```



```
df_clean = data.drop(['link', 'thông tin', 'NPU', 'Hiệu năng xử lý AI (TOPS)', 'Tính năng khác',  
                     'Màn hình cảm ứng', 'Độ sáng SDR', 'DTS', 'Tản nhiệt', 'Khe đọc thẻ nhớ', 'Độ phủ màu',  
                     'Kích thước', 'Dày(mm')], axis=1)  
  
df = df_clean.dropna()
```

```
df.shape  
(793, 34) ✓
```

Thương hiệu	0
Tên sp	0
Giá	0
Tên khách hàng	0
Rating	0
Hài lòng	0
Đánh giá	0
RAM	0
Loại RAM	0
Tốc độ Bus RAM	0
Hỗ trợ RAM tối đa	0
Ổ cứng	0
Công nghệ CPU	0
Số nhân	0
Số luồng	0
Tốc độ CPU	0
Tốc độ tối đa	0
Màn hình	0
Độ phân giải	0
Tần số quét	0
Công nghệ màn hình	0
Card màn hình	0
Công nghệ âm thanh	0
Cổng giao tiếp	0
Kết nối không dây	0
Webcam	0
Đèn bàn phím	0
Chất liệu	0
Thông tin Pin	0
Hệ điều hành	0
Thời điểm ra mắt	0
Dài(mm)	0
Rộng(mm)	0
Trọng Lượng(kg)	0

3

# TRỰC QUAN HÓA DỮ LIỆU

# TRỰC QUAN HÓA DỮ LIỆU

Tổng Doanh Thu  
**14 tỷ**

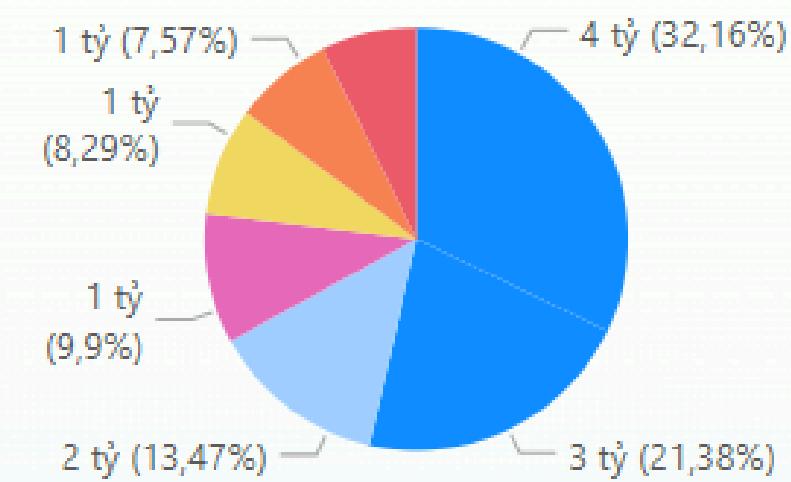
**124**

Tổng sản phẩm

**793**

Tổng số đơn đã bán

Tổng Doanh thu theo Thương hiệu



- Asus
- Hp
- Acer
- Dell
- Lenovo
- MSI
- MacBook

**17,50 triệu**

Giá TB

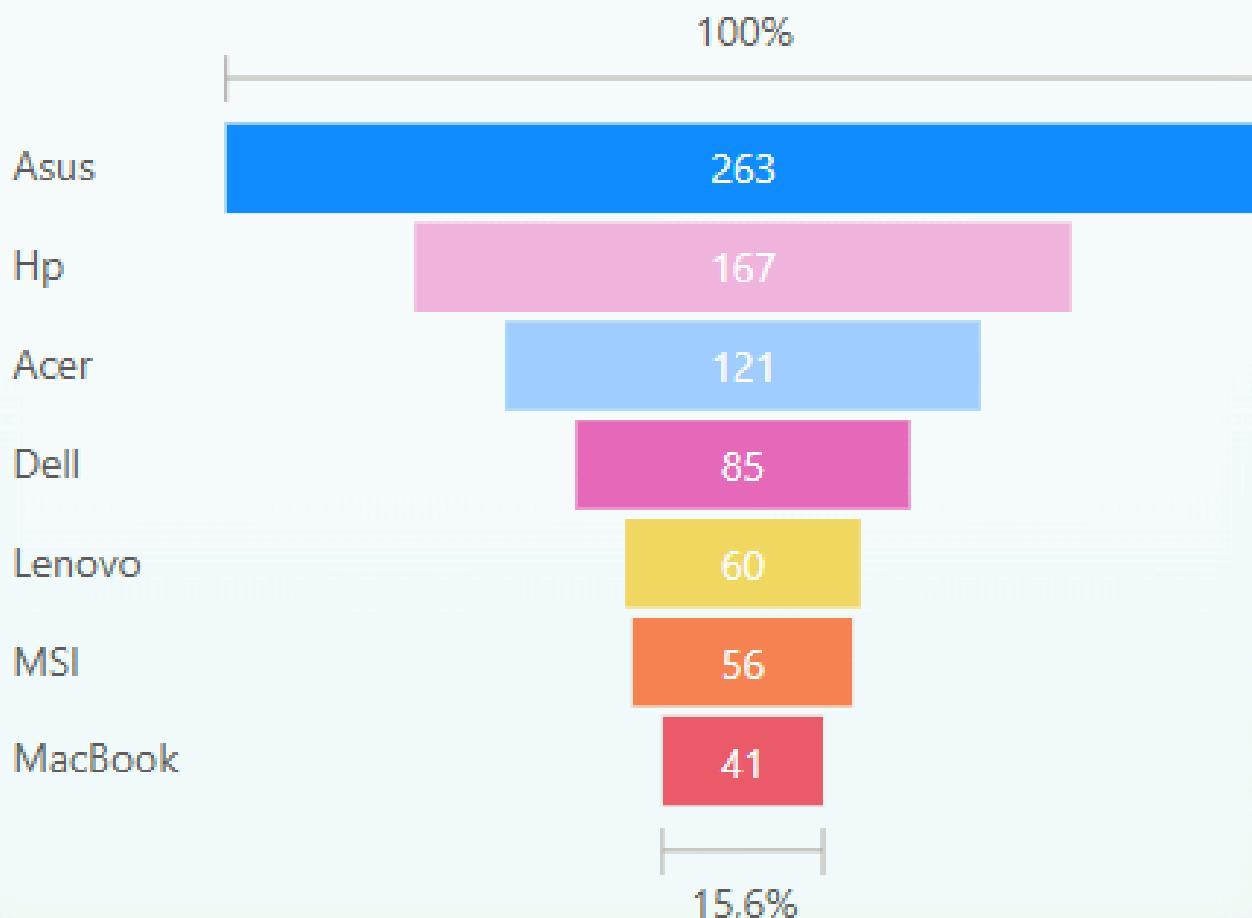
**3,96**

Rating TB

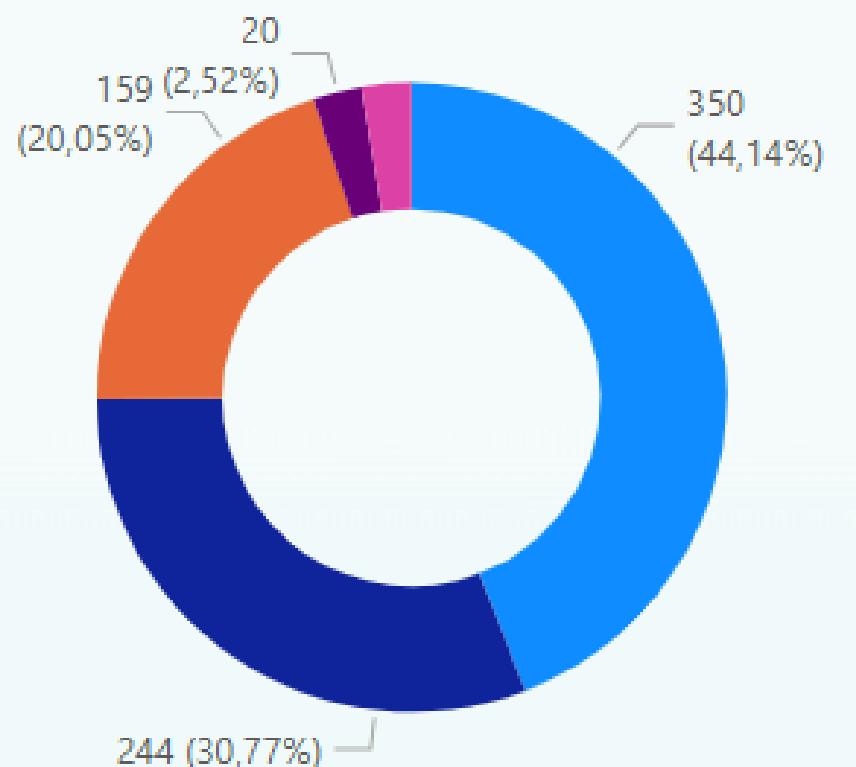
**0,59**

Hài lòng TB

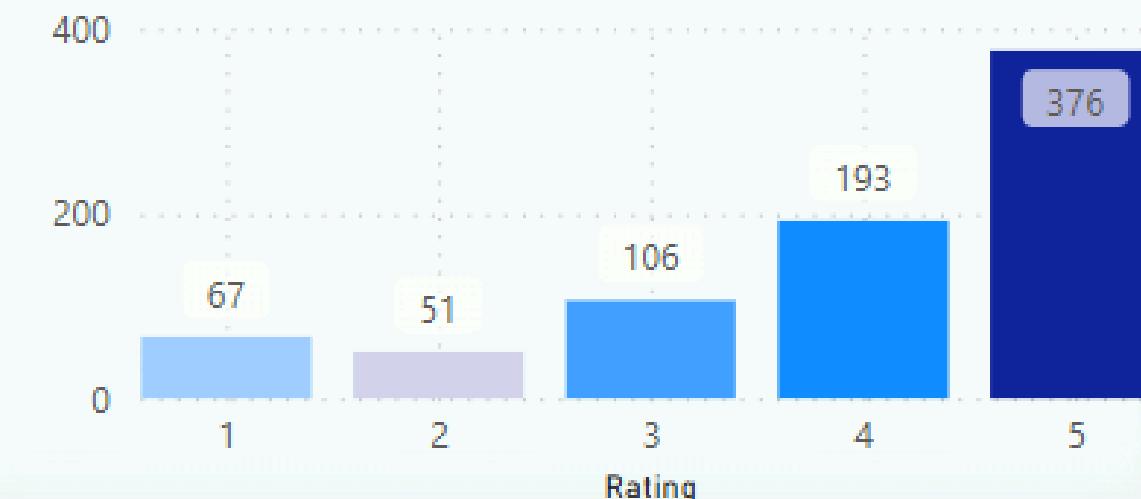
Tổng số lượng sản phẩm bán ra theo Thương hiệu



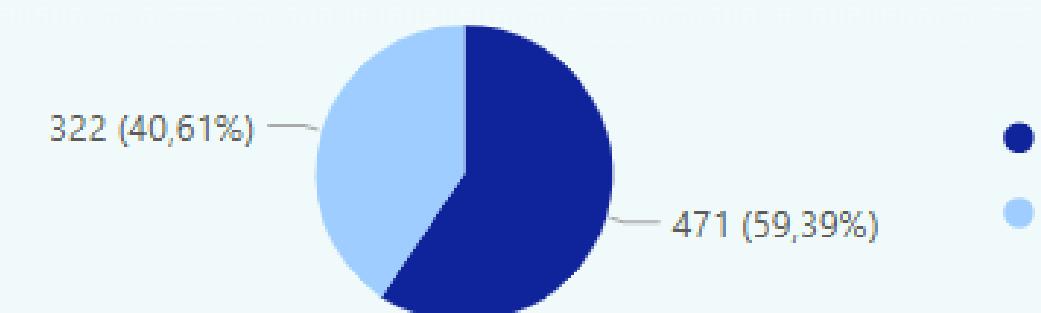
Tổng số lượng sản phẩm bán ra theo Năm



Rating

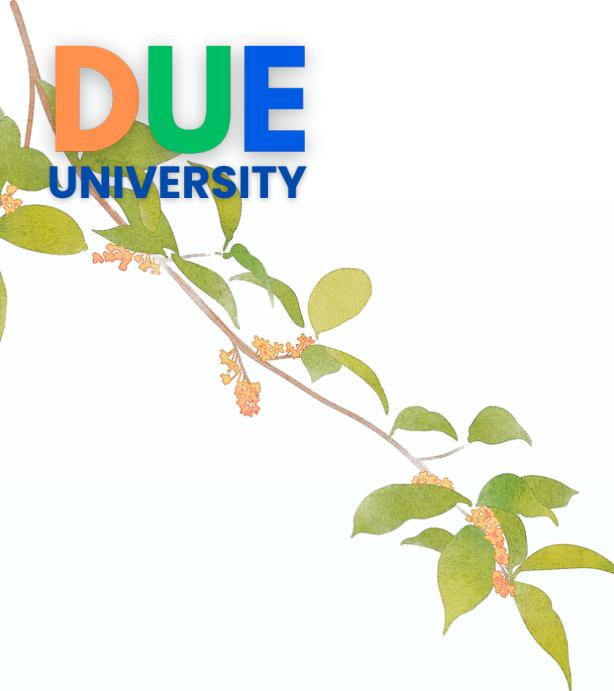


Hài lòng



# TOP SP ĐƯỢC ƯA CHUỘNG NHẤT DỰA TRÊN CHỈ SỐ KẾT HỢP RATING & MỨC ĐỘ HÀI LÒNG

Thương hiệu	Tên sp	Rating TB	Hài lòng TB
<input type="checkbox"/> Acer	Laptop MSI Gaming Thin 15 B12UCX i5 12450H/16GB/512GB/4GB RTX2050/144Hz/Balo/Win11 (2046VN)	5,00	1,00
<input type="checkbox"/> Asus	Laptop MSI Gaming Sword 16 HX B14VFKG i7 14700HX/16GB/1TB/8GB RTX4060/240Hz/Balo/Win11 (045VN)	5,00	1,00
<input type="checkbox"/> Dell	Laptop Lenovo Gaming LOQ 15ARP9 R7 7435HS/24GB/512GB/6GB RTX4050/144Hz/Win11 (83JC0040VN)	5,00	1,00
<input type="checkbox"/> Hp	Laptop HP 240 G9 i3 1215U/8GB/512GB/Win11 (AG2J5AT)	5,00	1,00
<input type="checkbox"/> Lenovo	Laptop HP 15 fd0303TU i3 1315U/8GB/512GB/Win11 (A2NL4PA)	5,00	1,00
<input type="checkbox"/> MacBook	Laptop Dell Vostro 15 3520 i5 1235U/16GB/512GB/120Hz/OfficeHS/KYHD/Win11 (i5U165W11GRU)	5,00	1,00
<input type="checkbox"/> MSI	Laptop Dell Inspiron 15 3530 i7 1355U/16GB/512GB/120Hz/OfficeHS/Win11 (71043888)	5,00	1,00
	Laptop Dell Inspiron 15 3530 i7 1355U/16GB/1TB/120Hz/OfficeHS/Win11 (P16WD)	5,00	1,00
	Laptop Asus Zenbook S 14 UX5406SA Ultra 7 258V/32GB/1TB/120Hz/Túi/OfficeHS/Win11 (PV140WS)	5,00	1,00
	Laptop Asus Vivobook 15 X1504ZA i5 1235U/16GB/1TB/Win11 (NJ1528W)	5,00	1,00
	Laptop Asus TUF Gaming A15 FA506NC R5 7535HS/16GB/512GB/4GB RTX3050/144Hz/Win11 (HN017W)	5,00	1,00
	Laptop Asus Gaming TUF A15 FA507NUR R7 7435HS/16GB/512GB/6GB RTX4050/144Hz/Win11 (LP057W)	5,00	1,00
	Laptop Apple MacBook Pro 16 inch M4 Pro 24GB/1TB	5,00	1,00
	Laptop Acer Aspire Lite 14 51M 36PN i3 1215U/8GB/512GB/Win11 (NX.KTWSV.001)	5,00	1,00
	Laptop Acer Aspire Lite 14 51M 36MH i3 1215U/8GB/256GB/Win11 (NX.KTVSV.001)	5,00	1,00
	<b>Tổng</b>	<b>5,00</b>	<b>1,00</b>
<b>Thời điểm ra m...</b>			
<input type="checkbox"/> 2020			
<input type="checkbox"/> 2021			
<input type="checkbox"/> 2022			
<input type="checkbox"/> 2023			
<input checked="" type="checkbox"/> 2024			



4

# XÂY DỰNG MÔ HÌNH

# CHUẨN BỊ DỮ LIỆU ĐỂ DỰ ĐOÁN

1.Import dữ liệu

3.Mã hóa các biến dạng chuỗi

5.Cân bằng dữ liệu

4.Sử dụng ma trận tương quan và lựa chọn ngưỡng > 0.05  
→ Số biến được chọn: 16

2.Chuyển đổi dữ liệu

4.Giảm chiều dữ liệu

6.Chuẩn hóa dữ liệu

## 5. Sử dụng SMOTE để cân bằng dữ liệu

Dữ liệu trước khi cân bằng: Hải lỏng

1	471
0	322

Name: count, dtype: int64

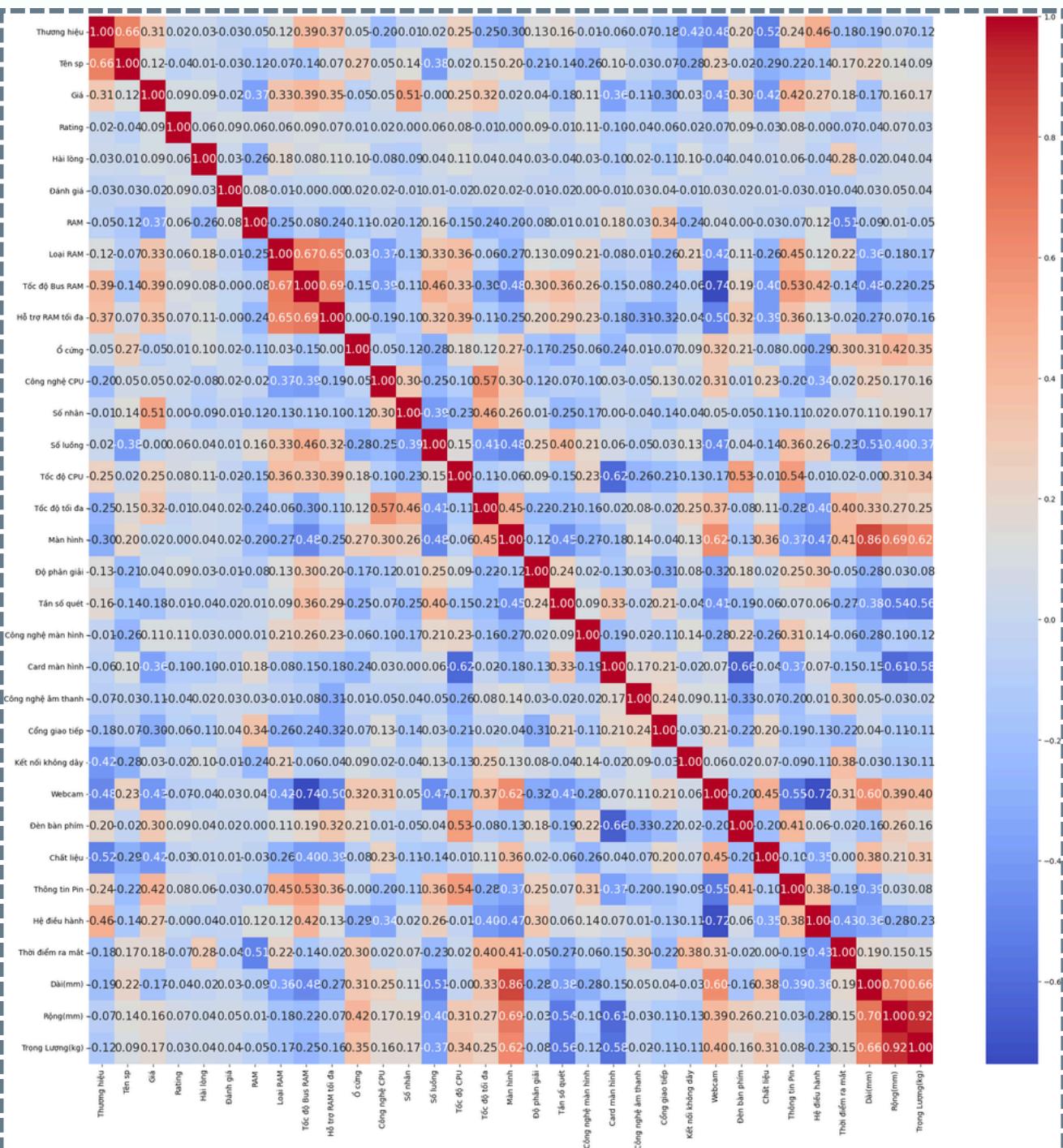
Dữ liệu sau khi cân bằng: Hải lỏng

0	471
1	471

Name: count, dtype: int64

	Hải lỏng
Hải lỏng	1.000000
Thời điểm ra mắt	0.284284
Loại RAM	0.181476
Kết nối không dây	0.131325
Tốc độ CPU	0.113752
Hỗ trợ RAM tối đa	0.110990
Ổ cứng	0.105996
Giá	0.088140
Tốc độ Bus RAM	0.071873
Thông tin Pin	0.060612
Rating	0.057633
Đèn bàn phím	0.044724
Số lượng	0.043647
Tốc độ tối đa	0.042390
Trọng Lượng(kg)	0.041872
Rộng(mm)	0.041736
Màn hình	0.038970
Đánh giá	0.031896
Thương hiệu	0.029100
Độ phân giải	0.026779
Công nghệ màn hình	0.026717
Công nghệ âm thanh	0.021434
Tên sp	0.012586
Chất liệu	0.006143
Dài(mm)	-0.016309
Tần số quét	-0.042645
Webcam	-0.042859
Hệ điều hành	-0.044802
Công nghệ CPU	-0.081351
Số nhân	-0.092689
Card màn hình	-0.101841
Công giao tiếp	-0.109659
RAM	-0.258379

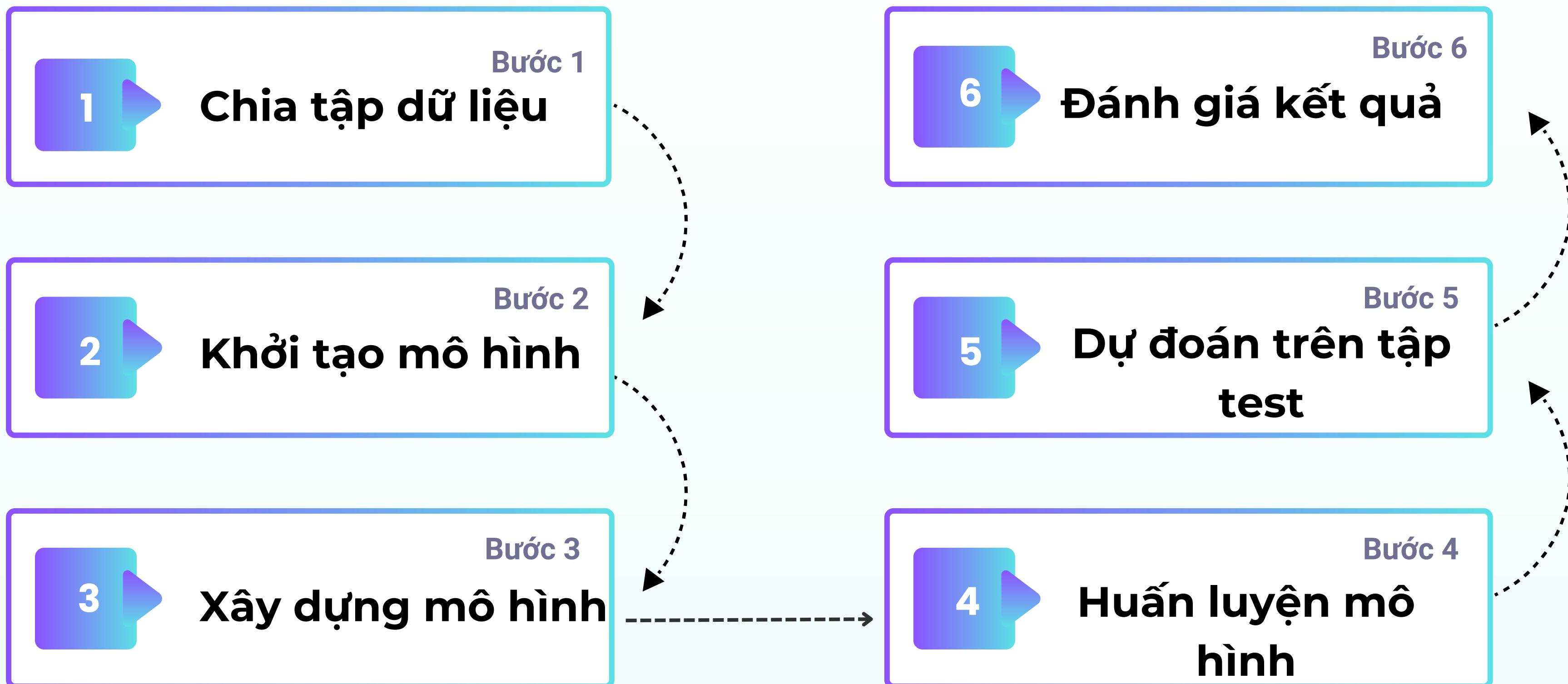
dtype: float64



## 6. Chuẩn hóa Min-max

	Giá	Rating	RAM	Loại RAM	Tốc độ Bus RAM	Hỗ trợ RAM tối đa	Ổ cứng	Công nghệ CPU	Số nhân	Tốc độ CPU	Card màn hình	Công giao tiếp	Kết nối không dây	Thông tin Pin	Thời điểm ra mắt
5	0.065574	0.25	1.0	0.333333	1.000000	0.375	0.368421	0.461538	0.444444	0.047619	0.904762	0.285714	0.000000	0.138889	0.50
361	0.188525	0.50	0.0	0.444444	0.153846	0.250	0.315789	0.564103	0.555556	0.476190	0.095238	0.142857	0.846154	0.277778	1.00
52	0.040984	1.00	1.0	0.333333	0.153846	0.375	0.578947	0.076923	0.333333	0.190476	0.666667	0.285714	0.692308	0.277778	0.75
147	0.339344	0.75	0.0	0.777778	0.923077	1.000	0.263158	0.205128	0.333333	0.238095	0.428571	0.428571	0.692306	1.000000	0.50
186	0.229508	0.75	0.0	0.277778	0.153846	0.375	0.052632	0.564103	0.555556	0.476190	0.095238	0.285714	0.153846	0.555556	0.75

# XÂY DỰNG MÔ HÌNH DỰ ĐOÁN



# Mô hình Hồi quy Logistic

Chia tập dữ liệu: 80% training - 20% testing

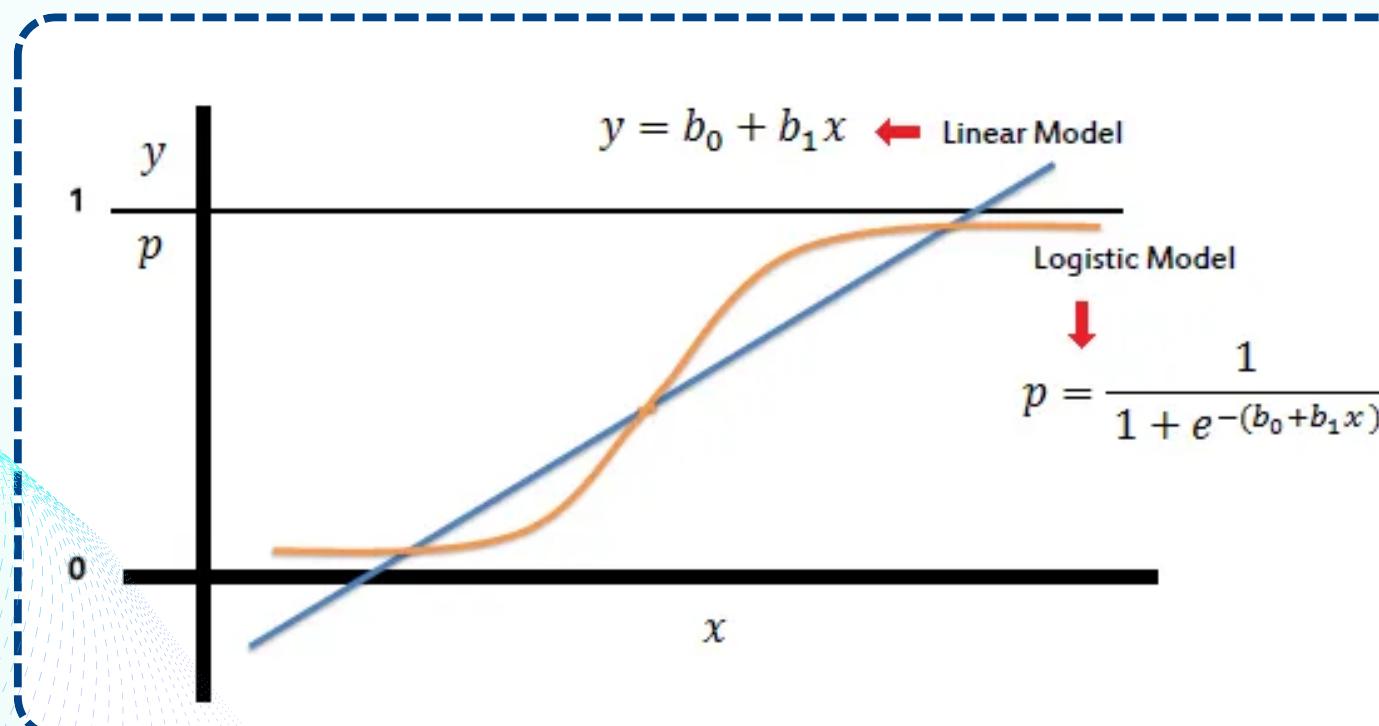
## Đặc điểm xây dựng



**penalty = L2**

**max\_iter = 100**

**tol = 1e-4(0.0001)**



Tên cột	Giá trị
$b_0$	-1.39956951
Giá	0.04456671
Rating	0.38176152
RAM	-0.6194309
Loại RAM	-0.03241243
Tốc độ Bus RAM	0.63179232
Hỗ trợ RAM tối đa	0.12414764
Ổ cứng	0.30679516
Công nghệ CPU	-0.28307281
Số nhân	-1.11243086
Tốc độ CPU	0.3102352
Card màn hình	0.06881999
Cổng giao tiếp	0.08565513
Kết nối không dây	-0.15913699
Thông tin Pin	0.22942948
Thời điểm ra mắt	2.14058529

- Ảnh hưởng mạnh:  $|trọng số| \geq 1.0$

- Ảnh hưởng vừa phải:  $|trọng số|$  từ 0.25 đến <1.0

- Ảnh hưởng nhỏ:  $|trọng số| < 0.25$

# Mô hình Random Forest

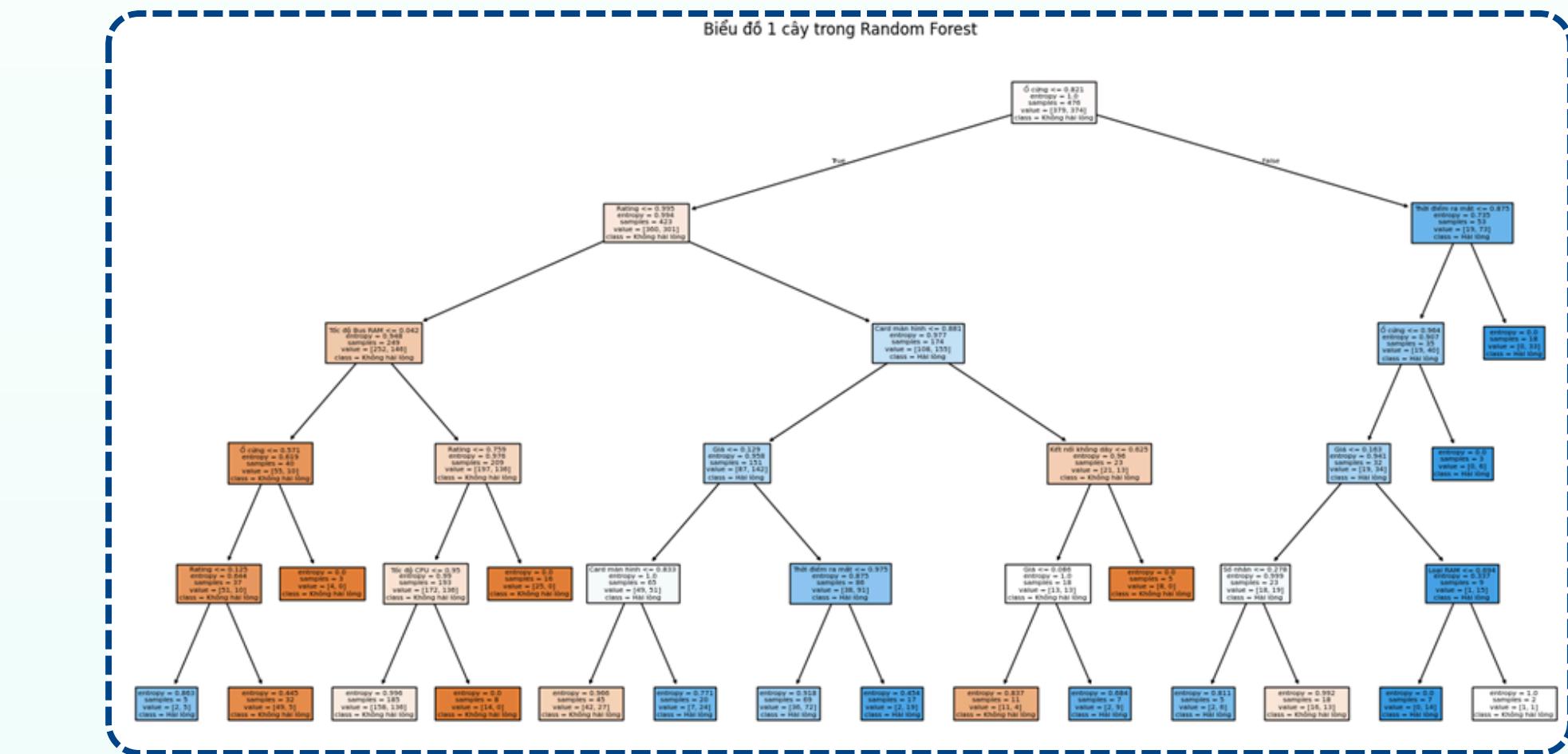
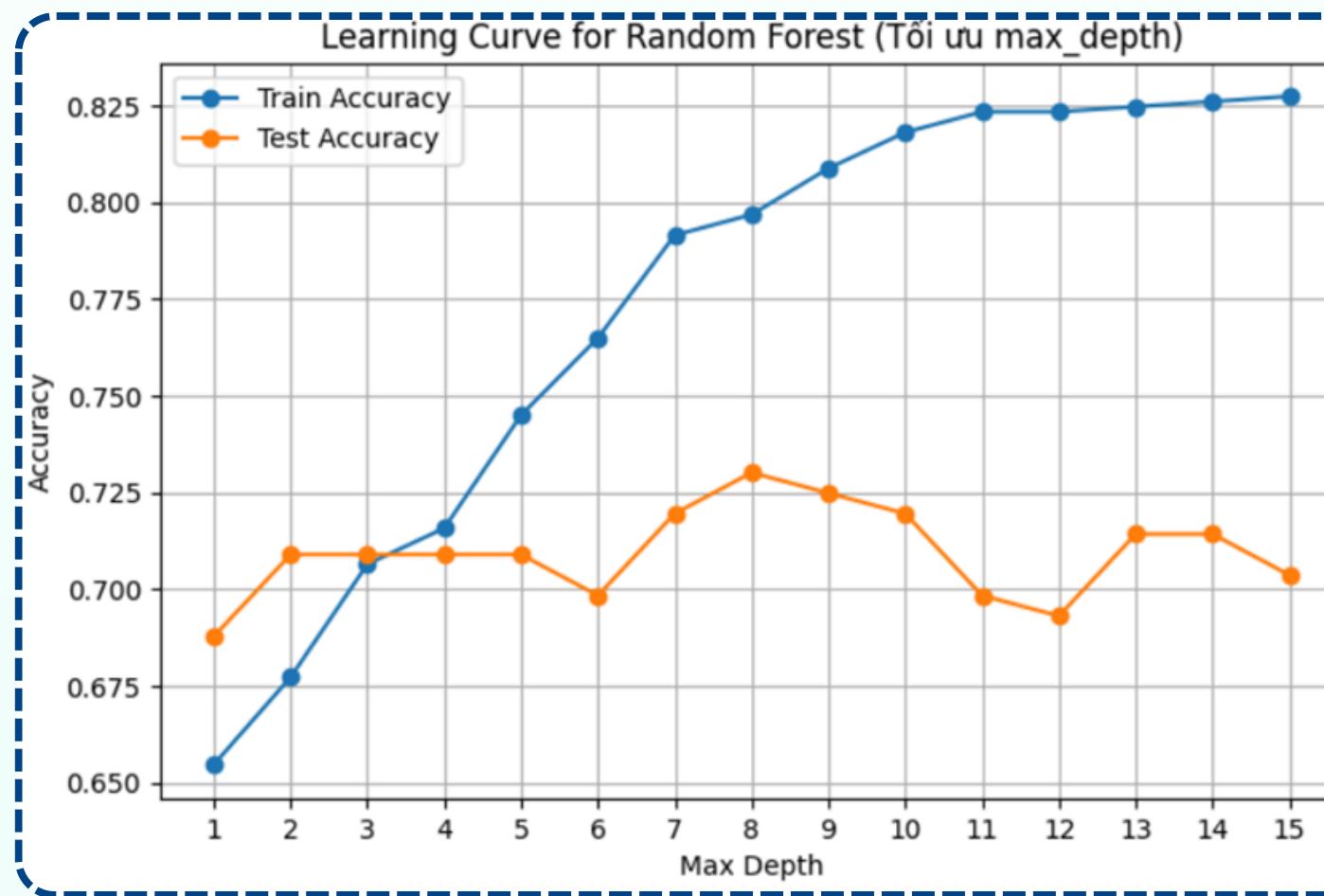
Chia tập dữ liệu: 80% training - 20% testing

## Đặc điểm xây dựng

criterion = "entropy"

max\_depth = 8

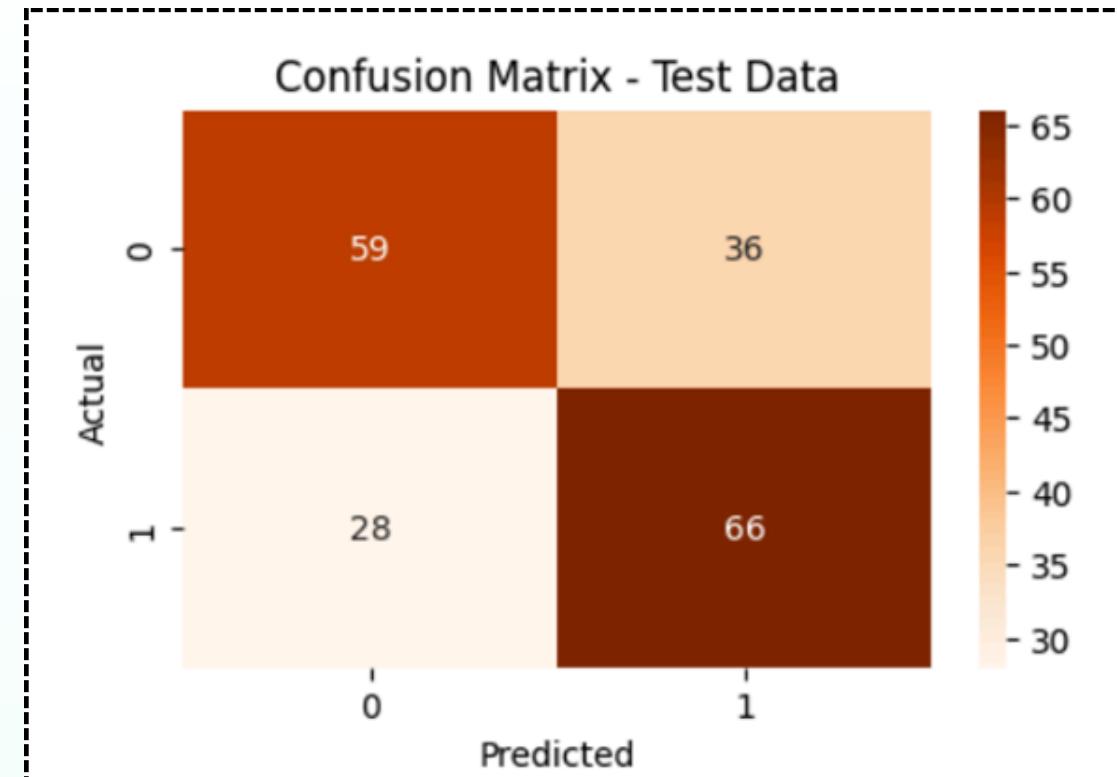
n\_estimators = 99



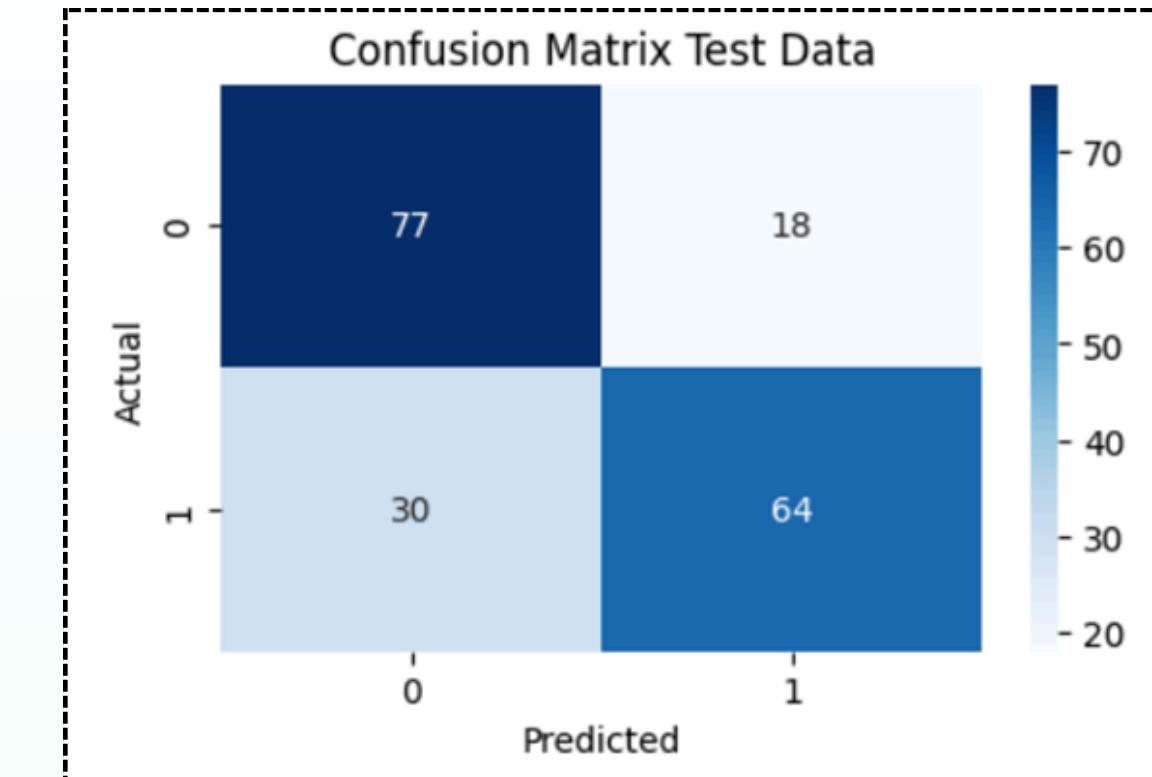


# SO SÁNH HAI MÔ HÌNH DỰ ĐOÁN

Hồi quy Logistic



Random Forests



## CHỈ SỐ

So sánh - Kết luận



## MÔ HÌNH

Accuracy

Precision

Recall

F1 - Score

AUC

Random Forests

**74.60%**

**78%**

**68%**

**73%**

**0.79**

Hồi quy Logistic

**66.14%**

**65%**

**70%**

**67%**

**0.72**

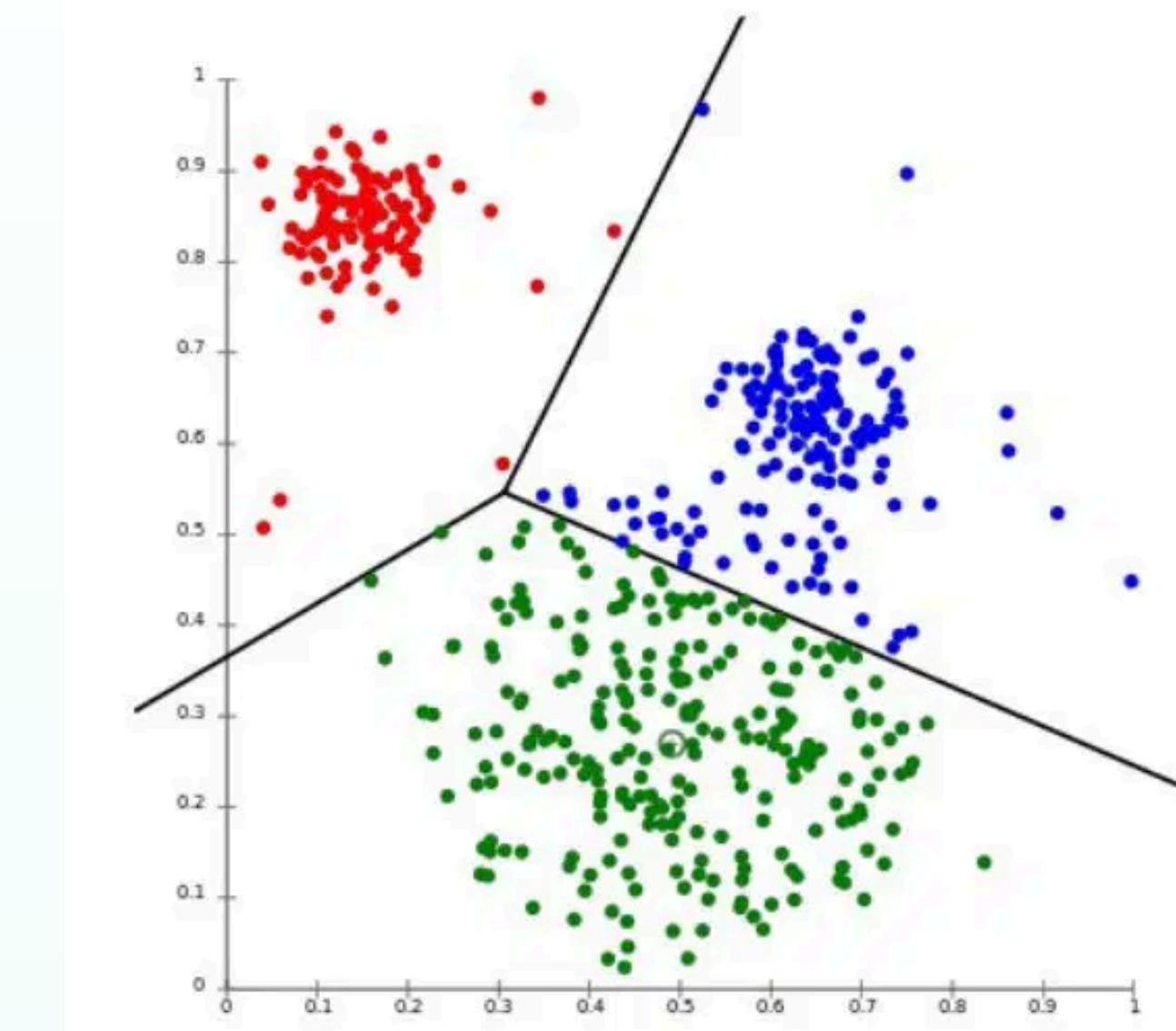
# K-MEANS PHÂN CỤM DỮ LIỆU

## 1. Ý tưởng thuật toán

- Phân chia dữ liệu thành K-cụm sao cho các điểm dữ liệu trong cụm có điểm tương đồng với nhau

## 2. Cách thức hoạt động

- Khởi tạo ngẫu nhiên k tâm cụm
- Gán nhãn cho các điểm dữ liệu
- Cập nhật tâm cụm
- Kiểm tra điều kiện dừng



# Chuẩn hóa dữ liệu

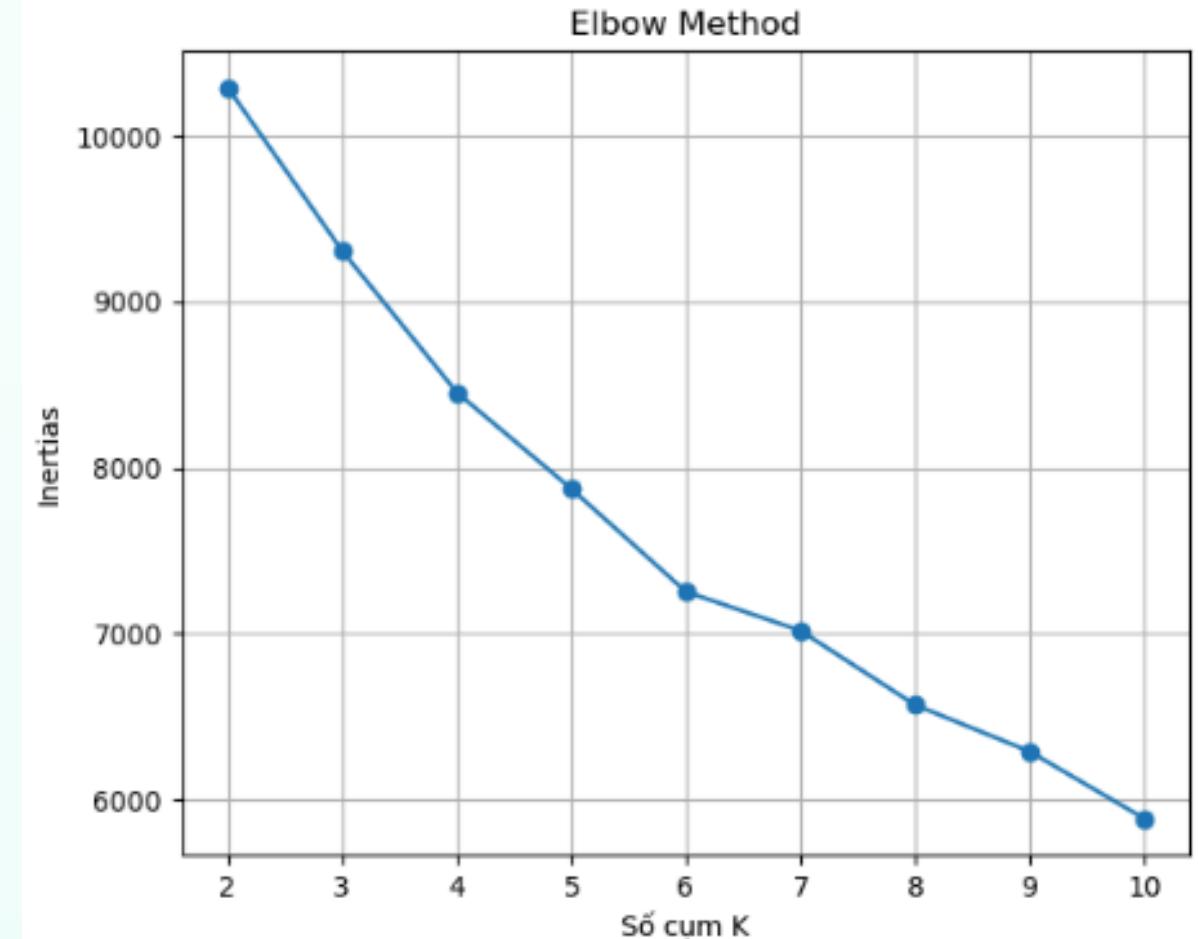
	Giá	Rating	Hài lòng	RAM	Loại RAM	Tốc độ Bus RAM	Hỗ trợ RAM tối đa	Ổ cứng	Công nghệ CPU	Số nhân	Tốc độ CPU	Card màn hình	Cổng giao tiếp	Kết nối không dây	Thông tin Pin	Thời điểm ra mắt
0	12490000	1.0	1	0	15	3	8	7	3	8	4	14	1	4	10	2024
1	12490000	1.0	1	0	15	3	8	7	3	8	4	14	1	4	10	2024
2	12490000	5.0	1	0	15	3	8	7	3	8	4	14	1	4	10	2024
3	12490000	3.0	1	0	15	3	8	7	3	8	4	14	1	4	10	2024
4	12490000	4.0	1	0	15	3	8	7	3	8	4	14	1	4	10	2024

```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
data_scaled = scaler.fit_transform(kmean)
```

Biểu đồ Elbow: điểm k là nơi xuất hiện điểm khuỷu tay => k=6 hoặc k=7

Biểu đồ Silhouette: điểm k tối ưu có giá trị Silhouette cao => k=9 hoặc k=10

=> K = 6



Sử dụng StandardScaler  
chuẩn hóa dữ liệu về phân  
phối chuẩn

## Tìm số cụm k tối ưu

## Đặc trưng từng cụm

Cụm	Số lượng	Đặc trưng	Nhóm khách hàng
0	204	<ul style="list-style-type: none"> <li>- Giá tầm trung</li> <li>- RAM 16 GB và ổ cứng 512 GB SSD NVMe - PCIe</li> <li>- Card tích hợp - Intel Iris Xe Graphics</li> <li>- Pin trung bình</li> </ul>	Sinh viên, nhân viên văn phòng, khách hàng không có nhu cầu về đồ họa, chơi game dung lượng nặng.
1	214	<ul style="list-style-type: none"> <li>- Giá rẻ nhất (13 triệu)</li> <li>- Ram 8 GB và ổ cứng 512 GB SSD NVMe PCIe (có thể nâng cấp)</li> <li>- Card tích hợp - Intel Iris Xe Graphics</li> </ul>	Sinh viên, người không có nhu cầu đồ họa, nhưng có ngân sách thấp
2	219	<ul style="list-style-type: none"> <li>- Giá trung bình cao (19.5 triệu)</li> <li>- Rating cao (4.17)</li> <li>- Hiệu năng cao với RAM 24 GB và CPU Alder Lake mạnh</li> <li>- Card rời - NVIDIA GeForce RTX 3050, 6 GB</li> <li>- Tốc độ Bus RAM cao và ổ cứng có thể nâng cấp.</li> </ul>	Khách hàng làm việc liên quan đến đồ họa, lập trình viên, nhu cầu chơi game, làm việc đa nhiệm cần hiệu năng tốt.

Cụm	Số lượng	Đặc trưng	Nhóm khách hàng
3	44	<ul style="list-style-type: none"> <li>- Laptop cao cấp với giá cao (31.2 triệu)</li> <li>- RAM mạnh 24 GB, loại RAM LPDDR5 tốc độ cao</li> <li>- Chip Intel Core i3 thế hệ 13 tiết kiệm điện</li> <li>- Card rời AMD</li> </ul>	Khách hàng thường xuyên làm việc di động cần mỏng nhẹ.
4	72	<ul style="list-style-type: none"> <li>- Giá trung bình thấp (13.7 triệu)</li> <li>- RAM 16 GB LPDDR5 tốc độ cao.</li> <li>- Ổ cứng 512 GB SSD NVMe PCIe(có thể nâng cấp)</li> <li>- Thời lượng pin tốt</li> </ul>	Sinh viên, người dùng mong muốn máy mượt, nhẹ, pin khỏe
5	40	<ul style="list-style-type: none"> <li>- Giá trung bình cao (23.4 triệu)</li> <li>- RAM 32 GB hiệu năng cao</li> <li>- Card tích hợp Apple GPU</li> <li>- Công nghệ CPU Apple M2 hiệu năng mạnh, tiết kiệm điện</li> </ul>	Khách hàng yêu thích dòng sản phẩm Apple, giao diện macOS Khách hàng làm việc trong lĩnh vực sáng tạo nội dung

# HỆ THỐNG ĐỀ XUẤT

## 1. Chuẩn bị dữ liệu

Thứ tự	Thương hiệu	Tên sp	Giá	Đánh giá	RAM	Loại RAM	Ổ cứng	Công nghệ CPU	Tốc độ CPU	Màn hình	Độ phân giải	Card màn hình	Chất liệu	Thông tin Pin	Hệ điều hành
0	Acer	Laptop Acer Aspire 3 A314-42P R3B3 R7 5700U/16...	12490000	Máy để chế độ sleep khoảng 12 tiếng sẽ bị nóng...	16 GB	LPDDR4X (Onboard)	512 GB SSD NVMe PCIe (Có thể tháo ra, lắp thanh...)	AMD Ryzen 7 - 5700U	1.8GHz	14"	WUXGA	Card tích hợp - AMD Radeon Graphics	Vỏ nhựa	3-cell, 50Wh	Windows 11 Home SL
1	Acer	Laptop Acer Aspire 3 A314-42P R3B3 R7 5700U/16...	12490000	Mua sản phẩm vào tháng 09/2024 đến tháng 02/20...	16 GB	LPDDR4X (Onboard)	512 GB SSD NVMe PCIe (Có thể tháo ra, lắp thanh...)	AMD Ryzen 7 - 5700U	1.8GHz	14"	WUXGA	Card tích hợp - AMD Radeon Graphics	Vỏ nhựa	3-cell, 50Wh	Windows 11 Home SL

Sử dụng các cột chứa thông tin mà khách hàng sẽ quan tâm

# HỆ THỐNG ĐỀ XUẤT

## 2. Xử lý văn bản

- Chuyển chữ hoa thành chữ thường
- Xóa ký tự đặc biệt
- Gộp các đơn vị số
- Xóa khoảng trắng
- Gộp các cột thành chuỗi văn bản

'acer rất tốt 8gb ddr4 2 khe 1 khe 8gb 1 khe rời 512gb ssd nvme pcie có thể tháo ra lắp thanh khác tối đa 1tb intel core i5 alder lake 1235u 1.3ghz 15.6 full hd 1920 x 1080 card tích hợp intel uhd graphics iris xe graphics chỉ hoạt động với ram kênh đôi vỏ nhựa 3cell 40wh windows 11 home sl'

Dữ liệu sau khi xử lý



# HỆ THỐNG ĐỀ XUẤT

## 3. Ma trận đặc trưng

- Sử dụng TF-IDF chuyển dữ liệu thành vector số học
- Tính độ tương đồng giữa các sản phẩm

```
from sklearn.metrics.pairwise import cosine_similarity
similarity_matrix = cosine_similarity(vector)
similarity_matrix

array([[1.          , 0.1474738 , 0.04711246, ..., 0.03877953, 0.03476593,
       0.07023735],
       [0.1474738 , 1.          , 0.04964471, ..., 0.09441334, 0.03597066,
       0.12041802],
       [0.04711246, 0.04964471, 1.          , ..., 0.01944744, 0.03487717,
       0.02480393],
       ...,
       [0.03877953, 0.09441334, 0.01944744, ..., 1.          , 0.2972698 ,
       0.54141652],
       [0.03476593, 0.03597066, 0.03487717, ..., 0.2972698 , 1.          ,
       0.37914813],
       [0.07023735, 0.12041802, 0.02480393, ..., 0.54141652, 0.37914813,
       1.        ]])
```



Càng gần 1 => 2 sản phẩm càng giống nhau

# HỆ THỐNG ĐỀ XUẤT

## 4. Đưa ra gợi ý

- Tạo hàm để xử lý các mô tả mới
- Tính độ tương đồng giữa sản phẩm mới và cũ
- Đưa ra gợi ý

```
recommend({'Thương hiệu': 'hp',
           'Giá': '20000000',
           'RAM': '16gb'})
```

Các sản phẩm tương tự:

Laptop Hp 15S Fq5162Tu I5 1235U/8Gb/512Gb/Win11 (7C134Pa) - Giá: 12990000  
Laptop Hp 240 G9 I5 1235U/8Gb/512Gb/Win11 (6L1Y2Pa) - Giá: 12990000  
Laptop Hp Pavilion 15 Eg2081Tu I5 1240P/16Gb/512Gb/Win11 (7C0Q4Pa) - Giá: 16890000  
Laptop Hp Pavilion 15 Eg2081Tu I5 1240P/16Gb/512Gb/Win11 (7C0Q4Pa) - Giá: 16890000  
Laptop Hp Pavilion 15 Eg2081Tu I5 1240P/16Gb/512Gb/Win11 (7C0Q4Pa) - Giá: 16890000  
Laptop Hp Pavilion 15 Eg3091Tu I7 1355U/16Gb/512Gb/Win11 (8C5L2Pa) - Giá: 20990000  
Laptop Hp Pavilion 15 Eg2081Tu I5 1240P/16Gb/512Gb/Win11 (7C0Q4Pa) - Giá: 16890000  
Laptop Hp Gaming Victus 15 Fa1139Tx I5 12450H/16Gb/512Gb/4Gb Rtx2050/144Hz/Win11 (8Y6W3Pa) - Giá: 17690000  
Laptop Hp Pavilion 15 Eg2081Tu I5 1240P/16Gb/512Gb/Win11 (7C0Q4Pa) - Giá: 16890000

5

# KẾT LUẬN & HƯỚNG PHÁT TRIỂN

## ĐẠT ĐƯỢC

- Tìm hiểu được các kỹ thuật học máy, ứng dụng vào xây dựng mô hình dự đoán sự hài lòng
- Đánh giá được hiệu quả của mô hình dự đoán thông qua các chỉ số.
- Thực hiện được việc phân cụm sản phẩm thành các cụm, mỗi cụm sẽ phù hợp với một nhóm khách hàng.
- Xây dựng được mô hình gợi ý sản phẩm có đặc điểm tương đồng với mong muốn của khách hàng.

## HẠN CHẾ

- Mô hình dự đoán vẫn chưa đạt được độ chính xác 100%, vẫn còn một số trường hợp dự đoán không chính xác.
- Chưa thể phân tích sâu các yếu tố ảnh hưởng đến sự hài lòng của khách hàng.
- Các điểm trong cụm vẫn chưa được chặt chẽ, các cụm chưa tách biệt rõ ràng với nhau.

## HƯỚNG PHÁT TRIỂN

- Tiếp tục nghiên cứu và cải tiến mô hình học máy để nâng cao độ chính xác và hiệu quả dự đoán.
- Ứng dụng nhiều kỹ thuật học máy khác nhau để khai thác tốt hơn về dữ liệu.
- Mở rộng phạm vi nghiên cứu, thu thập thêm dữ liệu, phân tích sâu hơn các yếu tố ảnh hưởng, phân biệt rõ ràng từng cụm với nhau.
- Phối hợp xây dựng một trang web để xuất cho khách hàng với giao diện dễ thao tác.



ĐẠI HỌC ĐÀ NẴNG  
TRƯỜNG ĐẠI HỌC KINH TẾ

Khoa: Thống kê - Tin học  
KHÓA 47

# CHÂN THÀNH CẢM ƠN QUÝ HỘI ĐỒNG !

Sinh viên trình bày:

**NGUYỄN THỊ DUYÊN  
NGUYỄN THỊ MÙI**

