

Trường Đại học Khoa học tự nhiên – Khoa Công nghệ thông tin.

# Đồ án thực hành cuối kỳ.

Introduction to Data Science – Nhập môn khoa học dữ liệu

Nhóm 9  
Tháng 10, 2024.

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

**KHOA CÔNG NGHỆ THÔNG TIN**



**ĐỀ CƯƠNG ĐỒ ÁN THỰC HÀNH**

**Bộ môn:** Nhập môn khoa học dữ liệu.

**Tên đề tài:**

*“Phân tích mối quan hệ giữa các loại gia vị, nguyên liệu và khả năng kết hợp của chúng trong các món ăn”.*

**STT nhóm:** 9.

**Thành viên:**

1. 22120384 – Nguyễn Đình Trí.
2. 22120398 – Vũ Hoàng Nhật Trường.
3. 22120412 – Nguyễn Anh Tường.
4. 22120424 – Nguyễn Ngọc Bảo Uyên.
5. 22120449 – Lê Nguyễn Huyền Vy.

### **Thông tin chung:**

- 1. Bộ môn:** Nhập môn khoa học dữ liệu.
- 2. Giảng viên lý thuyết:** Thầy Lê Ngọc Thành.
- 3. Giảng viên thực hành:** Thầy Lê Nhựt Nam.
- 4. Mã lớp:** 22\_21.
- 5. STT nhóm:** 9.
- 6. Danh sách thành viên:**
  - a. 22120384 – Nguyễn Đình Trí.
  - b. 22120398 – Vũ Hoàng Nhật Trường.
  - c. 22120412 – Nguyễn Anh Tường.
  - d. 22120424 – Nguyễn Ngọc Bảo Uyên.
  - e. 22120449 – Lê Nguyễn Huyền Vy.
- 7. Link github repository:** [“Click here to go to our github repository.”](#)

## **I. Chủ đề thực hiện:**

Phân tích mối liên hệ giữa các loại gia vị, nguyên liệu và cách phối hợp chúng trong các món ăn.

**Ví dụ:** Đây là loại gia vị được sử dụng phổ biến nhất trong các món ăn?

## **II. Nguồn dữ liệu:**

[Kokotaru – Chia sẻ công thức món ăn và những điều thú vị trong cuộc sống.](#)

## **III. Kế hoạch thực hiện:**

**Bước 1:** Xác định vấn đề: cần phải phân tích và hoàn thành các mô hình ở mục **IV. Kết quả kỳ vọng.**

**Bước 2:** “Cào” dữ liệu từ trang web kokotaru về.

*Dữ liệu cào về bao gồm:* Tên món ăn | Nguyên liệu của món ăn đó

**Bước 3:** Làm sạch dữ liệu và chuẩn hóa dữ liệu.

**Bước 4:** Khám phá, đánh giá, phân tích dữ liệu.

**Bước 5:** Xây dựng tập huấn luyện, mô hình và đánh giá mô hình.

**Bước 6:** Trực quan hóa và báo cáo kết quả.

**Bước 7:** Phát triển và bảo trì các kết quả đã đạt được theo kỳ vọng.

## **IV. Kết quả kỳ vọng:**

### **IV.I Phân tích dữ liệu:**

#### **1. Phân tích tần suất sử dụng gia vị, nguyên liệu:**

- Xác định tần suất xuất hiện của mỗi loại gia vị, nguyên liệu trong các món ăn.
- Từ đó, có thể tìm ra các loại gia vị, nguyên liệu phổ biến nhất được sử dụng trong nhiều món ăn, cũng như các gia vị, nguyên liệu ít xuất hiện hơn.

#### **2. Phân nhóm món ăn theo loại gia vị, nguyên liệu chính:**

- Dựa trên các gia vị, nguyên liệu, có thể nhóm các món ăn lại theo loại gia vị, nguyên liệu chính (ví dụ: nhóm các món có ớt, gừng, hay tỏi làm gia vị chính).
- Điều này giúp ta hiểu được sự liên quan giữa các món ăn dựa trên sự tương đồng về gia vị, nguyên liệu.

### **3. Phân tích tương quan giữa các gia vị, nguyên liệu:**

- Phân tích xem các gia vị, nguyên liệu thường xuyên đi kèm với nhau. Ví dụ, nếu tỏi và ớt thường xuất hiện cùng nhau, suy ra giữa có mối quan hệ trong việc tạo hương vị.
- Phân tích xem các loại gia vị, nguyên liệu nào thường xuyên không đi kèm với nhau. Nguyên nhân?

### **4. Đánh giá độ phức tạp của món ăn dựa trên số lượng gia vị, nguyên liệu:**

- Phân loại và đánh giá các món ăn dựa trên độ phức tạp của gia vị, nguyên liệu.
- Có thể có 5 mức độ đánh giá tương ứng với số ngôi sao.

## **IV.II Phát triển mô hình ( KỶ VỌNG ):**

### **1. Mô hình gợi ý gia vị thay thế, nguyên liệu:**

Phát triển mô hình gợi ý gia vị, nguyên liệu thay thế dựa trên các món ăn tương tự hoặc các món sử dụng các gia vị, nguyên liệu gần giống nhau.

Mô hình này có thể sử dụng Word2Vec hoặc cosine similarity để tìm ra các gia vị, nguyên liệu thay thế phù hợp trong cùng một nhóm.

### **2. Mô hình phân cụm món ăn theo gia vị, nguyên liệu:**

Sử dụng các thuật toán như K-Means Clustering để phân cụm món ăn dựa trên các loại gia vị, nguyên liệu. Mô hình sẽ giúp chia món ăn thành các nhóm khác nhau dựa trên sự tương đồng về thành phần gia vị, nguyên liệu.

### **3. Mô hình đề xuất món ăn dựa trên gia vị, nguyên liệu có sẵn:**

Phát triển hệ thống đề xuất món ăn dựa trên các gia vị, nguyên liệu người dùng có sẵn.

Sử dụng các kỹ thuật như collaborative filtering hoặc content-based filtering để đề xuất những món ăn có thể nấu dựa trên các gia vị, nguyên liệu mà người dùng nhập vào.