

# FINAL PROJECT

## Handling Real-world Problems

---

**Môn học:** Nhập môn Khoa học dữ liệu

**Nhóm:** 9



# Nội dung

O1

Giới thiệu

O3

Data Preprocessing

O5

Data Modeling

O2

Data Collection

O4

Data Exploration



# O1

## GIỚI THIỆU NHÓM

22120384

NGUYỄN ĐÌNH TRÍ

22120398

VŨ HOÀNG NHẬT TRƯỜNG

22120412

NGUYỄN ANH TƯỜNG

22120424

PHẠM NGỌC BẢO UYÊN

22120449

LÊ NGUYỄN HUYỀN VY



# O2

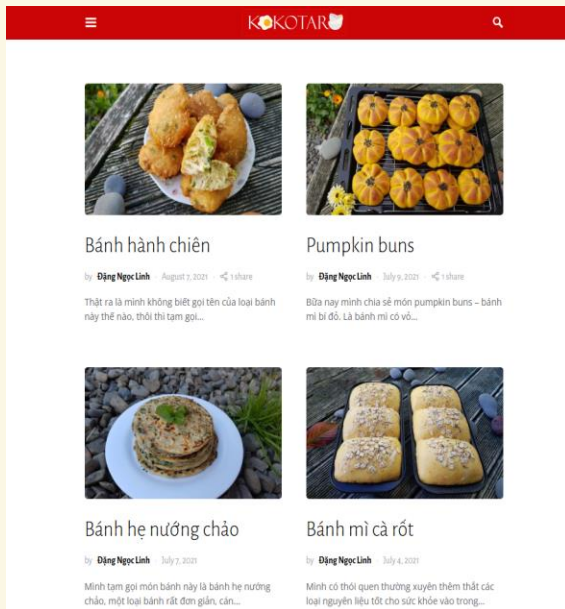
---

## DATA COLLECTION



# Cào dữ liệu

Kokotaru.com



Kitchenart.com



Selenium



HTML

# Kết quả

```
1 Title: Nước Soda Chanh Sả
2 Ingredients:
3 1/4 cup vỏ chanh xanh (từ 8 quả chanh)
4 20 lá chanh
5 2 củ sả, thái rối
6 3/4 cup (180ml) nước đường (tỉ lệ đường- nước 1:1)
7 6 cups (1,5l) nước soda, chia đều ra 12 cốc
8 -----
9 Title: Thịt Gà Hầm Rau Củ
10 Ingredients:
11 1 thìa canh dầu oliu
12 1 củ hành tây, thái nhỏ hạt lựu
13 500g thịt đùi gà, lọc xương, da
14 300g khoai tây bi
15 425ml nước dùng rau củ
16 350g súp lơ xanh
17 Một nắm hành lá, húng tây thái nhỏ
18 140g đậu Hà Lan
19 -----
20 Title: Cá Nướng Giòn Với Sốt Tartare
21 Ingredients:
22 2/3 cup bột chiên xù (chúng tôi dùng hiệu panko)
23 1/4 cup phomai parmesan bào nhỏ
24 2 thìa canh (tbs) lá mùi tây dẹt ( flat-leaf parsley), cắt nhỏ
25 1/3 cup bột mì đa dụng
26 1 quả trứng gà, đánh tan
27 4 phile cá vược (hoặc cá quả), rút xương
28 Dầu oliu
29 60g salad rau, cà chua bi trộn dầu dấm
30 Sốt tartare
31 2/3 cup sữa chua không đường
32 1 thìa cà phê (tsp) mù tạt Dijon
33 1 thìa canh (tbs) nụ bạch hoa (capers), cắt nhỏ
34 2-3 quả dưa chuột bao tử ngâm dấm, cắt nhỏ
35 -----
```

**Kokotaru: 399 / 468 bài viết**

**Kitchenart: 682 / 720 bài viết**

**Tổng cộng: 1081 bài viết**





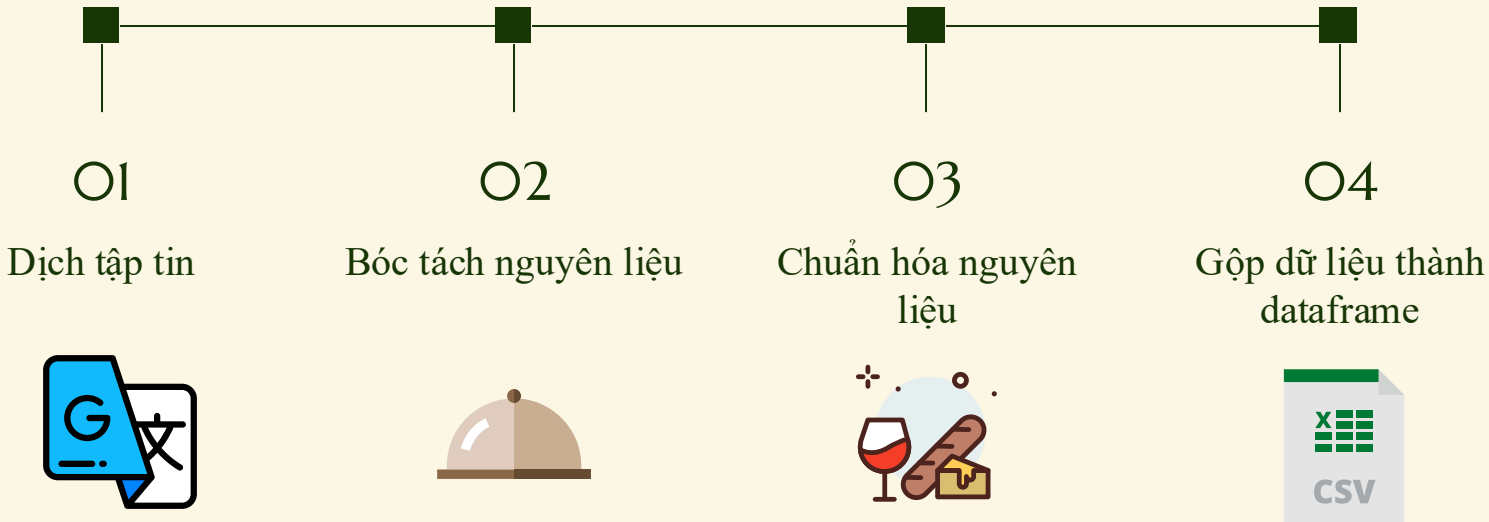
# O3

---

## DATA PREPROCESSING



# Các bước thực hiện

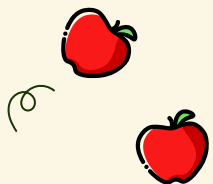






# Tập dữ liệu

Name of dish	soda	turmeric	herb	yogurt	butter	beef bone	chili	lemor	beef fillet	chipotle	sr	cocoa	pow	tea	po	green	cashew	cherry	tom
10 common	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11 ways to	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12 types of	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14 ways to	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14 ways to	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4 ways to n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6 delicious	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9 delicious	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Amaretti à	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Answers to	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Apple Muff	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Apple and	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Apple crisp	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Apple raisi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Apple scor	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Apple, Ras	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Apricot ch	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bacon Toa	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Baked Zuc	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Banana Ch	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Banana co	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0





# O4

---

## DATA EXPLORATION



# Tìm hiểu về tập dữ liệu

01

Dữ liệu gồm mấy dòng, mấy cột?



Ý nghĩa mỗi dòng? Dòng trùng?

02

03

Ý nghĩa mỗi cột? Kiểu dữ liệu mỗi cột?



# Tìm hiểu về tập dữ liệu

04

Cột số: Min, Max? Giá trị thiếu?...



Cột phân loại: Giá trị thiếu? Phân biệt?


05

06


Kiểm tra tập dữ liệu phù hợp chưa?



# Phân tích dữ liệu

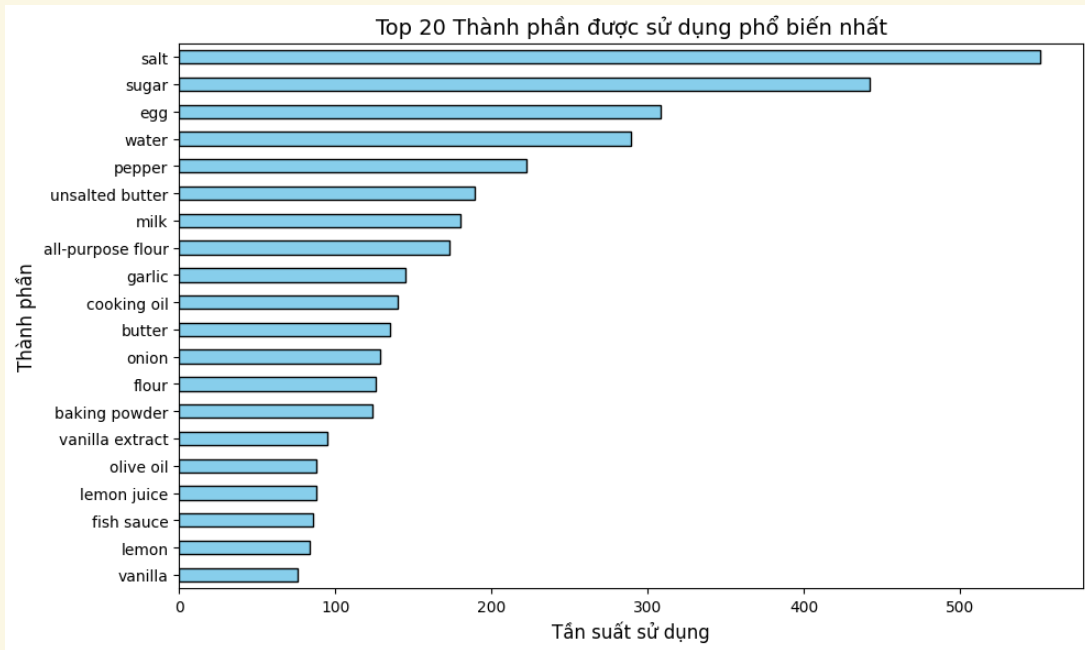


Phân tích sự phổ biến và khả năng kết hợp của các nguyên liệu thành phần?



Phân tích độ phức tạp của món ăn dựa trên số lượng nguyên liệu cũng như độ phổ biến của chúng?

# PHÂN TÍCH SỰ PHỔ BIẾN VÀ KHẢ NĂNG KẾT HỢP?

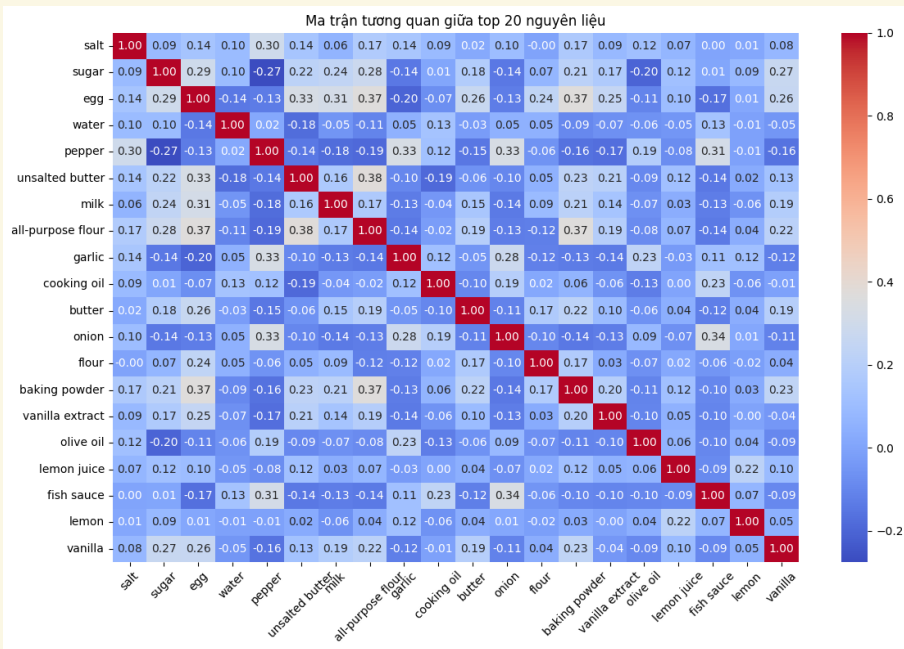


## 1 SỰ PHỔ BIẾN

Muối, đường: gia vị không thể thiếu trong các món ăn

Ưu tiên dự trữ các nguyên liệu có tần suất sử dụng cao

# PHÂN TÍCH SỰ PHỔ BIẾN VÀ KHẢ NĂNG KẾT HỢP?



2

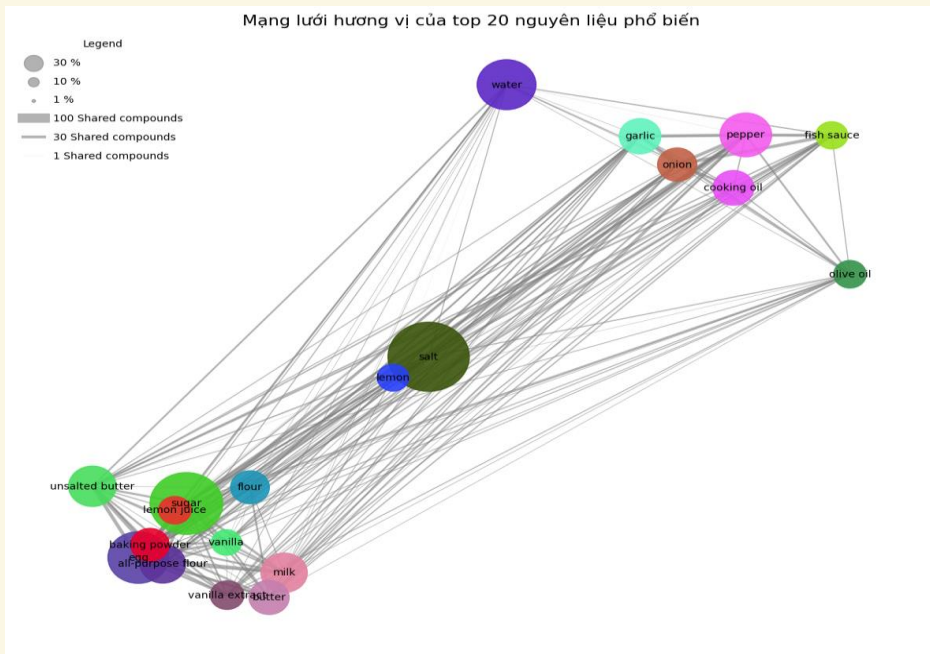
SỰ KẾT HỢP

Dựa trên mối tương quan mà chia các nhóm món ăn:

➤ **Nhóm làm bánh:** bột mỳ, bột nở,...

➤ **Nhóm món mặn:** nước mắm, tỏi, tiêu, hành,...

# PHÂN TÍCH SỰ PHỔ BIẾN VÀ KHẢ NĂNG KẾT HỢP?



2

## SỰ KẾT HỢP

Nguyên liệu được phân cụm thành: nguyên liệu mặn, nguyên liệu làm bánh

Đường liên kết giữa 2 nguyên liệu thể hiện mức độ chia sẻ cao



# PHÂN TÍCH SỰ PHỔ BIẾN VÀ KHẢ NĂNG KẾT HỢP?

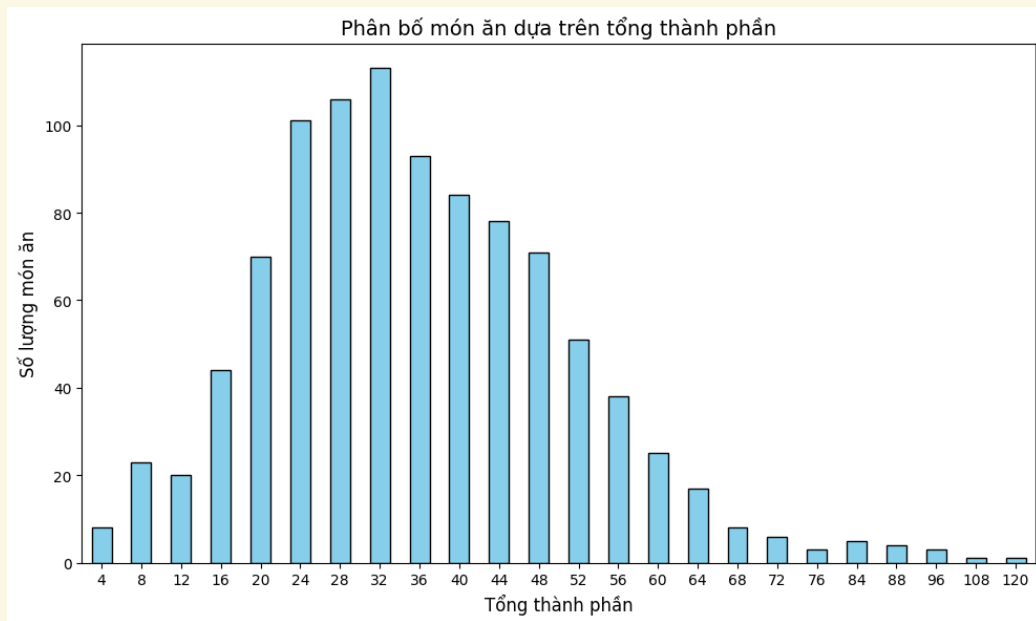
3

**SỰ ĐẶC TRƯNG**

Một vài món ăn chỉ có 1 nguyên liệu chỉ xuất hiện một lần trong món đó

	Name of dish	Distinctive Ingredients
0	Answers to questions about moon cake ingredients	d ingredients
1	Blackberry crumble cheesecake bars	weetbix
2	Chocolate, raspberry and ginger chocolate mous...	gelatin leaf
3	Coconut Cookies	please give test
4	Cookies with jam filling – Rosenmunnar	fruit jam according

# PHÂN TÍCH ĐỘ PHỨC TẠP CỦA MÓN ĂN

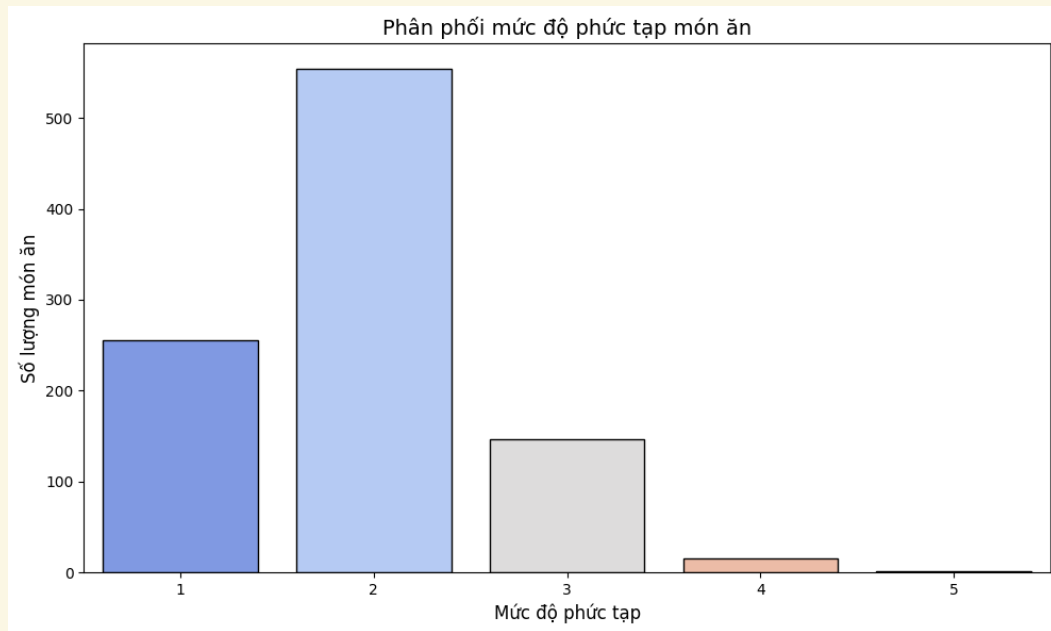


1 SỐ LƯỢNG NL

Phân phối lệch phải

Đa số các món ăn tập trung ở số lượng nguyên liệu vừa (24-40 nguyên liệu)

# PHÂN TÍCH ĐỘ PHỨC TẠP CỦA MÓN ĂN

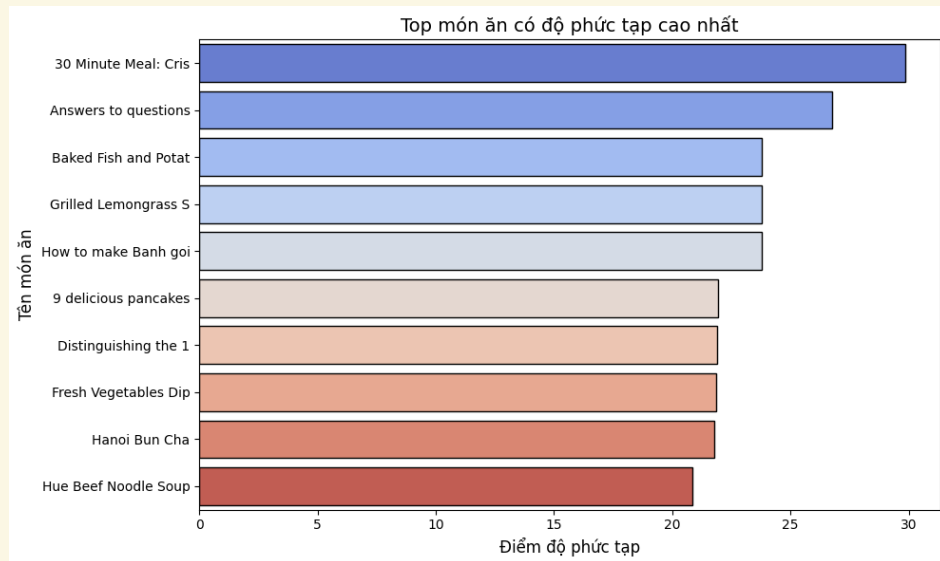


**2** ĐỘ PHỨC TẠP

Phân phối lệch phải

Đa số các món ăn có độ phức tạp dễ và trung bình (1-2 sao)

# Một vài món ăn có độ phức tạp cao

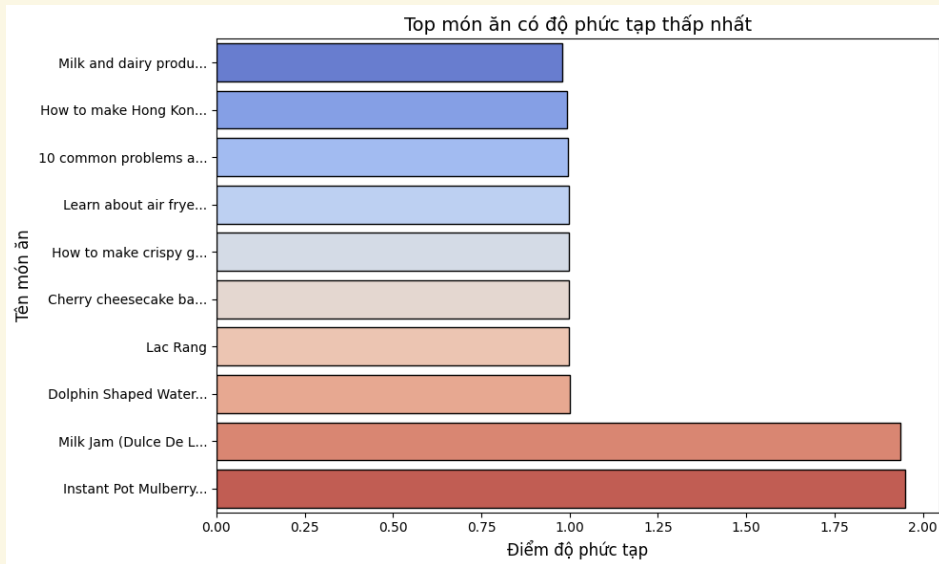


Số lượng nguyên liệu trong mỗi món ăn trên:

1. Món 30 Minute Meal: Crispy Fried Chicken, Salad, Green Curry, Noodle Soup: 60
2. Món Answers to questions about moon cake ingredients: 54
3. Món Baked Fish and Potatoes with Tartar Sauce: 48
4. Món Grilled Lemongrass Spring Rolls: 48
5. Món How to make Banh goi: 48
6. Món 9 delicious pancakes that don't require an oven: 44
7. Món Distinguishing the 10 most popular types of cheese: 44
8. Món Fresh Vegetables Dipped in Simple Sweet and Sour Sauce: 44
9. Món Hanoi Bun Cha: 44
10. Món Hue Beef Noodle Soup Cooked in Instant Pot Pressure Cooker: 42

» *Các món ăn có độ phức tạp cao thường là các món ăn có chứa nhiều nguyên liệu (và có thể chứa các nguyên liệu ít phổ biến)*

# Một vài món ăn có độ phức tạp thấp

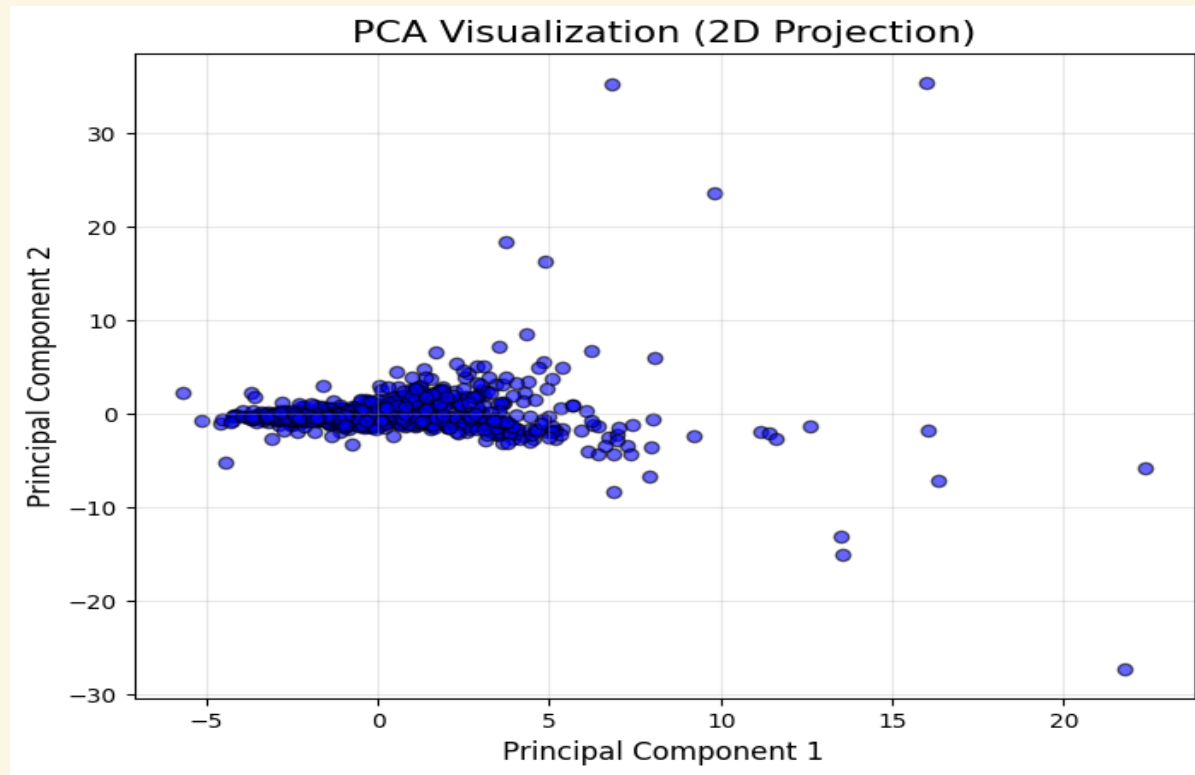


Số lượng nguyên liệu trong mỗi món ăn trên:

1. Món Milk and dairy products: 2
2. Món How to make Hong Kong Pandan Sponge Cake: 2
3. Món 10 common problems and mistakes when making bread: 2
4. Món Learn about air fryer: 2
5. Món How to make crispy grilled smoked meatloaf - Bacon puffs: 2
6. Món Cherry cheesecake bars: 2
7. Món Lac Rang: 2
8. Món Dolphin Shaped Watermelon: 2
9. Món Milk Jam (Dulce De Leche): 4
10. Món Instant Pot Mulberry Syrup: 4

» Các món đơn giản chỉ cần 1-2 nguyên liệu là có thể hoàn thành

# Xác định thành phần chính của tập dữ liệu






O5

---


# DATA MODELING




# Mô hình



Hệ thống gợi ý nguyên liệu thay thế cho các nguyên liệu bị thiếu, sao cho phù hợp với các nguyên liệu hiện có



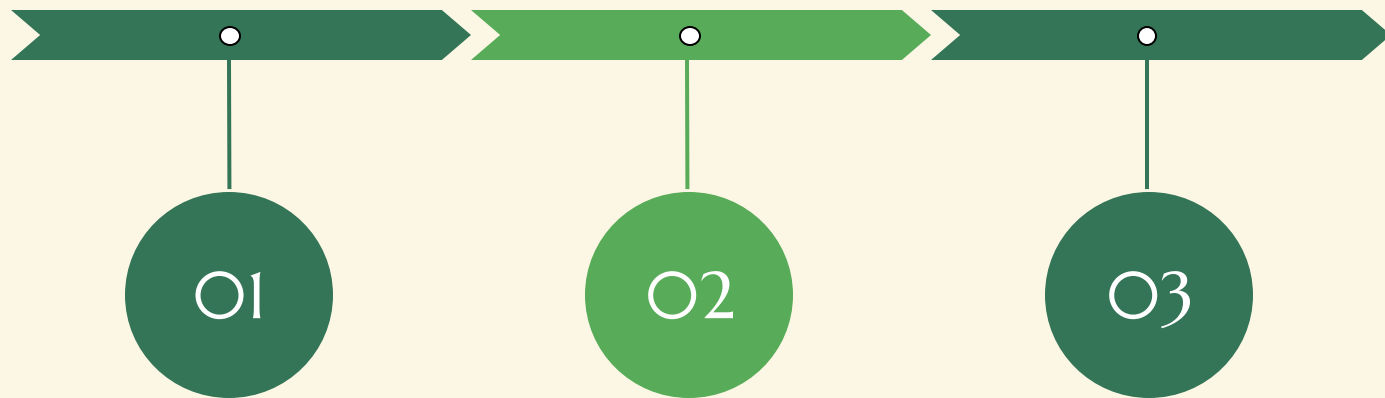
Hệ thống đề xuất các món ăn thay thế, giúp dễ dàng khám phá các món ăn có đặc điểm chung.



Hệ thống tìm kiếm và đề xuất các món ăn phù hợp dựa trên danh sách nguyên liệu có sẵn.



# HỆ THỐNG GỢI Ý NGUYÊN LIỆU THAY THẾ



Xây dựng tập ý nghĩa

Xây dựng mô hình

Đánh giá mô hình

# HỆ THỐNG GỢI Ý NGUYÊN LIỆU THAY THẾ



## Context – Công dụng

Topping, làm ngọt, gia vị, trang trí,...



## Flavor – Vị

Ngọt, đắng, chua, cay,...



## Smell - Mùi

Mùi thảo mộc, cay, khói, tanh,...

## Type – Loại

Trái cây, rau củ, thịt, hạt,...



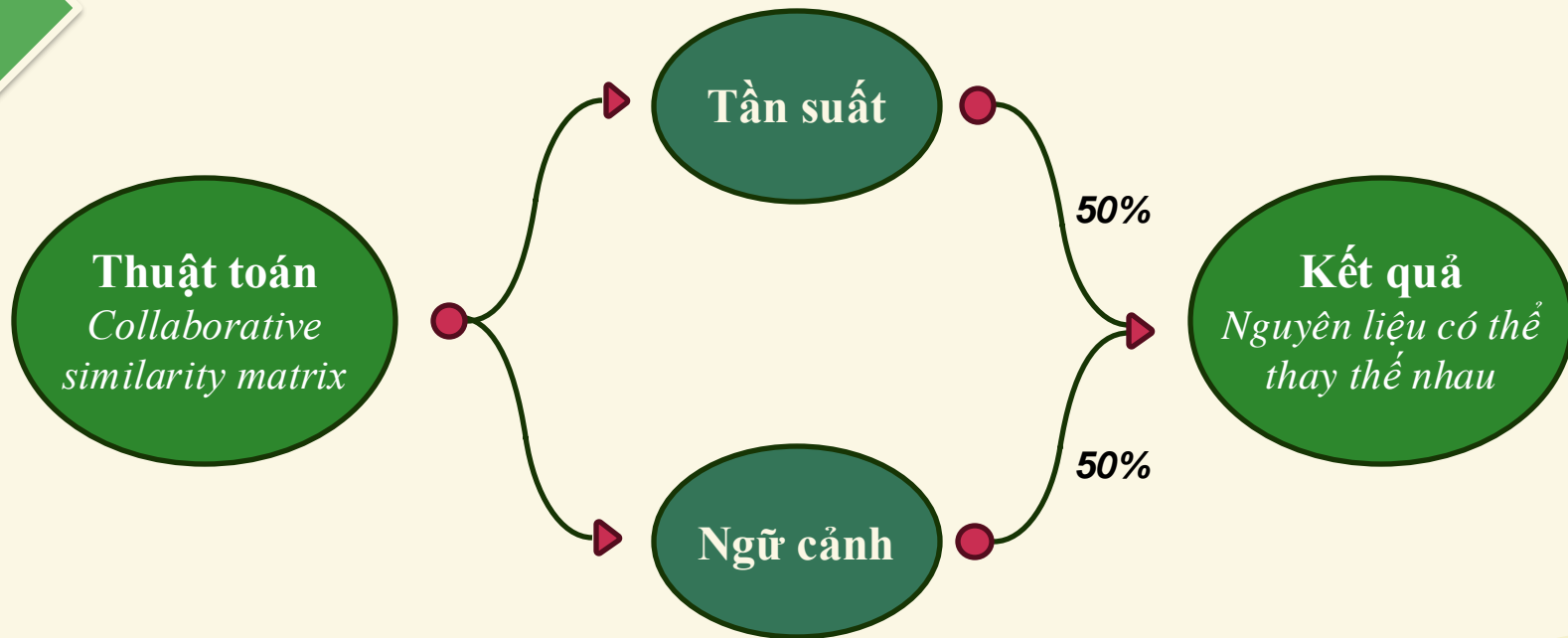
LM Studio - Model: QwQ 2.5 14B Instruct.



# HỆ THỐNG GỢI Ý NGUYÊN LIỆU THAY THẾ



O2



# HỆ THỐNG GỢI Ý NGUYÊN LIỆU THAY THẾ



O2

```
1 current_ingredients = ['active yeast', 'agave nectar', 'all-purpose flour']
2 missing_ingredients = ['brown sugar', 'bacon']
3
4 recommendations = recommend_replacements(current_ingredients, missing_ingredients)
5 recommendations

{'brown sugar': ['sugar', 'maple syrup', 'icing sugar'],
 'bacon': ['pacetta bacon', 'ham', 'pork loin']}
```

» Có thể thay thế **đường nâu** bằng đường, bằng **maple syrup**, hoặc **đường bột** (đường xay – icing sugar)

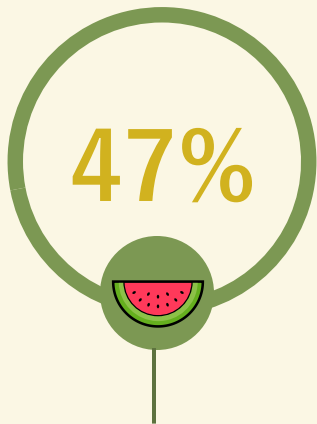
» Có thể thay thế **thịt xông khói** (bacon) bằng **thịt xông khói pacetta**, **thịt nguội** hoặc **thịt lợn loin**



# HỆ THỐNG GỢI Ý NGUYÊN LIỆU THAY THẾ



03



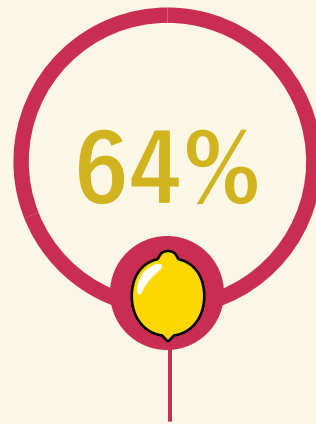
**Precision**

Nguyên liệu được gợi ý khá chính xác



**Recall**

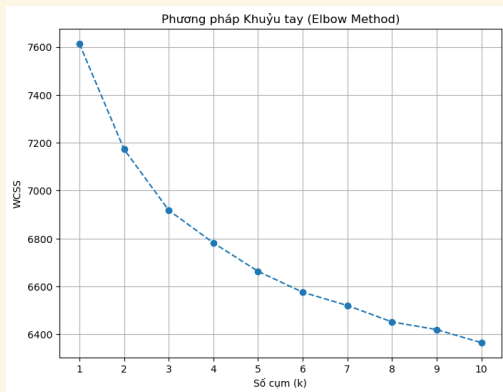
Gợi ý được tất cả nguyên liệu kỳ vọng



**F1**

Cân bằng tốt giữa Precision và Recall

# HỆ THỐNG ĐỀ XUẤT CÁC MÓN ĂN THAY THẾ



CỤM MÓN  
CHÍNH

Mỳ pasta tôm, gà nướng  
thảo mộc, gà hầm rau củ...

ELBOW  
METHOD

CỤM MÓN  
BÁNH

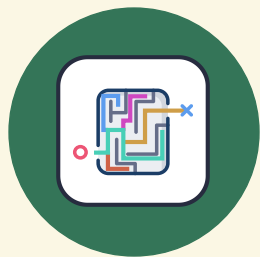
Pancake sữa hạt, cookies  
hạnh nhân, ...

CỤM MÓN  
ĂN VẶT

Trà vải, trà tắc, phô mai, ...

Sử dụng phương pháp **Khuỷu tay** để xác định số cụm

# HỆ THỐNG ĐỀ XUẤT CÁC MÓN ĂN THAY THẾ



## THUẬT TOÁN

*K-Means Clustering*



## GIAO DIỆN

Gõ tên      Món ăn  
món ăn            thay thế

# HỆ THỐNG ĐỀ XUẤT CÁC MÓN ĂN THAY THẾ

Hệ thống đề xuất món ăn thay thế

Input recipes:

Apple Muffins

ENTER

1. Let's Make Easter Bread
2. How to make strawberry crumble pie
3. How to make savory pepper and cheddar cheese muffins
4. How to Make Irresistibly Delicious and Chewy Singapore Cream Puffs
5. Crispy black sesame almond cookies



# HỆ THỐNG ĐỀ XUẤT CÁC MÓN ĂN THAY THẾ

## Đánh giá mô hình

6917.19

### Intertia

Cải thiện độ chặt chẽ các  
cụm khi  $k=3$

0.045

### Silhouette Score

Các cụm vẫn còn chồng lấn,  
các điểm dữ liệu không  
được phân tách rõ ràng

4.03

### Davies-Bouldin Score

Các cụm chưa thực sự phân  
biệt tốt

# HỆ THỐNG TÌM KIẾM & ĐỀ XUẤT MÓN ĂN



## THUẬT TOÁN

*K-Nearest Neighbors*



## GIAO DIỆN

Gõ tên  
nguyên liệu  Món ăn

# HỆ THỐNG TÌM KIẾM & ĐỀ XUẤT MÓN ĂN



Hệ thống gợi ý món ăn

**Input ingredients:**

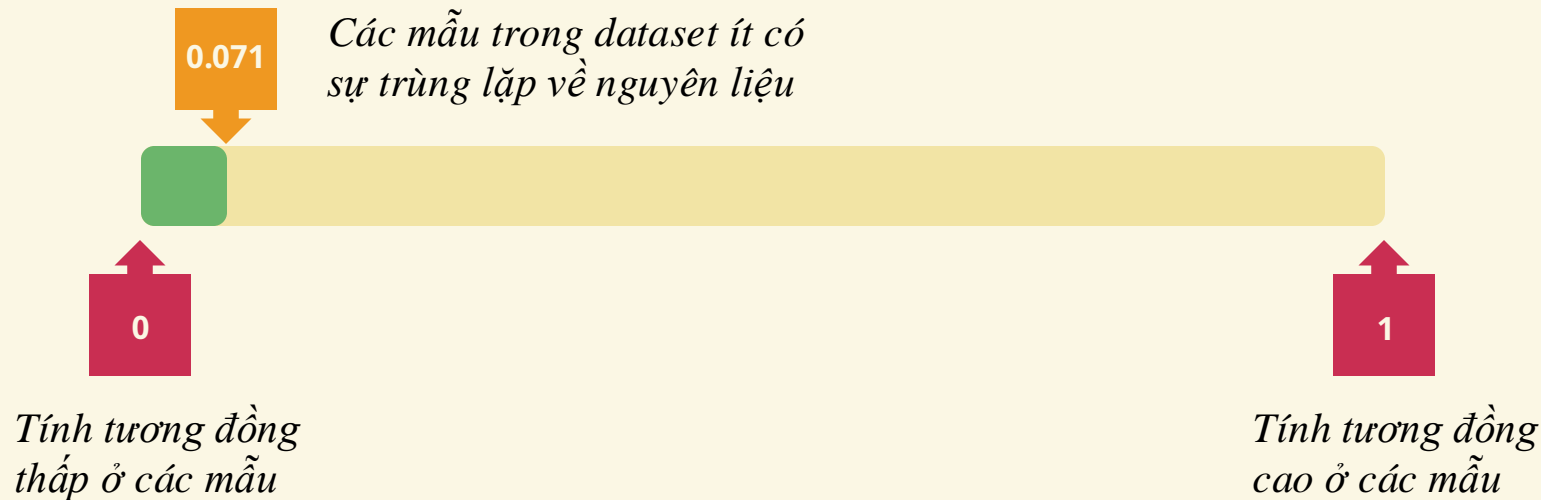
beef

**ENTER**

1. How to make beef stew bread
2. Making Steak Super Easy And Delicious Without Oil, Without Spitting
3. Korean-Style Grilled Beef Skewers with Cosori Air Fryer
4. Grilled Beef Skewers With Vegetables, Smoked, Not Inferior To Charcoal,
5. Instant Pot Beef and Pickled Cucumber Soup

# HỆ THỐNG TÌM KIẾM & ĐỀ XUẤT MÓN ĂN

## Đánh giá mô hình



**Mean Pairwise Jaccard Similarity**

# Thanks!

---

Cảm ơn thầy và các bạn đã lắng nghe!

**Chi tiết đồ án, truy cập:**



[IntroDS-FinalPro-HandlingReal-WorldProblems](#)

