

# Xây dựng hệ thống khuyến nghị bữa ăn

Bùi Hồng Thiên, Nguyễn Văn Hữu Nghĩa and Nguyễn Hoài Bảo

Trường Đại Học Công Nghệ Thông Tin - Đại Học Quốc Gia  
TP.HCM, Việt Nam.

Contributing authors: [19520970@gm.uit.edu.vn](mailto:19520970@gm.uit.edu.vn);  
[19521900@gm.uit.edu.vn](mailto:19521900@gm.uit.edu.vn); [19520405@gm.uit.edu.vn](mailto:19520405@gm.uit.edu.vn);

## Tóm tắt nội dung

Nhóm chúng em thực hiện một hệ khuyến nghị bữa ăn cho người dùng. Do không có nguồn dữ liệu phù hợp nên chúng em quyết định tự xây dựng bộ dữ liệu. Nguồn dữ liệu chính chúng em thu thập từ trang web [www.cet.edu.vn](http://www.cet.edu.vn). Phương pháp nhóm chúng em sử dụng là content base và KNN. Hai phương pháp đưa ra khuyến nghị bữa ăn cho người dùng theo cách khác nhau. Do đánh giá được hệ khuyến nghị ở bài toán này rất khó nên nhóm chúng em chỉ đưa ra những phương pháp chủ quan và mang tính tương đối.

**Keywords:** Recommendation system - Hệ khuyến nghị - Bữa ăn

## 1 Giới thiệu

Thực phẩm luôn là trung tâm của cuộc sống con người. Ngoài không khí chúng ta thở, thức ăn là vật chất duy nhất, mà con người đưa vào cơ thể. Thuở sơ khai, con người phải đối mặt với nhiệm vụ xác định và thu thập thức ăn cho sự sống còn. Tại hiện tại, sự lựa chọn chế độ ăn uống đang trở nên quan trọng trong việc đáp ứng nhu cầu đa dạng, chẳng hạn như dinh dưỡng cơ bản, calo, hương vị, sức khỏe,... . Theo Tổ chức Đái tháo đường Quốc tế Liên bang, khoảng 415 triệu người trên toàn thế giới bị bệnh tiểu đường, và tỷ lệ mắc bệnh tiểu đường được dự đoán là tăng thêm hơn 50% vào năm 2040, trở thành một mối đe dọa lớn đối với sức khỏe toàn cầu. Yếu tố dinh dưỡng là một trong những yếu tố chính nguyên nhân của sự gia tăng đáng kể tỷ lệ

mắc bệnh béo phì và tiểu đường. Nghiên cứu về gánh nặng bệnh tật toàn cầu cũng chỉ ra các yếu tố chế độ ăn uống như là một đóng góp chính cho mức độ suy dinh dưỡng, béo phì và thừa cân, chế độ ăn uống không hợp lý dẫn đến 11 triệu ca tử vong sớm có thể tránh được mỗi năm.

Hệ khuyến nghị đang nổi lên như một lĩnh vực mới để cải thiện những vấn đề này. Là một nhiệm vụ quan trọng trong tính toán thực phẩm, thực phẩm giới thiệu ý định tìm mặt hàng thực phẩm phù hợp cho người dùng để đáp ứng nhu cầu cá nhân của họ, và do đó đóng một vai trò quan trọng trong sự lựa chọn chế độ ăn uống của con người. Một chế độ ăn uống cân bằng là rất quan trọng để duy trì sức khỏe thể chất của một người. Tuy nhiên, các chất dinh dưỡng cần được ăn vào rất khác nhau tùy thuộc vào sở thích ăn uống cá nhân và tình trạng sức khỏe. Do đó, làm thế nào để cung cấp khuyến nghị thực phẩm cá nhân theo yêu cầu cá nhân khác nhau là rất quan trọng. Thập kỷ qua đã chứng kiến sự phát triển nhanh chóng của dịch vụ internet và thiết bị di động. Nó đã được thuận tiện hơn để mọi người truy cập một lượng lớn nội dung đa phương tiện trực tuyến nội dung thực phẩm từ nhiều nguồn khác nhau, chẳng hạn như diễn đàn, mạng xã hội phương tiện truyền thông, các trang web chia sẻ công thức và các trang web đánh giá của khách hàng. Mặc dù sự tăng trưởng này cho phép người dùng có nhiều sự lựa chọn hơn, nhưng nó cũng mang đến vấn đề cho người dùng lựa chọn các mặt hàng thực phẩm ưa thích từ hàng nghìn ứng viên. Vì vậy, khuyến nghị thực phẩm là ngày càng trở nên cần thiết để phục vụ tiềm năng khổng lồ nhu cầu dịch vụ và có thể giúp người dùng dễ dàng khám phá một tập hợp nhỏ các mặt hàng thực phẩm thú vị và phù hợp với họ. So với khuyến nghị trong các lĩnh vực khác, khuyến nghị thực phẩm có những đặc điểm riêng. Ví dụ, thực phẩm sở thích là một bước quan trọng đối với khuyến nghị thực phẩm. Tuy nhiên, sở thích thực phẩm liên quan đến nhiều yếu tố khác nhau, chẳng hạn như sở thích mùi vị, sự khác biệt về nhận thức, hạn chế về nhận thức, sự quen thuộc về văn hóa và thậm chí cả ảnh hưởng di truyền. Do đó, nó làm cho việc học sở thích thực phẩm chính xác hơn khó khăn. Hơn nữa, khuyến nghị thực phẩm nên xem xét thêm thông tin ngữ cảnh. Bên cạnh thông tin bối cảnh cơ bản được chụp từ các thiết bị di động quen thuộc chẳng hạn như thời gian, thông tin về vị trí và môi trường (ví dụ: nhiệt độ và PM2.5), các tín hiệu khác nhau liên quan đến trạng thái cơ thể, chẳng hạn như các bước đã thực hiện, nhịp tim, chất lượng giấc ngủ, tăng tốc cơ thể và thậm chí cả tình cảm các trạng thái cũng có thể được ghi lại từ các thiết bị cảm biến mới, chẳng hạn như đồng hồ đeo tay, thiết bị theo dõi thể dục và vòng đeo tay. Tín hiệu có thể mô tả toàn diện tình trạng cơ thể thực tế của người dùng trong thời gian thực và có tầm quan trọng thiết yếu đối với thực phẩm sự giới thiệu.

Nghiên cứu đa phương tiện hiện tại đã đạt được tiến bộ lớn trong việc cải thiện trải nghiệm và hiệu suất đề xuất trong nhiều lĩnh vực như phim ảnh và POI, nhưng phần lớn bị tụt hậu trong miền thực phẩm. Theo hiểu biết tốt nhất của chúng em, mặc dù các công trình liên quan về khuyến nghị thực phẩm có nhận được nhiều sự chú ý hơn trong cộng đồng đa phương tiện, có rất ít đánh giá có hệ thống, cung cấp một thống nhất khuôn khổ và tóm tắt toàn

diện về những nỗ lực hiện tại trong khuyến nghị thực phẩm. Vì tiềm năng to lớn trong thực phẩm khuyến nghị, nên chúng em quyết định thực hiện đề tài khuyến nghị bữa ăn cho người dùng. Các mục tiêu của đề tài này là như sau: Xây dựng bộ dữ liệu (Phần 3); Giới thiệu các phương pháp nhóm áp dụng (Phần 4); Các kết quả nhóm thực nghiệm được (Phần 5) và phần cuối cùng là Kết Luận (Phần 6).

## 2 Công trình liên quan

Khuyến nghị ẩm thực là 1 lĩnh vực rộng lớn đòi hỏi nhiều nghiên cứu và cũng đồng thời chiếm được sự chú ý của giới nghiên cứu nhờ khả năng thay đổi linh hoạt dựa trên nhu cầu, sức khỏe của cá nhân. Đề án khuyến nghị ẩm thực này sử dụng content-based và KNN.

Trong [1], các nhà nghiên cứu đã sử dụng phương pháp content-based, dựa trên sở thích, lịch sử xem của người dùng về 1 món đồ thời trang từ đó rút trích ra các đặc trưng như loại đồ (váy, áo, quần...), màu sắc, kiểu dáng... Dù cho đây là khuyến nghị trong lĩnh vực thời trang nhưng áp dụng cùng phương thức ta có thể khuyến nghị các bữa ăn hoàn chỉnh dựa trên sở thích của người dùng về các món ăn.

Phương pháp còn lại là sử dụng KNN để xây dựng 1 hệ khuyến nghị đồ ăn như [2]. Tuy nhiên, thay vì cân nhắc nhiều yếu tố như calories, đường, chất béo để khuyến nghị cho người bệnh, khuyến nghị ở đề án này nhắm đến những người khỏe mạnh hoặc không có chế độ ăn khắc nghiệt nên bỏ qua các yếu tố trên mà dựa vào sở thích của người dùng và flowchart rút ra từ bài nghiên cứu.

## 3 Bộ dữ liệu

Bộ dữ liệu gồm tập train và test được thu thập bằng 2 phương thức khác nhau. Bộ dữ liệu train bao gồm 970 món ăn và thức uống của người Việt được crawl từ 2 nguồn là <https://bepmina.vn/> và <https://www.cet.edu.vn/>. Có tổng cộng 8 thuộc tính cho mỗi món ăn, chi tiết trong bảng 1

Thuộc tính	Định nghĩa
ID món ăn	ID của món ăn, mỗi ID là độc nhất nhằm phân biệt các món ăn với nhau
Tên món ăn	Tên của món ăn
Thể loại món ăn	Món ăn có thể là ăn vặt, bún, canh, cháo,....
Nguyên liệu chính	Thành phần chính tạo nên món ăn
ID bữa ăn	Một món ăn có thể được dùng vào buổi sáng, trưa, tối hoặc cho ăn tiệc. ID bữa ăn được lấy bằng cách tính tổng các số sau với 1: bữa sáng, 2: bữa trưa, 3: bữa tối, 4: tiệc. Ví dụ món ăn được vào buổi sáng và trưa sẽ có ID là 3 (1+2)
Thể loại bữa ăn	Món chay hay mặn
Giá ước tính	Giá ước tính cần để chế biến món ăn
Vùng miền	Nguồn gốc (đặc sản) của món ăn, có thể là Bắc, Trung, Nam hoặc nguồn gốc từ nước ngoài

**Bảng 1:** Định nghĩa chi tiết về các thuộc tính trong bộ dữ liệu

Trong khi đó, bộ dữ liệu người dùng được thu thập bằng cách khảo sát 20 người ngẫu nhiên nhằm chọn ra 100 bữa ăn lý tưởng theo sở thích của từng người. Mỗi người tham gia khảo sát phải chọn ra 5 bữa ăn, trong đó mỗi bữa ăn phải đảm bảo ít nhất 1 món tráng miệng.

## 4 Phương pháp

### 4.1 KNN

#### 4.1.1 Định nghĩa

K-nearest neighbor (KNN) là một trong những thuật toán học có giám sát đơn giản nhất trong Machine Learning. Ý tưởng của KNN là tìm ra output của bộ dữ liệu dựa trên thông tin của những dữ liệu training gần nó nhất.

#### 4.1.2 Quy trình làm việc của thuật toán KNN

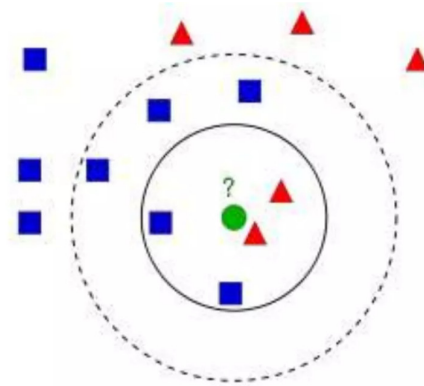
**Bước 1:** Xác định tham số  $K$  = số láng giềng gần nhất

**Bước 2:** Tính khoảng cách đối tượng cần phân lớp với tất cả các đối tượng trong data trainning

**Bước 3:** Sắp xếp khoảng cách theo thứ tự tăng dần và xác định  $K$  láng giềng gần nhất với đối tượng cần phân lớp

**Bước 4:** Lấy tất cả các lớp của  $K$  láng giềng gần nhất

**Bước 5:** Dựa vào phần lớn lớp của  $K$  để xác định lớp cho đối tượng cần phân lớp



**Hình 1:** Mô tả thuật toán KNN

#### 4.1.3 Hướng tiếp cận

Để áp dụng phương pháp KNN vào đề tài của chúng em, chúng em tạo ra một bộ dữ liệu mới từ bộ dữ liệu ban đầu. Chúng em tạo ra bộ dữ liệu mới theo các bước như sau:

**Bước 1:** Lấy ra một món ăn từ bộ dữ liệu làm món ăn chính

**Bước 2:** Chọn thêm 3 món để tạo thành 1 bữa ăn sao cho phải có món tráng miệng và món nước và không có 2 món trùng nhau.

**Bước 3:** Với mỗi món chính ta gán được 10 bàn ăn khác nhau

**Bước 4:** Lặp lại các bước 1-2-3 cho đến khi hết các món trong bộ dữ liệu cũ  
Sau khi hoàn thành bộ dữ liệu của chúng em có 970 dòng và 4 cột. Do từ bộ dữ liệu ban đầu mỗi món chính tạo ra 10 bữa ăn khác nhau nên bộ dữ liệu mới lớn hơn bộ dữ liệu cũ gấp 10 lần

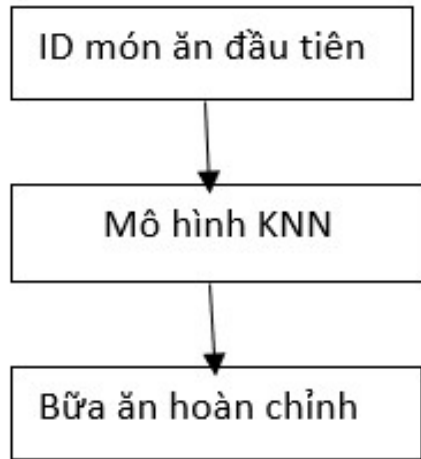
x	y_1	y_2	y_3
12.0	850.0	760.0	176.0
12.0	845.0	777.0	954.0
12.0	841.0	282.0	161.0
12.0	852.0	87.0	946.0
12.0	852.0	763.0	157.0

**Hình 2:** Bộ dữ liệu dùng cho thuật toán KNN

Số thứ tự	Tên thuộc tính	Kiểu dữ liệu	Mô tả
1	x	float	ID món ăn chính
2	y_1	float	ID món ăn thứ 2
3	y_2	float	ID món ăn thứ 3
4	y_3	float	ID món ăn thứ 4

**Bảng 2:** Mô tả chi tiết bộ dữ liệu được sử dụng cho phương pháp KNN

#### 4.1.4 Triển khai



**Hình 3:** Mô tả bài toán

**Input:** ID món ăn đầu tiên do người dùng chọn

**Output:** Bữa ăn hoàn chỉnh gồm ID 4 món ăn

Chúng em sử dụng mô hình KNN với  $K=7$  và bộ dữ liệu mới được tạo để thực hiện bài toán

## 4.2 Content-base

### 4.2.1 Định nghĩa

Content-based là một phương pháp mà ở đó sẽ dựa vào thông tin các món ăn qua đó sẽ đánh giá các đặc tính tương đồng của các món ăn. Qua đó sẽ đưa những món ăn có thể kết hợp được với nhau lên khuyến nghị qua đó sẽ tạo ra một bữa ăn đầy đủ. .

### 4.2.2 Hướng tiếp cận

Với mục đích là giúp tạo thành một bữa ăn hoàn chỉnh với các món ăn được khuyến nghị tuần tự các món ăn dựa trên sở thích của các người dùng tương

tự nhau

Một bữa ăn đầy đủ ở đây bao gồm 2 món chính với các thể loại món ăn đa dạng, 1 món nước gồm canh, lẩu,... dùng kèm với 1 món tráng miệng bao gồm các loại nước uống, kem, rau câu,... Với sự kết hợp như trên sẽ tạo ra một bữa ăn cơ bản với đầy đủ các thành phần

### 4.2.3 Vận hành

Để áp dụng khuyến nghị dựa trên nội dung vào bài toán của thì chúng em đã tiến hành dựa trên 2 cơ sở:

**Thành phần đầu vào là một sở thích của người dùng:**

- + Hệ thống sẽ khuyến nghị một danh sách món ăn ban đầu dựa trên sở thích của người dùng.
- + Sau khi chọn một món ăn thì hệ thống sẽ kiểm tra xem món ăn đó thuộc thể loại món ăn nào để giảm số lượng biến đếm của món ăn đó.
- + Tiến hành tính toán lại để đưa ra một danh sách món ăn khác.(Nếu thể loại món ăn nào đã đạt tới giới hạn tối đa thì thể loại đó sẽ không còn xuất hiện nữa ở trên danh sách)
- + Tiến hành chọn cho đến khi đủ 4 món để hoàn thành một bữa ăn hoàn chỉnh.

**Thành phần đầu vào là một món ăn mà người dùng chọn:**

- + Hệ thống sẽ khuyến nghị một danh sách một món ăn ban đầu dựa vào thể loại món ăn mà người dùng đã chọn.
- + Sau khi chọn một món ăn thì hệ thống sẽ kiểm tra xem món ăn đó thuộc thể loại món ăn nào để giảm số lượng biến đếm của món ăn đó.
- + Tiến hành tính toán lại để đưa ra một danh sách món ăn khác.(Nếu thể loại món ăn nào đã đạt tới giới hạn tối đa thì thể loại đó sẽ không còn xuất hiện nữa ở trên danh sách)
- + Tiến hành chọn cho đến khi đủ 4 món để hoàn thành một bữa ăn hoàn chỉnh.

### 4.2.4 Triển khai

(Hình 4)

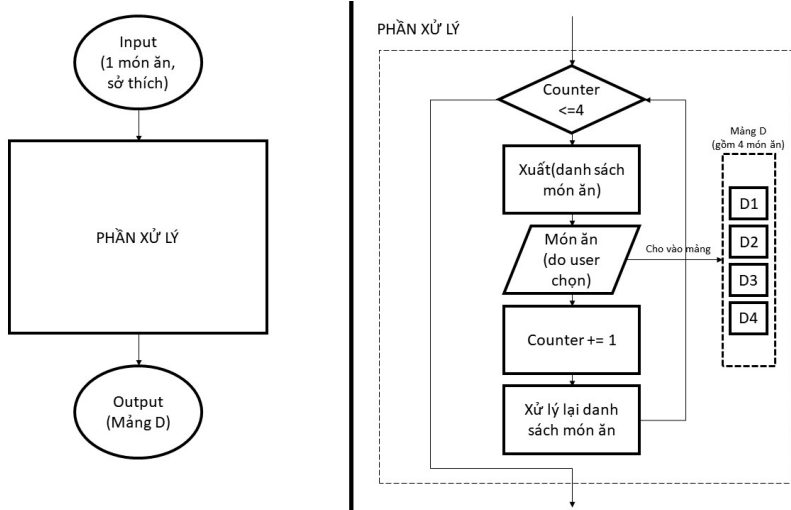
Về cơ bản thì 2 phần trên của content-based chỉ khác nhau ở cách xử lý phần đầu vào và phần còn lại có sự tương đồng với nhau rất lớn

## 5 Thực nghiệm

### 5.1 KNN

#### 5.1.1 Đánh giá

Chúng em chia bộ dữ liệu dùng cho mô hình KNN theo tỉ lệ 80-20 để thực hiện đánh giá. Do đánh giá mô hình KNN cho bài toán này là khó đối với chúng em



**Hình 4:** Mô tả cách áp dụng content-based

nên chúng em xin phép thầy cho nhóm em đánh giá theo phương pháp chủ quan như sau. Khi mô hình cho ra một bữa ăn hoàn chỉnh thì xét xem bữa ăn đó đạt chính xác khi thỏa những yêu cầu sau: Phải có 4 món khác nhau và phải có 2 món tráng miệng và món nước. Nếu thỏa yêu cầu trên thì mô hình cho ra một bữa ăn chính xác. Và sau khi kiểm tra trên tập test thì mô hình đạt chính xác 92%.

### 5.1.2 Demo

```

# Nhập món đầu tiên
input = 'Tép rang lá chanh'

# dự đoán
predict(input)

```

[31]

... Menu đề xuất là:

ID món ăn	Tên món ăn	Thể loại món ăn	Giá ước tính
0	858.0 Súp Tôm Dinh Dưỡng	Súp	150000
1	62.0 Canh Chua Cá Điêu Hồng	Canh-Lẩu	135000
2	140.0 Chè bí đỏ	Tráng miệng	25000

Tổng số tiền ước tính cho thực đơn của quý khách là: 310000 vnd

**Hình 5:** Demo mô hình KNN



## 5.2 Content base

### 5.2.1 Đánh giá

Mặc dù có ý tưởng thực hiện nhưng nhóm vẫn chưa có thể hoàn toàn kết nối được các thể loại món ăn do nguồn dữ liệu làm tay khá là hạn chế. Thêm vào đó sự thiếu hụt về kiến thức lĩnh vực này và còn thiếu một nguồn tham khảo hợp lý giữa các thể loại khi phối hợp với nhau để tạo nên một bữa ăn nên về cơ bản nhóm chỉ có thể xuất ra một bàn ăn hoàn chỉnh chứ các món ăn vẫn chưa thực sự có sự liên kết với nhau như ý muốn. Qua đó để đánh giá độ chính xác thì nếu để cho ra một bữa ăn hoàn chỉnh thì độ chính xác rất cao, mà còn theo các thể loại sao cho phù hợp thì độ chính xác gần như rất thấp do các vấn đề đã nêu trên

### 5.2.2 Demo



Hình 6: Chạy thử mô hình khuyến nghị dựa trên nội dung

## 6 Kết luận

Các hệ thống khuyến nghị đang dần đóng vai trò quan trọng trong cuộc sống hiện nay và là bước ngoặt của nhiều ngành và ở đây cũng có thể về lĩnh vực đồ ăn. Khi có quá nhiều món ăn và nhiều người vẫn chưa biết họ phải ăn gì và nhu cầu đòi hỏi một bữa ăn phù hợp. Theo ý tưởng của chúng em thì đã đưa ra được một hệ thống khuyến nghị dựa trên KNN và content based với

một số điều chỉnh. Tuy nhiên vẫn còn rất nhiều thách thức trong sự phù hợp giữa các món ăn trong bữa ăn được đề xuất. Tương lai nếu có thể thì có thể phát triển một hệ khuyến nghị mạnh mẽ hơn bằng cách cải thiện hiểu biết về lĩnh vực ẩm thực

## Tài liệu

- [1] Rafieian, B., Costa-jussà, M.R.: E-commerce content and collaborative-based recommendation using k-nearest neighbors and enriched weighted vectors. In: Proceedings of Workshop on Natural Language Processing in E-Commerce, pp. 1–10 (2020)
- [2] Dhali, D.: A food recommendation system combining inceptionv3 and knn considering calorie estimated from images. PhD thesis, Dublin, National College of Ireland (2021)