



Welcome!

Welcome to our presentation



Đại học Quốc gia Thành phố Hồ Chí Minh
Trường Đại học Công Nghệ Thông Tin



Tên đề tài:

Xây dựng bộ dữ liệu dự đoán giá laptop cũ thu thập từ trang thương mại điện tử FPT Shop

Môn học:

Thu thập và tiền xử lý dữ liệu - DS103.M21

Giảng viên hướng dẫn:

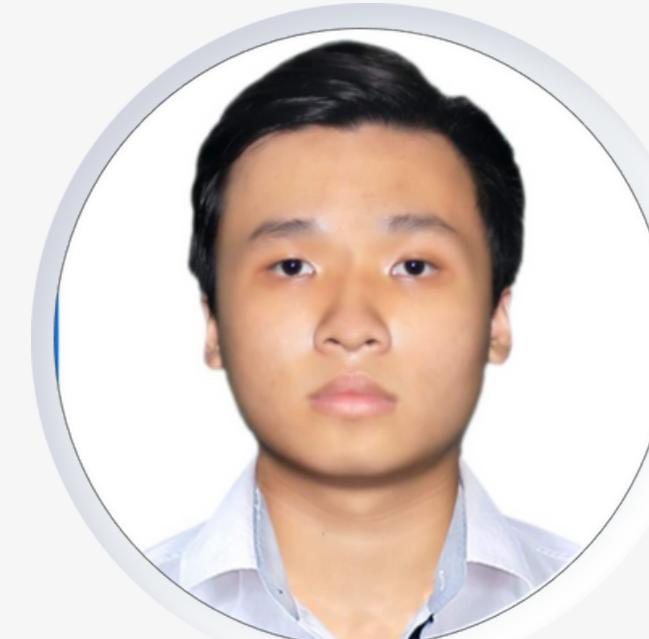
TS. Nguyễn Gia Tuấn Anh

ThS. Lưu Thanh Sơn

Team Member



Nguyễn Hùng Trung Hiếu
20521323



Trịnh Thế Hiện
20521310



Lưu Quang Tiến Hoàng
20521342



Nguyễn Thành Nhàn
20521701

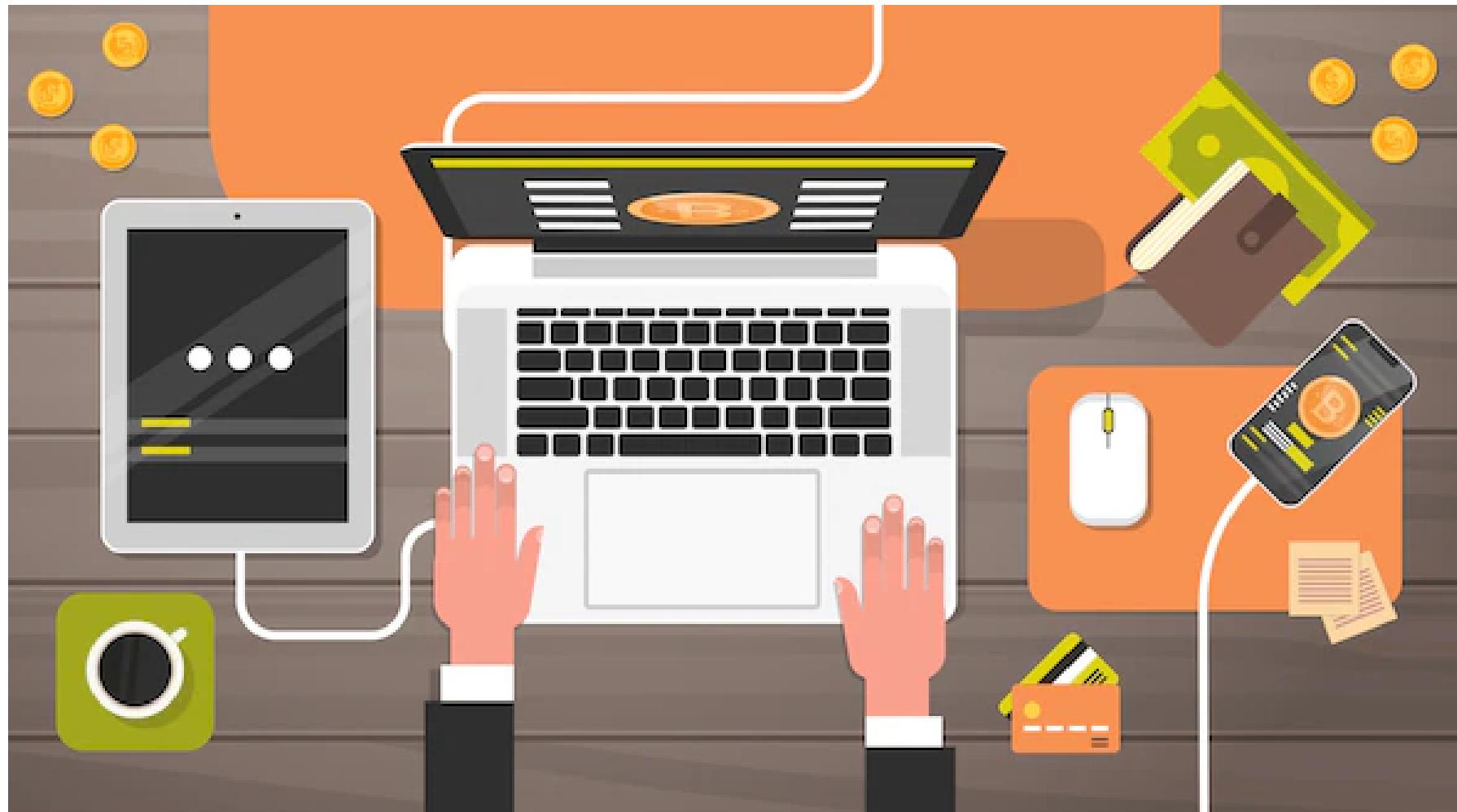
CONTENT



- 1 Giới thiệu
- 2 Tổng quan về bộ dữ liệu
- 3 Thu thập và xử lý dữ liệu
- 5 Phương pháp máy học
- 6 Kết luận và hướng phát triển
- 7 Tài liệu tham khảo

GIỚI THIỆU

- *Bối cảnh*



- Thời đại công nghệ ngày càng phát triển mạnh mẽ
- Đời sống con người đang nhận được sự ủng hộ và thúc đẩy rất lớn
- Laptop là một thứ không thể thiếu để phục vụ cho học tập và công việc

GIỚI THIỆU

- *Bối cảnh*

Không phải bất kì ai cũng có đủ:

- Kinh tế để có thể trang bị những sản phẩm mới
- Có đủ chuyên môn để đánh giá một chiếc laptop.

Chọn đề tài "*Xây dựng bộ dữ liệu dự đoán giá laptop cũ thu thập từ trang thương mại điện tử FPT Shop*"



GIỚI THIỆU

- *Định nghĩa bài toán*

Task: Dự đoán giá laptop cũ.

E: Gồm 2 tập train và test.

P: Tỉ lệ dự đoán chính xác giá của laptop cũ.

Input: Thông số của một chiếc laptop cũ.

Output: Nhãn giá được dự đoán cho chiếc laptop đó.

GIỚI THIỆU

- Mục tiêu

- Nghiên cứu xây dựng được bộ dữ liệu sạch với độ tin cậy cao
- Dự đoán được giá laptop cũ với độ chính xác khoảng ... %
- Nghiên cứu và thực hiện được quy trình các bước để xây dựng một bộ dữ liệu.



GIỚI THIỆU

- *Ứng dụng*

- a
- b
- c

CONTENT



- 1 Giới thiệu
- 2 Tổng quan về bộ dữ liệu
- 3 Thu thập và xử lý dữ liệu
- 5 Phương pháp máy học
- 6 Kết luận và hướng phát triển
- 7 Tài liệu tham khảo

CONTENT



- 1 Giới thiệu
- 2 Tổng quan về bộ dữ liệu
- 3 Thu thập và xử lý dữ liệu
- 5 Phương pháp máy học
- 6 Kết luận và hướng phát triển
- 7 Tài liệu tham khảo

Tổng quan về bộ dữ liệu

- Nguồn thu thập

- Dữ liệu được thu thập từ trang thương mại điện tử FPT Shop

Nguồn: <https://fptshop.com.vn/>



Tổng quan về bộ dữ liệu

- Thông tin về bộ dữ liệu

- Có 1573 dòng dữ liệu
- File Raw Data có hơn 70 thuộc tính
- Sau khi xử lý chỉ giữ lại 38 thuộc tính phù hợp với mô hình máy học

	Tên máy	Giá hiện tại	Giá gốc	Hạn bảo hành	Link	Thiết kế & Trọng lượng	Thiết kế & Trọng lượng	Bộ xử lý.Công nghệ	Bộ xử lý.Loại CPU	Bộ xử lý.Tốc độ CPL	Bộ xử lý.
0	MacBook Pro 16" 20	50.402.000	69.999.000	11/10/2022	https://fptshop.com.v	2	["Nhựa", "Nhôm", "Nhôm"]	Core i9	9th-gen		2.3
1	MacBook Pro 14" 20	50.149.000	65.990.000	08/02/2023	https://fptshop.com.v	1.604	["Nhựa", "Nhôm", "Nhôm"]	M1 Pro			
2	Laptop Dell XPS 13	43.802.000	59.990.000	10/12/2022	https://fptshop.com.v	1.32	["Nhôm"]	Core i7	1165G7		2.8
3	MacBook Pro 16" 20	43.197.000	59.999.000	11/10/2022	https://fptshop.com.v	2	["Nhôm"]	Core i7	9th-gen		2.6
4	MacBook Pro 16" 20	43.197.000	59.999.000	01/12/2022	https://fptshop.com.v	2	["Nhôm"]	Core i7	9th-gen		2.6
5	MacBook Pro 16" 20	43.197.000	59.999.000	01/12/2022	https://fptshop.com.v	2	["Nhôm"]	Core i7	9th-gen		2.6
6	MacBook Pro 16" 20	43.197.000	59.999.000	01/12/2022	https://fptshop.com.v	2	["Nhôm"]	Core i7	9th-gen		2.6
7	MacBook Pro 16" 20	43.197.000	59.999.000	01/12/2022	https://fptshop.com.v	2	["Nhôm"]	Core i7	9th-gen		2.6
8	MacBook Pro 16" 20	43.197.000	59.999.000	01/12/2022	https://fptshop.com.v	2	["Nhôm"]	Core i7	9th-gen		2.6
9	MacBook Pro 16" 20	43.197.000	59.999.000	01/12/2022	https://fptshop.com.v	2	["Nhôm"]	Core i7	9th-gen		2.6
10	MacBook Pro 16" 20	43.197.000	59.999.000	01/12/2022	https://fptshop.com.v	2	["Nhôm"]	Core i7	9th-gen		2.6
11	MacBook Pro 16" 20	43.197.000	59.999.000	01/12/2022	https://fptshop.com.v	2	["Nhôm"]	Core i7	9th-gen		2.6
12	MacBook Pro 16" 20	43.197.000	59.999.000	01/12/2022	https://fptshop.com.v	2	["Nhôm"]	Core i7	9th-gen		2.6

Tổng quan về bộ dữ liệu

- Thông tin về bộ dữ liệu

Các thuộc tính như sau:

STT	Tên thuộc tính	STT	Tên thuộc tính	STT	Tên thuộc tính	STT	Tên thuộc tính
1	Hãng máy	11	Bộ xử lý.Số luồng	21	Màn hình.Màn hình cảm ứng	31	Lưu trữ.Số khe SSD/HDD còn lại
2	Giá hiện tại	12	RAM.Dung lượng RAM	22	Đồ họa.Card onboard.Hãng	32	Lưu trữ.SSD.Dung lượng
3	Giá gốc	13	RAM.Loại RAM	23	Đồ họa.Card onboard.Model	33	Giao tiếp & kết nối.Cổng giao tiếp
4	Hạn bảo hành	14	RAM.Số khe RAM còn lại	24	Đồ họa.Card onboard.Xung nhịp	34	Bàn phím & TouchPad.Đèn bàn phím
5	Thiết kế & Trọng lượng.Trọng lượng sản phẩm	15	RAM.Hỗ trợ RAM tối đa	25	Đồ họa.Card onboard.Xung nhịp tối đa	35	Màn hình.Tần số quét
6	Thiết kế & Trọng lượng.Chất liệu	16	Màn hình.Kích thước màn hình	26	Đồ họa.Card onboard.Bộ nhớ	36	Thông tin pin & Sạc.Dung lượng pin
7	Bộ xử lý.Công nghệ CPU	17	Màn hình.Độ phân giải	27	Đồ họa.Card rời.Hãng	37	Hệ điều hành
8	Bộ xử lý.Loại CPU	18	Màn hình.Loại màn hình	28	Đồ họa.Card rời.Model		
9	Bộ xử lý.Tốc độ CPU	19	Màn hình.Tấm nền	29	Đồ họa.Card rời.Bộ nhớ		
10	Bộ xử lý.Số nhân	20	Màn hình.Độ phủ màu	30	Lưu trữ.Kiểu ổ cứng		

CONTENT



- 1 Giới thiệu
- 2 Tổng quan về bộ dữ liệu
- 3 Thu thập và xử lý dữ liệu
- 4 Phương pháp máy học
- 5 Kết luận và hướng phát triển
- 6 Tài liệu tham khảo

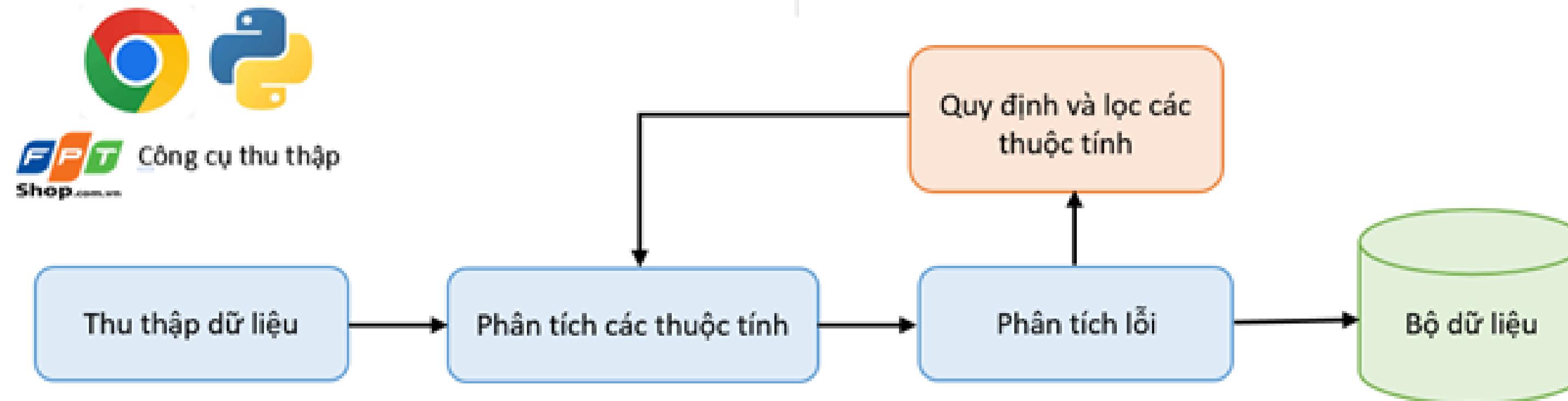
Thu thập và xử lý dữ liệu

- *Cách thức thu thập dữ liệu*

- Thu thập trên trang thương mại điện tử FPT Shop.
- Sử dụng tiện ích Google Chrome, Google Colab, Python Visual Code
- Thu được khoảng từ 10 – 20 dữ liệu trong 1 lần tải

Thu thập và xử lý dữ liệu

- Quy trình thu thập dữ liệu



Thu thập và xử lý dữ liệu

- *Quy trình xử lý dữ liệu*

Thu thập và xử lý dữ liệu

- Đánh giá tập dữ liệu

- Có nhiều thuộc tính, Có nhiều kiểu giá trị khác nhau, cần được tiền xử lý.
- Tập dữ liệu có kích thước nhỏ, nhưng đủ để thực hiện nghiên cứu.
- Tập train và tập test được chia theo đúng tỉ lệ 8:2

Thu thập và xử lý dữ liệu

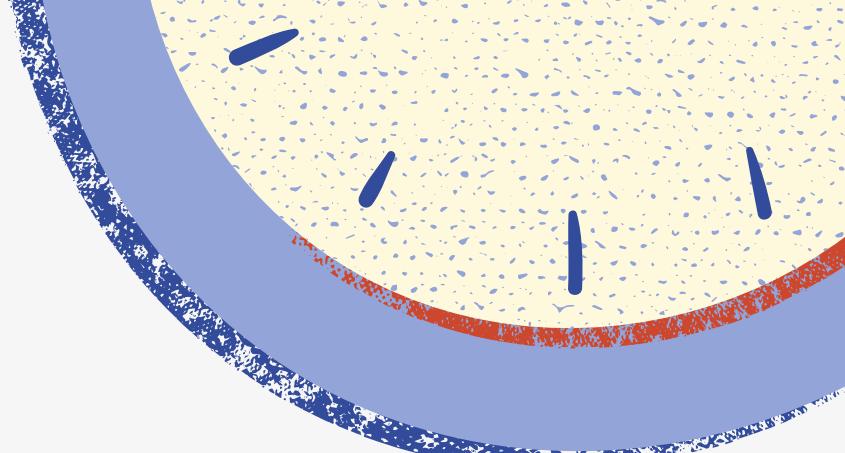
- *Thách thức trong quá trình xây dựng bộ dữ liệu*

- Trang web thu thập dữ liệu là trang web động
- Không đảm bảo sự đồng đều về chất lượng và nhất quán giữa các giá trị
- Có một số lượng dữ liệu bị trùng sau khi tải
- Bị lẫn dữ liệu rác, link quảng cáo trong quá trình tải

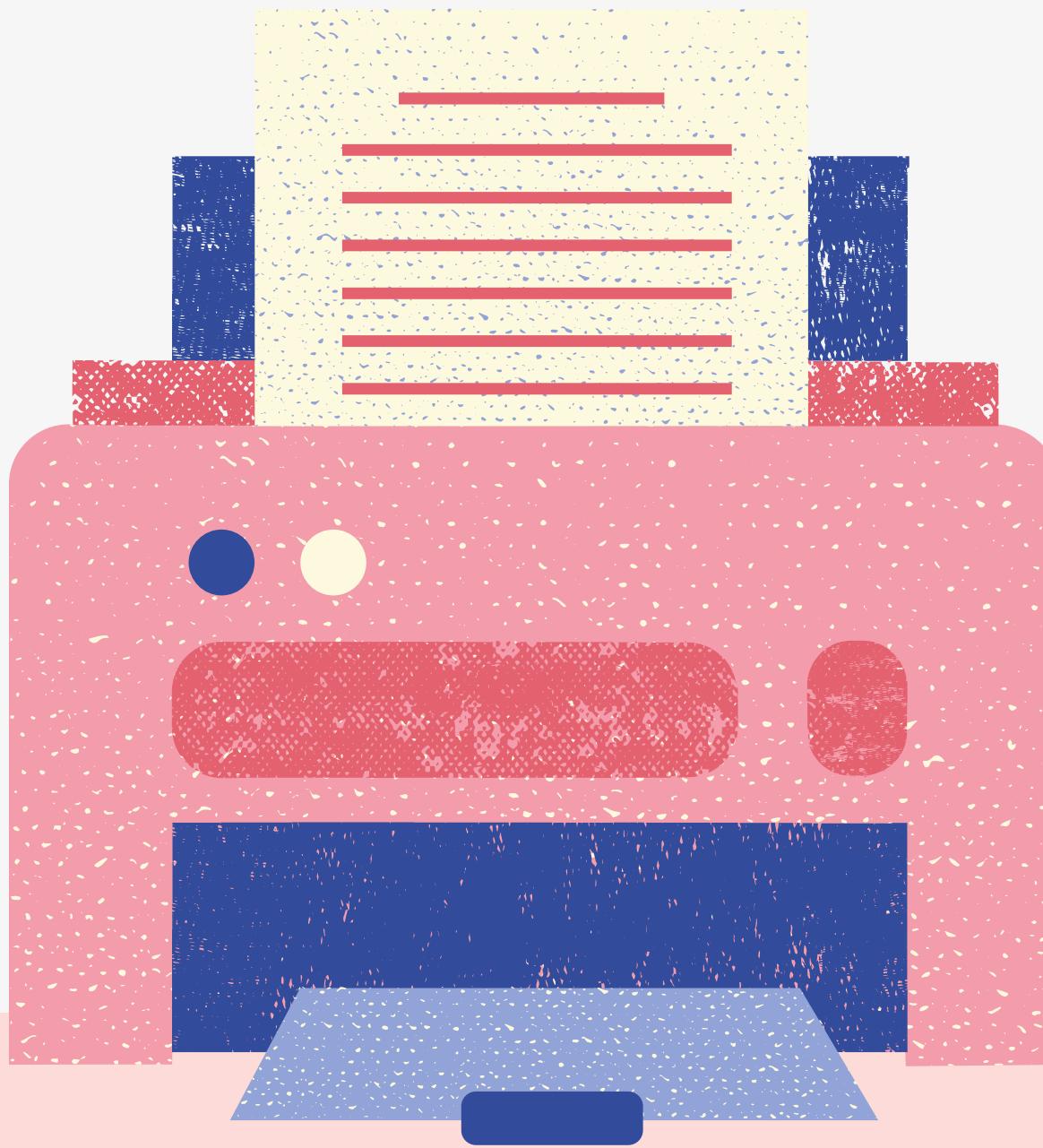
CONTENT



- 1 Giới thiệu
- 2 Tổng quan về bộ dữ liệu
- 3 Thu thập và xử lý dữ liệu
- 4 Phương pháp máy học
- 5 Kết luận và hướng phát triển
- 6 Tài liệu tham khảo



*Thank you
For Listening*



Do you have any questions?

