**.**

**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT THÀNH PHỐ HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**

□□□□

Logo

Description automatically generated

**Môn học: NHẬP MÔN KHOA HỌC DỮ LIỆU**

**PHÂN TÍCH HÀM LƯỢNG DINH DƯỠNG MÓN ĂN UỐNG TỪ THỰC ĐƠN CÁC NHÀ HÀNG ĐỂ ĐƯA RA ĐỀ XUẤT PHÙ HỢP NHU CẦU NGƯỜI DÙNG.**

**BÁO CÁO ĐỒ ÁN CUỐI KỲ**

**Mã lớp học:**  INDS331085\_23\_2\_01

**GVHD: Lê Minh Tân**

**SVTH MSSV**

**Trần Minh Hoàng 21110461**

**Lâm Duy Bảo 21110373**

**Phan Nguyên Định 21110423**

Tp. Hồ Chí Minh, tháng 4 năm 2024

**MỤC LỤC**

[**CHƯƠNG 1: GIỚI THIỆU ĐỀ TÀI** 3](#_Toc150375147)

[**1. Giới thiệu** 3](#_Toc150375148)

[**2. Lợi ích** 3](#_Toc150375149)

[**3. Phương pháp** 4](#_Toc150375150)

[**4. Phạm vi dự án** 5](#_Toc150375151)

[**CHƯƠNG 2: GIỚI THIỆU VỀ BỘ DỮ LIỆU** 6](#_Toc150375158)

[**1.** **Giới thiệu 3 bộ dữ liệu** 6](#_Toc150375159)

[**1.1. Bộ 1** 6](#_Toc150375167)

[**1.2. Bộ 2** 12](#_Toc150375167)

[**1.3. Bộ 3** 17](#_Toc150375167)

[**2. So sánh sâu 3 bộ dữ liệu** 21](#_Toc150375160)

[**3. Chọn ra dữ liệu làm đề tài** 27](#_Toc150375161)

[**4. Kết luận về 3 bộ dữ liệu** 28](#_Toc150375162)

[**CHƯƠNG 3: DEMO** 28](#_Toc150375165)

[**1.** **Báo cáo** 29](#_Toc150375159)

[**2. So sánh** 30](#_Toc150375160)

[**3. Kết luận** 31](#_Toc150375161)

**CHƯƠNG 1: GIỚI THIỆU ĐỀ TÀI**

# **1. Giới thiệu**

Trong thời đại ngày nay, sự quan tâm đến việc duy trì một lối sống lành mạnh và chăm sóc sức khỏe cá nhân đang ngày càng tăng lên. Trong khi đó, việc lựa chọn món ăn phù hợp với nhu cầu dinh dưỡng và thể chất trở thành một phần không thể thiếu trong cuộc sống hàng ngày của mỗi người. Đồng thời, việc dành thời gian để nấu ăn tại nhà không phải lúc nào cũng là lựa chọn phù hợp, đặc biệt đối với những người bận rộn.

Trong bối cảnh này, việc phân tích hàm lượng dinh dưỡng của các món ăn từ thực đơn của các nhà hàng trở nên vô cùng quan trọng và hữu ích. Một dự án phân tích hàm lượng dinh dưỡng món ăn từ thực đơn các nhà hàng có thể giúp đưa ra các đề xuất phù hợp với nhu cầu dinh dưỡng và thị hiếu của người dùng.

Bằng cách thu thập và phân tích dữ liệu từ thực đơn của các nhà hàng, chúng ta có thể tạo ra một cơ sở dữ liệu đa dạng về các loại món ăn và thông tin dinh dưỡng liên quan như lượng calo, protein, carbohydrate, chất béo và các dưỡng chất khác. Dựa trên thông tin này, hệ thống có thể đưa ra các đề xuất món ăn phù hợp với nhu cầu dinh dưỡng cũng như sở thích cá nhân của người dùng.

# **2. Lợi ích**

Dự án "Phân tích hàm lượng dinh dưỡng của món ăn từ thực đơn các nhà hàng để đưa ra đề xuất phù hợp với nhu cầu người dùng" là một phần không thể thiếu trong bối cảnh ngày nay, khi mà sự quan tâm đến dinh dưỡng và lối sống lành mạnh ngày càng được nhấn mạnh. Với số lượng lớn các nhà hàng và món ăn được cung cấp trên các nền tảng trực tuyến, việc lựa chọn món ăn phù hợp với nhu cầu dinh dưỡng và thị hiếu cá nhân trở nên quan trọng hơn bao giờ hết.

Dự án này không chỉ giúp tiết kiệm thời gian và công sức của người tiêu dùng trong quá trình tìm kiếm món ăn, mà còn giúp họ có cái nhìn tổng quan và đúng đắn về hàm lượng dinh dưỡng của các món ăn từ các nhà hàng khác nhau. Đồng thời, thông qua việc đề xuất các món ăn phù hợp với nhu cầu dinh dưỡng và thị hiếu của người dùng, dự án còn góp phần vào việc nâng cao nhận thức về dinh dưỡng và lối sống lành mạnh cho cả cộng đồng.

Đối với các nhà hàng, dự án này không chỉ là một công cụ hữu ích để tối ưu hóa thực đơn và tăng cường trải nghiệm của khách hàng, mà còn là một cơ hội để quảng bá sản phẩm của họ một cách hiệu quả, tạo ra sự kết nối mạnh mẽ và đồng thuận với khách hàng thông qua việc cung cấp các món ăn phù hợp và đáp ứng đúng đắn các nhu cầu dinh dưỡng và thị hiếu của họ.

# **3. Phương pháp**

Phương pháp được áp dụng trong dự án này bao gồm một loạt các bước từ thu thập dữ liệu, phân tích, xây dựng mô hình, đến thử nghiệm và đánh giá hiệu suất, sau đó là triển khai và tối ưu hóa. Đầu tiên, chúng tôi sẽ tiến hành thu thập dữ liệu từ các thực đơn của các nhà hàng, bao gồm thông tin về các món ăn, thành phần, cách chế biến, và các thông tin dinh dưỡng như lượng calo, protein, carbohydrate, chất béo, và các vitamin và khoáng chất khác.

Sau khi thu thập dữ liệu, chúng tôi sẽ tiến hành phân tích để hiểu rõ về hàm lượng dinh dưỡng của mỗi món ăn và tìm ra các yếu tố quan trọng ảnh hưởng đến sự lựa chọn của người dùng. Phân tích này sẽ bao gồm các phương pháp thống kê để xác định các xu hướng và mối quan hệ giữa các biến.

Tiếp theo, chúng tôi sẽ phát triển một mô hình dự đoán để đề xuất các món ăn phù hợp dựa trên dữ liệu về hàm lượng dinh dưỡng và sở thích của người dùng. Mô hình này có thể sử dụng các kỹ thuật học máy như học sâu, học máy có giám sát, hoặc học máy không giám sát tùy thuộc vào cách tiếp cận cụ thể.

Sau khi xây dựng mô hình, chúng tôi sẽ tiến hành các bài kiểm tra và đánh giá hiệu suất của hệ thống khuyến nghị. Điều này bao gồm việc sử dụng dữ liệu thử nghiệm để đánh giá độ chính xác của mô hình và thu thập phản hồi từ người dùng thực tế để cải thiện hệ thống.

Cuối cùng, sau khi mô hình đã được thử nghiệm và đánh giá, chúng tôi sẽ triển khai hệ thống vào môi trường thực tế và tiếp tục tối ưu hóa dựa trên phản hồi và dữ liệu mới. Quá trình này có thể bao gồm việc điều chỉnh mô hình, cải thiện thuật toán, và tối ưu hóa các thông số để đảm bảo rằng hệ thống hoạt động một cách hiệu quả và chính xác nhất có thể. Phương pháp này kết hợp các kỹ thuật phân tích dữ liệu, học máy, và đánh giá hiệu suất để xây dựng và triển khai một hệ thống khuyến nghị dinh dưỡng hiệu quả. Chúng tôi hy vọng rằng dự án này sẽ mang lại lợi ích lớn cho cả người dùng và các nhà hàng, giúp cải thiện sức khỏe và trải nghiệm ăn uống của mọi người.

**4. Phạm vi dự án**

Dự án "Phân tích hàm lượng dinh dưỡng của món ăn từ thực đơn các nhà hàng để đưa ra đề xuất phù hợp với nhu cầu người dùng" sẽ tập trung vào một loạt các mục tiêu cụ thể và sẽ có phạm vi như sau:

**Thu thập dữ liệu:**

Thu thập thông tin từ các thực đơn của các nhà hàng, bao gồm các món ăn, thành phần, cách chế biến, và thông tin dinh dưỡng chi tiết.

Xác định các nguồn dữ liệu đáng tin cậy và đa dạng để đảm bảo tính toàn vẹn và độ phong phú của dữ liệu.

**Phân tích dữ liệu:**

Tiến hành phân tích chi tiết về hàm lượng dinh dưỡng của mỗi món ăn từ dữ liệu thu thập.

Xác định các yếu tố quan trọng ảnh hưởng đến sự lựa chọn của người dùng, bao gồm cả yếu tố dinh dưỡng và sở thích cá nhân.

**Xây dựng mô hình:**

Phát triển một mô hình dự đoán có khả năng đề xuất các món ăn phù hợp dựa trên dữ liệu dinh dưỡng và sở thích cá nhân của người dùng.

Sử dụng các kỹ thuật học máy như học sâu, học máy có giám sát, hoặc học máy không giám sát để xây dựng mô hình dự đoán hiệu quả.

**Thử nghiệm và đánh giá hiệu suất:**

Tiến hành các bài kiểm tra để đánh giá độ chính xác và hiệu suất của hệ thống khuyến nghị.

Sử dụng dữ liệu thử nghiệm và thu thập phản hồi từ người dùng để cải thiện và tối ưu hóa mô hình.

**Triển khai và tối ưu hóa:**

Triển khai hệ thống vào môi trường thực tế và tiếp tục tối ưu hóa dựa trên phản hồi và dữ liệu mới.

Điều chỉnh mô hình, cải thiện thuật toán, và tối ưu hóa các thông số để đảm bảo tính hiệu quả và chính xác của hệ thống.

Dự án sẽ tập trung vào việc kết hợp các phương pháp phân tích dữ liệu và học máy để xây dựng và triển khai một hệ thống khuyến nghị dinh dưỡng hiệu quả, nhằm cung cấp lợi ích cho cả người dùng và các nhà hàng.

**CHƯƠNG 2: GIỚI THIỆU VỀ BỘ DỮ LIỆU**

# **1. Giới thiệu 3 bộ dữ liệu:**

## **1.1 Bộ 1:**

**Mô tả:**

Tên: Starbucks.csv

Mô tả: Hàm lượng dinh dưỡng của các loại đồ uống có trong menu của chuỗi cửa hàng cà phê Starbucks

**Trường thông tin:**

A screenshot of a computer screen

Description automatically generatedA screenshot of a computer screen

Description automatically generated

**Số liệu thống kê:**

Số hàng: 242

Số cột: 18

Số lượng sản phẩm theo nhóm:

A bar graph with different colored bars

Description automatically generated

A table with numbers and numbers

Description automatically generated

**Chất lượng của bộ dữ liệu:**

Nguồn: <https://www.kaggle.com/datasets/henryshan/starbucks>

Thời gian cập nhật gần nhất: 3 tháng trước

Số giá trị không rỗng: 4355

Số giá trị rỗng : 1

Số giá trị 'varies' (không sử dụng được): 12

Các giá trị trùng lặp và số lượng của chúng:

Beverage\_category:

Tazo® Full-Leaf Red Tea Latte (Vanilla Rooibos)Soymilk 4

Tazo® Chai Tea LatteSoymilk 4

CappuccinoSoymilk 4

Cappuccino2% Milk 4

Caramel Macchiato2% Milk 4

Caramel MacchiatoSoymilk 4

White Chocolate Mocha (Without Whipped Cream)2% Milk 4

White Chocolate Mocha (Without Whipped Cream)Soymilk 4

Vanilla Latte (Or Other Flavoured Latte)Soymilk 4

Vanilla Latte (Or Other Flavoured Latte)2% Milk 4

Hot Chocolate (Without Whipped Cream)2% Milk 4

Tazo® Chai Tea Latte2% Milk 4

Hot Chocolate (Without Whipped Cream)Soymilk 4

Caffè Mocha (Without Whipped Cream)Soymilk 4

Caffè Latte2% Milk 4

Tazo® Full-Leaf Red Tea Latte (Vanilla Rooibos)2% Milk 4

Tazo® Green Tea Latte2% Milk 4

Tazo® Green Tea LatteSoymilk 4

Tazo® Full-Leaf Tea LatteSoymilk 4

Tazo® Full-Leaf Tea Latte2% Milk 4

Caffè Mocha (Without Whipped Cream)2% Milk 4

Caffè LatteSoymilk 4

Caramel (Without Whipped Cream)Whole Milk 3

Iced Brewed Coffee (With Milk & Classic Syrup)2% Milk 3

Mocha (Without Whipped Cream)Soymilk 3

CoffeeSoymilk 3

CoffeeWhole Milk 3

Iced Brewed Coffee (With Milk & Classic Syrup)Soymilk 3

Java Chip (Without Whipped Cream)Whole Milk 3

Caramel (Without Whipped Cream)Soymilk 3

Mocha (Without Whipped Cream)Whole Milk 3

Java Chip (Without Whipped Cream)Soymilk 3

Strawberries & Crème (Without Whipped Cream)Whole Milk 3

Strawberries & Crème (Without Whipped Cream)Soymilk 3

CoffeeGrande Nonfat Milk 2

CoffeeTall Nonfat Milk 2

CoffeeVenti Nonfat Milk 2

Name: count, dtype: int64

Số lượng giá trị độc nhất trong cột 'Merged\_Column': 149

**Phân tích chuyên sâu về bộ dữ liệu:**

Bộ dữ liệu này chứa thông tin dinh dưỡng cho các món trong thực đơn của Starbucks.

Cột **'Beverage\_category'** phân loại loại đồ uống, chẳng hạn như cà phê, trà hoặc sinh tố.

Cột **'Beverage'** cung cấp tên cụ thể của từng thức uống, ví dụ như Caramel Macchiato

hoặc Green Tea Latte.

Cột **'Beverage\_prep'** mô tả chi tiết về phương pháp pha chế đồ uống, bao gồm cả việc

nó được phục vụ nóng hay lạnh, và các thành phần hoặc lớp phủ bổ sung như kem đánh

hoặc xi-rô.

**Cột 'Calories'** liệt kê tổng hàm lượng calo của mỗi loại đồ uống, cung cấp cái nhìn tổng quan về năng lượng được cung cấp bởi từng thức uống.

Ba cột tiếp theo, **'Total Fat (g)', 'Trans Fat (g)',** và **'Saturated Fat (g)'**, cung cấp thông tin về hàm lượng chất béo trong mỗi loại đồ uống. Những thông tin này đặc biệt quan

trọng đối với những người quan tâm đến lượng chất béo trong chế độ ăn uống của họ vì

lý do sức khỏe.

Cột **'Sodium (mg)'** cho biết lượng natri trong mỗi loại đồ uống, cung cấp thông tin quan

trọng cho những người theo dõi lượng natri tiêu thụ hàng ngày của họ.

Cột **'Total Carbohydrates (g)'** cung cấp tổng hàm lượng carbohydrate, bao gồm đường, rất hữu ích cho những người quản lý bệnh tiểu đường hoặc theo chế độ ăn low-carb.

Cột **'Cholesterol (mg)'** liệt kê lượng cholesterol trong mỗi loại đồ uống, một yếu tố quan trọng đối với những người quan tâm đến sức khỏe tim mạch của họ.

Cột **‘Dietary Fiber (g)’** : Đây là lượng chất xơ trong mỗi đơn vị đồ uống, được đo bằng

đơn vị gram (g). Chất xơ là một yếu tố quan trọng trong chế độ ăn uống, giúp cải thiện

tiêu hóa và duy trì sức khỏe ruột.

Cột **‘Sugars (g)’**: Đây là lượng đường trong mỗi đơn vị đồ uống, được đo bằng đơn vị gram (g). Việc tiêu thụ đường quá mức có thể gây ra các vấn đề về sức khỏe như béo phì và bệnh tiểu đường.

Cột **‘Protein (g)’**: Đây là lượng protein trong mỗi đơn vị đồ uống, được đo bằng đơn vị gram (g). Protein là một phần quan trọng của một chế độ ăn uống cân đối, giúp cơ thể phục hồi và phát triển cơ bắp.

Cột **‘Vitamin A (% DV)**’: Phần trăm giá trị hàng ngày (% DV) của vitamin A trong mỗi đơn vị đồ uống. Vitamin A là một chất dinh dưỡng quan trọng cho sức khỏe của mắt, da và hệ miễn dịch.

Cột **‘Vitamin C (% DV)**’: Phần trăm giá trị hàng ngày (% DV) của vitamin C trong mỗi đơn vị đồ uống. Vitamin C giúp hỗ trợ hệ miễn dịch và cải thiện sức khỏe của da.

Cột **‘Calcium (% DV)’**: Phần trăm giá trị hàng ngày (% DV) của canxi trong mỗi đơn vị đồ uống. Canxi là một khoáng chất quan trọng giúp duy trì sức khỏe xương và răng.

Cột **‘Iron (% DV)’**: Phần trăm giá trị hàng ngày (% DV) của sắt trong mỗi đơn vị đồ uống. Sắt là một thành phần chính của hồng cầu và giúp cung cấp năng lượng cho cơ thể.

Cuối cùng là cột **‘Caffeine (mg)**’: Đây là lượng caffeine trong mỗi đơn vị đồ uống, được đo bằng đơn vị milligram (mg). Caffeine là một chất kích thích tự nhiên có thể giúp tăng cường tinh thần và tăng cường sự tỉnh táo.

Bộ dữ liệu này là một tài nguyên quý giá cho các nhà nghiên cứu, chuyên gia dinh dưỡng và người tiêu dùng quan tâm đến sức khỏe và dinh dưỡng. Nó cung cấp cái nhìn tổng quan về hàm lượng dinh dưỡng của các món đồ uống tại Starbucks, từ đó cho phép thực hiện phân tích và so sánh chi tiết. Nhờ đó, chúng ta có thể hiểu rõ hơn về các yếu tố dinh dưỡng và tác động của chúng đối với sức khỏe.

## **Bộ 2:**

**Mô tả:**

Tên: dairy-queen-menu-nutrition.csv

Mô tả: Hàm lượng dinh dưỡng của các món ăn có trong menu của chuỗi cửa hàng thức ăn nhanh Dairy Queen

**Trường thông tin:**

A screenshot of a computer screen

Description automatically generatedA screenshot of a computer

Description automatically generated

**Số liệu thống kê:**

Số hàng: 229

Số cột: 13

A table with numbers and numbers

Description automatically generated

Phân phối lượng đường của Blizzards, Shakes và Sundaes ở các kích cỡ khác nhau:

A graph of sugar distribution

Description automatically generatedA graph of different sizes of sugar

Description automatically generated

A graph of sugar distribution

Description automatically generated

**Chất lượng của bộ dữ liệu:**

Nguồn:

https://www.kaggle.com/code/mattop/dairy-queen-menu-nutrition-eda

Thời gian cập nhật gần nhất: 1 năm trước

Số giá trị không rỗng: 2959

Số giá trị rỗng: 18

Các giá trị trùng lặp và số lượng của chúng:

Menu Item

Hot Dog 2

Hamburger 2

Name: count, dtype: int64

Số lượng giá trị độc nhất trong cột 'Menu Item': 227

**Phân tích chuyên sâu về bộ dữ liệu:**

Cột **‘Menu Item’**: Tên cụ thể của món ăn hoặc đồ uống.

Cột **‘Calories (kcal)’**: Số lượng calo mà mỗi món cung cấp. Calo là đơn vị đo năng lượng từ thức ăn và đồ uống, và việc theo dõi lượng calo là quan trọng để duy trì cân nặng và sức khỏe tổng thể.

Cột **‘Fat Calories (kcal)’**: Số calo được cung cấp từ chất béo trong mỗi món. Chất béo là một nguồn năng lượng quan trọng, nhưng việc tiêu thụ quá nhiều chất béo, đặc biệt là

chất béo bão hòa và chất béo chuyển hóa, có thể góp phần vào tăng cân và các vấn đề sức

khỏe liên quan đến tim mạch.

Ba cột tiếp theo, **'Total Fat (g)', 'Trans Fat (g)',** và **'Saturated Fat (g)'**, cung cấp thông tin về hàm lượng chất béo trong mỗi món. Những thông tin này đặc biệt quan

trọng đối với những người quan tâm đến lượng chất béo trong chế độ ăn uống của họ vì

lý do sức khỏe.

Cột **‘Cholesterol (mg)’**: Lượng cholesterol trong mỗi món, được đo bằng miligam. Cholesterol là một loại chất béo mà cơ thể cần để sản xuất hormon và tạo cấu trúc tế bào, nhưng việc tiêu thụ quá mức có thể góp phần vào tăng nguy cơ mắc các bệnh tim mạch.

Cột **‘Sodium (mg)’**: Lượng natri trong mỗi món, được đo bằng miligam. Natri là một khoáng chất cần thiết cho cơ thể, nhưng việc tiêu thụ quá mức có thể gây ra các vấn đề sức khỏe như tăng huyết áp.

Cột **‘Carbohydrates (g)’**: Tổng lượng carbohydrate trong mỗi món, được đo bằng gram. Carbohydrate là một nguồn năng lượng chính cho cơ thể, nhưng việc tiêu thụ quá mức có thể dẫn đến tăng cân và các vấn đề sức khỏe liên quan đến đường huyết.

Cột **‘Fiber (g)’**: Lượng chất xơ trong mỗi món, được đo bằng gram. Chất xơ là quan trọng cho sức khỏe tiêu hóa và có thể giúp giảm nguy cơ mắc bệnh tim mạch và tiểu đường.

Cột **‘Sugars (g)’**: Tổng lượng đường trong mỗi món, được đo bằng gram. Đường là một loại carbohydrate đơn đường và việc tiêu thụ quá mức có thể gây ra tăng cân và các vấn đề sức khỏe khác như đái tháo đường.

Cột **‘Protein (g)’**: Lượng protein trong mỗi món, được đo bằng gram. Protein là một phần quan trọng của một chế độ ăn lành mạnh và cần thiết cho sự phát triển và sửa chữa các tế bào trong cơ thể.

Cột **‘Allergens:’** Các chất gây dị ứng có thể có trong mỗi món, bao gồm M (Sữa), E (Trứng), W (Đậu), P (Đậu phộng), SF (Hải sản), và T (Hạt). Các chất này có thể gây ra phản ứng dị ứng hoặc vấn đề sức khỏe khác đối với những người có khả năng dị ứng hoặc cảm giác mẫn cảm.

Dữ liệu này cung cấp một cái nhìn toàn diện về hàm lượng dinh dưỡng của các món ăn và đồ uống tại Dairy Queen. Thông qua việc phân tích các chỉ số như calo, chất béo, carbohydrate, protein, cholesterol và natri, người đọc có thể hiểu được tác động của từng món lên sức khỏe của họ. Đặc biệt, việc liệt kê các chất gây dị ứng cũng rất hữu ích đối với những người có những ràng buộc dinh dưỡng đặc biệt hoặc cảm giác mẫn cảm. Dữ liệu này có thể hỗ trợ nhà nghiên cứu, chuyên gia dinh dưỡng và người tiêu dùng trong việc đưa ra các quyết định thông minh về chế độ ăn uống và lối sống lành mạnh.

## **1.3 Bộ 3:**

**Mô tả:**

Tên: ms\_annual\_data\_2022.csv

Mô tả: Hàm lượng dinh dưỡng của các loại thực phẩm có trong menu của một số chuỗi cửa hàng lớn năm 2022.

**Trường thông tin:**

A screen shot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Số liệu thống kê:**

Số hàng: 26238

Số cột: 14

A graph with blue bars

Description automatically generated

A screenshot of a data sheet

Description automatically generated

**Chất lượng của bộ dữ liệu:**

Nguồn:

https://www.menustat.org/uploads/1/4/1/6/141624194/ms\_annual\_data\_2022.xlsx

Số giá trị không rỗng: 356607

Số giá trị rỗng: 10725

Các giá trị trùng lặp và số lượng của chúng:

item\_name

Coffee 23

Sprite 22

Diet Coke 21

Cheddar Cheese 18

Grilled Chicken 18

..

French Fries, Kids 2

Bud Light Lime 2

Whipped Cream 2

Sprite Zero, 32 fl oz 2

Sprite Zero, 22 fl oz 2

Name: count, Length: 1579, dtype: int64

Số lượng giá trị độc nhất trong cột 'item\_name': 22886

**Phân tích chuyên sâu về bộ dữ liệu:**

Cột **‘food-category’:** Tên loại của món ăn hoặc đồ uống.

Cột **‘restaurant’** : Tên các nhà hàng của món ăn, thức uống đó

Cột **‘item\_name’** : Tên đồ ăn hoặc thức uống

Cột **‘item**-**description**’ : Mô tả món ăn hoặc thức uống đó

Cột **‘Calories (kcal)’**: Đây là lượng calo trong mỗi phần thực phẩm. Nó là một chỉ số quan trọng để người tiêu dùng đánh giá mức độ năng lượng của một món ăn.

Ba cột tiếp theo, **'Total Fat (g)', 'Trans Fat (g)',** và **'Saturated Fat (g)'**, cung cấp thông tin về hàm lượng chất béo trong mỗi món. Những thông tin này đặc biệt quan

trọng đối với những người quan tâm đến lượng chất béo trong chế độ ăn uống của họ vì

lý do sức khỏe.

Cột **‘Cholesterol (mg)’**: Lượng cholesterol trong mỗi phần. Cholesterol là một yếu tố quan trọng đối với sức khỏe tim mạch và cơ bản, nhưng nên được tiêu thụ trong giới hạn.

Cột **‘Sodium (mg)’**: Lượng natri trong mỗi phần. Sử dụng quá nhiều natri có thể dẫn đến tăng huyết áp và các vấn đề về sức khỏe tim mạch.

Cột ‘**Carbohydrates (g)’**: Lượng carbohydrate trong mỗi phần, bao gồm cả đường và chất xơ. Carbohydrate là nguồn năng lượng chính cho cơ thể.

Cột **‘Dietary Fiber (g)’**: Lượng chất xơ dinh dưỡng trong mỗi phần. Chất xơ giúp cải thiện tiêu hóa và làm giảm nguy cơ bệnh tim mạch.

Cột **‘Sugar (g)’**: Lượng đường trong mỗi phần. Sử dụng quá nhiều đường có thể gây ra các vấn đề sức khỏe như tăng cân, răng sâu và nguy cơ mắc bệnh tiểu đường.

Cột **‘Protein (g)’**: Lượng protein trong mỗi phần. Protein là thành phần cần thiết cho sự phát triển cơ bắp và sự phục hồi của cơ thể.

Tổng quan về dữ liệu dinh dưỡng của các món ăn trong menu của một số chuỗi cửa hàng lớn năm 2022 cho thấy sự đa dạng và phong phú trong các lựa chọn ăn uống. Bộ dữ liệu cung cấp thông tin chi tiết về lượng calo, chất béo, natri, carbohydrate, protein, chất xơ và đường trong mỗi món ăn, cùng với thông tin về các chất gây dị ứng có thể có.

Phân tích sâu hơn các thuộc tính dinh dưỡng giúp chúng ta hiểu rõ hơn về ảnh hưởng của từng thành phần đối với sức khỏe. Điều này giúp người tiêu dùng có thể làm quen với thông tin dinh dưỡng và đưa ra quyết định thông minh khi chọn lựa món ăn phù hợp với nhu cầu dinh dưỡng và yêu cầu sức khỏe của họ.

Tuy nhiên, để đảm bảo sự chính xác và đầy đủ, việc phân tích cần được kết hợp với kiến thức chuyên môn trong lĩnh vực dinh dưỡng và y tế. Đồng thời, thông tin về chất gây dị ứng cũng rất quan trọng để đảm bảo an toàn cho người tiêu dùng.

# **2. So sánh sâu 3 bộ dữ liệu**

**Giống:**

* Cả ba bộ dữ liệu đều cung cấp thông tin dinh dưỡng chi tiết về các món ăn và đồ uống trong menu của họ.
* Mỗi bộ dữ liệu đều có các cột chứa thông tin về calo, chất béo, carbohydrate và protein, cho phép người tiêu dùng đánh giá mức độ dinh dưỡng của mỗi món ăn.
* Cả ba bộ dữ liệu đều cung cấp thông tin về kích thước phục vụ và các loại thực phẩm, giúp người tiêu dùng lựa chọn dựa trên nhu cầu của họ.Mỗi bộ đều có những cột dữ liệu không đầy đủ hoặc không sử dụng cho đề tài nhưng hoàn toàn có thể xử lý bằng các phương pháp ghép cột, xóa cột đơn giản.
* Đều là dữ liệu từ các nhà hàng cửa hàng nổi tiếng nên dễ tiếp cận đối với đối

tượng của đề tài

**Khác:**

**starbucks.csv:**

* Toàn bộ là đồ uống
* Phân loại theo nhóm đồ uống
* Thêm thông tin về Vitamin A, Vitamin C, Cafein, Beverage\_prep ( Cách pha chế)
* Cùng một món nhưng thay đổi cách pha chế, liều lượng thì thành phần dinh dưỡng cũng thay đổi cho nên cần ghép 2 cột tương ứng trong bộ dữ liệu này dễ thống kê hơn.
* Khá nhiều dữ liệu trùng do đặc điểm trên.
* Dữ liệu khá mới
* Bộ dữ liệu từ Starbucks cung cấp thông tin về hàm lượng dinh dưỡng của các đồ uống có trong menu của chuỗi cửa hàng cà phê. Điều này cho thấy sự chăm sóc đặc biệt đến chất lượng và thông tin dinh dưỡng của sản phẩm. Dữ liệu này có cấu trúc rõ ràng với nhiều cột thông tin, giúp người dùng dễ dàng tìm hiểu về các thành phần và giá trị dinh dưỡng của các sản phẩm Starbucks. Tuy nhiên, việc phát hiện một số giá trị rỗng và giá trị không hợp lệ như "varies" có thể cần phải được xử lý để đảm bảo tính toàn vẹn của dữ liệu.

**dairy-queen-menu-nutrition.csv:**

* Đa dạng thực phẩm, chủ yếu là thức ăn nhanh.
* Phân loại theo kích cỡ.
* Không có dữ liệu rỗng hay không sử dụng được.
* Gần như không có dữ liệu trùng lặp.
* Dữ liệu khá mới
* Bộ dữ liệu từ Dairy Queen tập trung vào thông tin dinh dưỡng của các món ăn trong menu của chuỗi cửa hàng thức ăn nhanh Dairy Queen. Dữ liệu này tương đối đơn giản với ít cột thông tin hơn so với Starbucks, tập trung chủ yếu vào các mục như Blizzards, Shakes và Sundaes. Sự tập trung này có thể phản ánh chính sách và mục tiêu kinh doanh của Dairy Queen, tập trung vào các sản phẩm phổ biến nhất hoặc có tiềm năng bán chạy nhất trong menu.

**ms\_annual\_data\_2022.csv:**

* Đa dạng thực phẩm tại nhiều nhà hàng khác nhau.
* Phân loại theo nhóm thực phẩm.
* Có nhiều cột dữ liệu null, không thể sử dụng cần loại bỏ.
* Khá nhiều dữ liệu trùng lặp do khác nhà hàng nhưng cùng tên món.
* Dữ liệu khá cũ (2022) nên độ chính xác có thể bị ảnh hưởng
* Bộ dữ liệu từ ms\_annual\_data\_2022 cung cấp thông tin về hàm lượng dinh dưỡng của các loại thực phẩm có trong menu của một số chuỗi cửa hàng lớn trong năm 2022. Với số lượng hàng hóa lớn và đa dạng, dữ liệu này mang tính phản ánh cao hơn về ngành công nghiệp ẩm thực. Tuy nhiên, việc phát hiện một số lượng lớn các giá trị rỗng cần được xử lý để đảm bảo tính chính xác của các phân tích tiếp theo.

**Các so sánh về dữ liệu giữa 3 bộ dữ liệu:**

A graph with blue and white bars

Description automatically generated

A graph of blue rectangles

Description automatically generated

A blue bar graph with white text

Description automatically generated

A graph with blue and white text

Description automatically generated with medium confidence

**Về khả năng cập nhật và mở rộng của ba bộ dữ liệu Starbucks.csv, dairy-queen-menu-nutrition.csv và ms\_annual\_data\_2022.csv, chúng ta cần xem xét các yếu tố như nguồn gốc của dữ liệu, tần suất cập nhật, và tính đa dạng của thông tin:**

Bộ dữ liệu từ Starbucks.csv được thu thập từ menu của chuỗi cửa hàng cà phê Starbucks, một thương hiệu lớn với sự phổ biến toàn cầu. Tuy nhiên, do tính chất cố định của menu, dữ liệu này có khả năng cập nhật thấp hơn so với các bộ dữ liệu khác. Thông tin về các loại đồ uống có thể thay đổi ít hơn so với các loại thực phẩm trong các nhà hàng fast food hoặc các chuỗi cửa hàng thực phẩm nhanh.

Dữ liệu từ dairy-queen-menu-nutrition.csv cũng có tính đặc thù trong việc cập nhật. Các món ăn trong menu của Dairy Queen có thể ít thay đổi hơn so với menu của các nhà hàng fast food khác nhưng cũng có thể chứa các phiên bản món ăn mới được giới thiệu theo mùa hoặc theo yêu cầu của thị trường. Tuy nhiên, việc cập nhật dữ liệu này có thể không thường xuyên và đòi hỏi theo dõi các thông tin cập nhật từ Dairy Queen.

Bộ dữ liệu từ ms\_annual\_data\_2022.csv có khả năng cập nhật và mở rộng cao nhất trong số ba bộ dữ liệu. Dữ liệu này thu thập thông tin từ một số chuỗi cửa hàng lớn và có khả năng bổ sung các thông tin mới theo thời gian. Nếu nguồn thông tin này tiếp tục được duy trì và cập nhật định kỳ, nó có thể cung cấp một nguồn tài nguyên phong phú và đáng tin cậy cho việc phân tích và đề xuất món ăn dựa trên nhu cầu dinh dưỡng của người dùng.

**Khi so sánh mức độ áp dụng của ba bộ dữ liệu trong đề tài "đề xuất món ăn phù hợp với nhu cầu dinh dưỡng người dùng", ta rút ra nhận xét sau:**

Bộ dữ liệu từ Starbucks.csv cung cấp thông tin về hàm lượng dinh dưỡng của các đồ uống có sẵn trong menu của chuỗi cửa hàng cà phê Starbucks. Tuy nhiên, với trọng tâm chủ yếu là đồ uống, dữ liệu này có thể hạn chế trong việc tạo ra các đề xuất món ăn đầy đủ và đa dạng cho nhu cầu dinh dưỡng. Thực đơn tại Starbucks thường tập trung vào các loại cà phê, trà và đồ uống có đường, với ít sự lựa chọn về thức ăn.

Bộ dữ liệu từ dairy-queen-menu-nutrition.csv tập trung vào hàm lượng dinh dưỡng của các món ăn trong menu của chuỗi cửa hàng thức ăn nhanh Dairy Queen. Với sự đa dạng hơn về loại món ăn như Blizzards, Shakes và Sundaes, dữ liệu này có thể cung cấp một phạm vi rộng hơn cho việc tạo ra các đề xuất món ăn phù hợp với nhu cầu dinh dưỡng. Tuy nhiên, cần phải chú ý đến tính chất của các món ăn này, có thể chứa nhiều đường và calo cao.

Bộ dữ liệu từ ms\_annual\_data\_2022.csv cung cấp thông tin về hàm lượng dinh dưỡng của các loại thực phẩm có trong menu của một số chuỗi cửa hàng lớn trong năm 2022. Với quy mô lớn và đa dạng về loại thực phẩm, dữ liệu này có thể cung cấp một nguồn tài nguyên phong phú cho việc tạo ra các đề xuất món ăn đa dạng và phù hợp với nhu cầu dinh dưỡng của người dùng. Tuy nhiên, cần xem xét việc lọc và chọn lọc thông tin để tạo ra các đề xuất có ý nghĩa và thực tiễn

**3. Chọn ra dữ liệu làm đề tài**

Chúng ta sẽ chọn bộ dữ liệu 3 (ms\_annual\_data\_2022.csv) làm đề tài vì nó có

những đặc điểm sau:

Đa dạng thực phẩm từ nhiều nhà hàng khác nhau: Bộ dữ liệu này cung cấp thông tin về

nhiều loại thực phẩm từ nhiều nhà hàng khác nhau, từ đồ ăn nhanh đến các món ăn hấp

dẫn từ nhà hàng cao cấp. Điều này mang lại sự đa dạng và phong phú cho việc nghiên

cứu và phân tích.

Phân loại theo nhóm thực phẩm: Dữ liệu được phân loại theo nhóm thực phẩm, giúp dễ

dàng hiểu biết và so sánh các loại thực phẩm trong cùng một nhóm, từ đó đưa ra các nhận định và kết luận hợp lý.

Dữ liệu có độ tin cậy tương đối: Mặc dù có một số cột dữ liệu rỗng và cần loại bỏ, nhưng dữ liệu trong bộ này có thể được xử lý để đảm bảo tính chính xác và độ tin cậy cao.

Dữ liệu có độ cũ phù hợp: Dù là dữ liệu từ năm 2022, nhưng vẫn có thể cung cấp thông tin hữu ích và đáng tin cậy cho việc nghiên cứu và phân tích hiện tại.

Khả năng phân tích sâu: Với sự đa dạng của dữ liệu và thông tin chi tiết về các loại thực phẩm và nhóm thực phẩm, bộ dữ liệu này mở ra nhiều cơ hội để thực hiện các phân tích sâu hơn về dinh dưỡng và thói quen ăn uống của người tiêu dùng.

Tóm lại, bộ dữ liệu ms\_annual\_data\_2022.csv được chọn làm đề tài vì tính đa dạng, thông tin phong phú và khả năng phân tích sâu về dinh dưỡng và thói quen ăn uống của người tiêu dùng.

# **3. Kết luận về 3 bộ dữ liệu:**

Các bộ dữ liệu được sử dụng trong dự án bao gồm Starbucks.csv, dairy-queen-menu-nutrition.csv và ms\_annual\_data\_2022.csv. Mỗi bộ dữ liệu có những đặc điểm riêng, nhưng đều cung cấp đủ thông tin về hàm lượng dinh dưỡng của các món ăn từ các nhà hàng khác nhau. Trong đó, Starbucks.csv chủ yếu tập trung vào các đồ uống với các thông tin chi tiết về cách pha chế, dairy-queen-menu-nutrition.csv tập trung vào thức ăn nhanh với phân loại theo kích cỡ, và ms\_annual\_data\_2022.csv cung cấp đa dạng thực phẩm từ nhiều nhà hàng khác nhau.

**Chương 3: DEMO**

1. **Báo cáo**

**BÁO CÁO**

Đầu tiên chúng ta tải dữ liệu từ trên google drive về và lọc dữ liệu.

Xóa hết các hàng có trạng thái null vì có thể ảnh hưởng tới quá trình đi tìm

Chuyển các cột dinh dưỡng sang kiểu số

Viết hàm tính cân nặng lý tưởng, nhu cầu năng lượng , nhu cầu riêng của các chất dinh dưỡng, chúng ta có thể tham khảo cách tính ở [CÁCH TÍNH NHU CẦU NĂNG LƯỢNG VÀ CÁC CHẤT DINH DƯỠNG. | PHAN THÀNH HIẾU (wordpress.com)](https://phanthanhhieu82.wordpress.com/2014/10/30/cach-tinh-nhu-cau-nang-luong-va-cac-chat-dinh-duong/?utm_source=zalo&utm_medium=zalo&utm_campaign=zalo)

Chúng ta sẽ viết **hàm đầu tiên**, đầu vào là dữ liệu, các cột dinh dưỡng mục tiêu. Ta sẽ chọn các chất dinh dưỡng có cách tính cụ thể như là calories, total\_fat, carbohydrate, protein và số lần lặp có thế là 1000 hay 2000…

Chọn ngẫu nhiên 3 món ăn rồi tính tổng các chất dinh dưỡng như trên, rồi chúng ta tìm độ lệch của chúng cộng lại, rồi chọn độ lệch nhỏ nhất. Chạy 1000 lần như vậy ta được bộ ba món ăn có thể làm thực đơn trong ngày.

Tiếp theo, chúng ta có thể nhập chiều cao, giới tính và mức độ lao động (nhẹ, trung bình hoặc nặng) . Dựa vào đó chúng ta tính được cân nặng lý tưởng, nhu cầu riêng của các chất dinh dưỡng. Chúng ta đưa các dữ liệu vào hàm chạy, chạy xong ta được thực đơn một ngày người này có thể ăn.

**Thuận toán thứ 2**, chúng ta dùng KNN .

Vẫn là như vậy chọn những đặc trưng cần thiết có thể tính được trên công thức.

Tạo mô hình KNN có sẵn, chúng ta có thể tìm ra k món ăn tùy vào người dùng muốn bao nhiêu bữa. Chia các giá trị mục tiêu thành k phần vì khi cộng lại ta sẽ được một số gần mục tiêu nhất. Ta tính toán tổ hợp tốt nhất dựa trên mô hình KNN bằng cách tìm k hàng xóm gần nhất của các giá trị mục tiêu nhỏ đó. Rồi ta lưu lại các tên món ăn đó, đề xuất cho người dùng.

Chúng ta cũng có thể tính sai số giữa đề xuất và mục tiêu.

1. **So sánh**

Giống nhau:

* Đều sử dụng sai số tương đối làm tiêu chuẩn đánh giá.
* Cùng sử dụng phương pháp tính lượng dinh dưỡng cần thiết dựa trên chiều cao, cân nặng để đưa ra đề xuất.
* Đều là những thuật toán hiệu quả để đề xuất món ăn (sai số tương đối khoảng 10%)

Khác nhau:

|  |  |  |
| --- | --- | --- |
|  | **Random** | **KNN** |
| Phương pháp | Thực hiện nhiều lần việc tính sai số 3 món ăn. Lấy sai số nhỏ nhất | Chọn 3 hàng xóm gần nhất với nhu cầu dinh dưỡng |
| Tính chất mẫu | Mẫu đầu ra có thể có số liệu khác nhau. | 3 mẫu gần giống nhau. |
| Sai số | Nhỏ hơn, không ổn định, phụ thuộc vào số lần thực hiện | Lớn hơn nhưng ổn định |
| Tốc độ chạy | Phụ thuộc vào số lần thực hiện | Phụ thuộc vào số lượng mẫu |
| Phạm vi áp dụng | Sử dụng đề đề xuất các món ăn trong 1 bữa | Đề xuất các món ăn trong nhiều bữa khác nhau |

1. **Kết luận**

Trong bối cảnh ngày nay, việc quan tâm đến sức khỏe và dinh dưỡng ngày càng trở nên quan trọng hơn bao giờ hết, và việc lựa chọn món ăn phù hợp với nhu cầu dinh dưỡng và thể chất là một phần không thể thiếu trong cuộc sống hàng ngày của mỗi người. Đối với những người bận rộn, việc dành thời gian để nấu ăn tại nhà không phải lúc nào cũng là sự lựa chọn hợp lý. Trong tình hình này, việc phân tích hàm lượng dinh dưỡng của các món ăn từ thực đơn của các nhà hàng trở nên vô cùng quan trọng và hữu ích.

Dự án "Phân Tích Hàm Lượng Dinh Dưỡng của Món Ăn từ Thực Đơn Các Nhà Hàng để Đưa Ra Đề Xuất Phù Hợp với Nhu Cầu Người Dùng" mang lại nhiều lợi ích đáng kể. Đầu tiên, nó giúp tiết kiệm thời gian và công sức của người tiêu dùng trong việc tìm kiếm món ăn phù hợp với nhu cầu dinh dưỡng và thị hiếu cá nhân. Thứ hai, thông qua việc đề xuất các món ăn phù hợp, dự án cung cấp một cái nhìn tổng quan và chính xác về hàm lượng dinh dưỡng của các món ăn từ các nhà hàng khác nhau, từ đó giúp nâng cao nhận thức về dinh dưỡng và lối sống lành mạnh cho cả cộng đồng. Cuối cùng, đối với các nhà hàng, dự án không chỉ là một công cụ hữu ích để tối ưu hóa thực đơn và tăng cường trải nghiệm của khách hàng mà còn là một cơ hội để quảng bá sản phẩm một cách hiệu quả và tạo ra sự kết nối mạnh mẽ với khách hàng thông qua việc cung cấp các món ăn đáp ứng đúng đắn các nhu cầu dinh dưỡng và thị hiếu của họ.

Dự án cũng đối mặt với nhiều thách thức trong quá trình triển khai và phát triển. Một trong những thách thức lớn nhất là xử lý dữ liệu lớn từ nhiều nguồn khác nhau. Việc này đòi hỏi sự chính xác và hiệu suất trong quá trình tiền xử lý và phân tích dữ liệu. Đồng thời, đảm bảo tính nhất quán của dữ liệu cũng là một thách thức đáng kể, vì phải đối mặt với việc thu thập và xử lý dữ liệu từ các nhà hàng và doanh nghiệp khác nhau. Cần phải có sự cẩn thận và kiểm soát chặt chẽ để đảm bảo dữ liệu thu thập là đầy đủ và chính xác.

Mặc dù đối diện với nhiều thách thức, dự án cũng mang lại nhiều cơ hội và triển vọng trong tương lai. Một trong những hướng phát triển tiềm năng là mở rộng phạm vi và tính ứng dụng của dự án. Bằng cách bao gồm nhiều nhà hàng và doanh nghiệp ẩm thực hơn, dự án có thể cung cấp một hệ thống phân tích dinh dưỡng toàn diện và đa dạng hơn. Đồng thời, việc phát triển các ứng dụng di động hoặc trang web giúp người tiêu dùng dễ dàng truy cập và tìm kiếm các món ăn phù hợp với nhu cầu dinh dưỡng của họ từ bất kỳ đâu. Cuối cùng, tích hợp tính năng tương tác xã hội và thu thập phản hồi từ người dùng giúp dự án ngày càng hoàn thiện và phù hợp với nhu cầu thực tế của cộng đồng.

Dự án không chỉ là một công cụ hữu ích trong việc tối ưu hóa lối sống lành mạnh, mà còn là một bước tiến quan trọng trong việc thúc đẩy nhận thức và hành động cho một cộng đồng khỏe mạnh hơn và hạnh phúc hơn. Bằng cách kết hợp sự sáng tạo và công nghệ, chúng ta không chỉ cung cấp thông tin về dinh dưỡng mà còn tạo ra một môi trường khuyến khích, nơi mọi người có thể thúc đẩy nhau để đạt được mục tiêu về sức khỏe và hạnh phúc